

# On the approximation of high-dimensional differential equations in the hierarchical Tucker format

Andreas Arnold · Tobias Jahnke

Received: 9 October 2012 / Accepted: 13 September 2013 / Published online: 15 October 2013  
© Springer Science+Business Media Dordrecht 2013

**Abstract** The hierarchical Tucker format is a way to decompose a high-dimensional tensor recursively into sums of products of lower-dimensional tensors. The number of degrees of freedom in such a representation is typically many orders of magnitude lower than the number of entries of the original tensor. This makes the hierarchical Tucker format a promising approach to solve ordinary differential equations for high-dimensional tensors. In order to propagate the approximation in time, differential equations for the parameters of the hierarchical Tucker format are derived from the Dirac-Frenkel variational principle. We prove an error bound for the dynamical approximation in the hierarchical Tucker format by extending previous results of Koch and Lubich for the non-hierarchical Tucker format.

**Keywords** High-dimensional differential equations · Tensor approximation · Hierarchical Tucker format · Dirac-Frenkel variational principle · Error bounds

**Mathematics Subject Classification (2010)** 15A69 · 65F99 · 65L05

## 1 Introduction

Differential equations on high-dimensional state spaces arise in many different fields and applications. Typical examples are the Black-Scholes equation for basket options in mathematical finance, the chemical master equation used in systems biology to

---

Communicated by Christian Lubich.

A. Arnold · T. Jahnke (✉)

Department of Mathematics, Institute for Applied and Numerical Mathematics, Karlsruhe Institute of Technology, Kaiserstr. 93, 76133 Karlsruhe, Germany  
e-mail: [jahnke@kit.edu](mailto:jahnke@kit.edu)

A. Arnold

e-mail: [arnold@kit.edu](mailto:arnold@kit.edu)

model the stochastic dynamics of a gene regulatory network, or the Schrödinger equation which describes the dynamics of the particles within a molecule according to the laws of quantum mechanics. Solving such problems numerically is notoriously difficult due to the fact that for standard discretizations the number of unknowns grows exponentially with respect to the dimension (“curse of dimensionality”). Hence, in high-dimensional approximations the main challenge is to compress the problem in such a way that the reduced equation can be solved numerically and nevertheless provides an acceptable approximation to the solution of the full problem.

If a multivariate function on a hyper-rectangle is discretized by an equidistant grid, the function values at the grid points can be regarded as the entries of a high-dimensional tensor. Hence, the approximation of high-dimensional functions and tensors is closely related, and for  $N_1, \dots, N_d \in \mathbb{N}$  we will consider tensors  $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$  as functions  $Y = Y(i_1, \dots, i_d)$  with  $i_k \in \{1, \dots, N_k\}$ . For a straightforward, entry-wise representation of  $Y$ , all  $\prod_{k=1}^d N_k$  values have to be stored. This number can be significantly reduced if  $Y$  can be approximated by a sum of products of univariate functions. Many such representations have been proposed, and an overview over the literature is given in [3, 9]. A particularly useful and popular representation is the orthogonal Tucker format

$$Y(i_1, \dots, i_d) \approx \tilde{Y}(i_1, \dots, i_d) = \sum_{j_1=1}^{r_1} \dots \sum_{j_d=1}^{r_d} a(j_1, \dots, j_d) \prod_{k=1}^d U_{j_k}^k(i_k), \tag{1.1}$$

see, e.g., [9, Sect. 4]. Here,  $a \in \mathbb{R}^{r_1 \times \dots \times r_d}$  is the core tensor of coefficients, and for every direction  $k \in \{1, \dots, d\}$ ,  $U_1^k, \dots, U_{r_k}^k$  is a set of univariate basis functions which satisfy the orthonormality conditions

$$\sum_{i_k=1}^{N_k} U_j^k(i_k)U_m^k(i_k) = \delta_{j,m} := \begin{cases} 1 & \text{if } j = m \\ 0 & \text{else} \end{cases}$$

for all  $j, m \in \{1, \dots, r_k\}$ . From the perspective of linear algebra, every

$$U_j^k : \{1, \dots, N_k\} \rightarrow \mathbb{R}$$

can be interpreted as a vector. Hence, (1.1) is a high-dimensional generalization of the well-known singular value decomposition, and  $\tilde{Y}$  is a low-rank-approximation of  $Y(i_1, \dots, i_d)$ , cf. [1]. For the representation (1.1) of  $\tilde{Y}$  only

$$\prod_{k=1}^d r_k + \sum_{k=1}^d r_k N_k$$

degrees of freedom have to be stored, which is a significant improvement if  $r_k \ll N_k$  for all  $k$ . Unfortunately, the fact that **all** possible combinations appear in the right-hand side of (1.1) causes again an exponential growth of the data (unless  $r_k$  decreases to 1 when  $k \rightarrow \infty$ ). For example, if  $r_1 = \dots = r_d = r$ , then the core tensor  $a(j_1, \dots, j_d)$  has  $r^d$  entries.

The hierarchical Tucker format avoids the disadvantageous growth of degrees of freedom by using the approach (1.1) in a recursive way. For example, with univariate functions  $U_{j_k}^{(k)} : \{1, \dots, N_k\} \rightarrow \mathbb{R}$  and coefficients  $\mathcal{B}_{m,j_1,j_2}^{(1,2)}, \mathcal{B}_{l,j_3,j_4}^{(3,4)} \in \mathbb{R}$ , one can define bivariate functions

$$U_m^{(1,2)}(i_1, i_2) = \sum_{j_1=1}^{r_1} \sum_{j_2=1}^{r_2} \mathcal{B}_{m,j_1,j_2}^{(1,2)} U_{j_1}^{(1)}(i_1) U_{j_2}^{(2)}(i_2)$$

$$U_l^{(3,4)}(i_3, i_4) = \sum_{j_3=1}^{r_3} \sum_{j_4=1}^{r_4} \mathcal{B}_{l,j_3,j_4}^{(3,4)} U_{j_3}^{(3)}(i_3) U_{j_4}^{(4)}(i_4),$$

and then construct a four-dimensional tensor via

$$Y(i_1, i_2, i_3, i_4) = \sum_{m=1}^{r_{(1,2)}} \sum_{l=1}^{r_{(3,4)}} \mathcal{B}_{m,l}^{(1,2,3,4)} U_m^{(1,2)}(i_1, i_2) U_l^{(3,4)}(i_3, i_4)$$

for some  $\mathcal{B}_{m,l}^{(1,2,3,4)} \in \mathbb{R}$ . This idea has first been used in quantum chemistry to solve high-dimensional Schrödinger equations [13, 15, 16] and later been investigated from a mathematical point of view in [2, 4, 12, 14]; see also [10] for a useful MATLAB toolbox. A formal definition of the hierarchical Tucker format will be given in Sect. 2. Due to the recursive definition, the hierarchical version is technically more involved than the standard Tucker format, but as we will discuss in Sect. 2, this avoids the exponential growth of the degrees of freedom.

The problem how a **given** tensor can be approximated in the hierarchical Tucker format up to a prescribed error tolerance has been studied in [2]. In the context of high-dimensional differential equations, however, an approximation for the **unknown** solution is sought after, and for time-dependent differential equations the question arises how an approximation in the hierarchical Tucker format can be propagated in time. For the non-hierarchical Tucker format, equations of motion for the core tensor and the basis functions have been derived in [8, 11] from the Dirac-Frenkel variational principle, and error bounds for the dynamical low-rank approximation in the non-hierarchical Tucker format have been proven in [8]. The main contribution of our article is to extend these results to the hierarchical Tucker format. We remark, however, that very similar results have independently and simultaneously been obtained in [12].

In Sect. 2 we introduce our notation, define the hierarchical Tucker format and discuss some of its properties. In Sect. 3, the hierarchical Tucker format is applied to approximate high-dimensional initial-value problems. In particular, differential equations for the parts of the representation are derived from the Dirac-Frenkel variational principle. The last three sections are devoted to the analysis of the accuracy of this approximation. The analysis is based on matricizations of tensors which enable a compact representation of certain projections defined in Sect. 4. The second important ingredient for the error analysis is curvature bounds which are proven in Sect. 5. Finally, an error bound for the approximation of high-dimensional differential equations in the hierarchical Tucker format is proven in Sect. 6.

## 2 Hierarchical low-rank representation of tensors: the $\mathcal{H}$ -Tucker format

A tensor  $T \in \mathbb{R}^{N_1 \times \dots \times N_d}$  is a  $d$ -dimensional array with  $\prod_{j=1}^d N_j$  entries. Defining the index sets

$$I_j = \{1, \dots, N_j\}, \quad j \in \{1, \dots, d\}$$

leads to the natural interpretation of tensors as multivariate functions

$$T : I_1 \times \dots \times I_d \rightarrow \mathbb{R}, \quad T = T(i_1, \dots, i_d).$$

This interpretation is advantageous as it simplifies the construction of high-dimensional tensors as products of low-dimensional ones which are considered as functions that only depend on a subset  $\mu = \{\mu_1, \dots, \mu_m\} \subseteq \{1, \dots, d\}$  of directions, i.e.

$$T : I_\mu \rightarrow \mathbb{R} \quad \text{with } I_\mu = I_{\mu_1} \times \dots \times I_{\mu_m}.$$

In order to define the multiplication of two tensors  $U : I_\mu \rightarrow \mathbb{R}$  and  $V : I_\nu \rightarrow \mathbb{R}$  for  $\mu, \nu \subseteq \{1, \dots, d\}$ , both tensors are interpreted as functions  $\tilde{U}$  and  $\tilde{V}$  in **all** variables  $\kappa = \{\kappa_1, \dots, \kappa_k\} = \mu \cup \nu$ . For example, if  $U = U(i_2, i_4, i_7)$  and  $V = V(i_2, i_5)$ , then we let  $\tilde{U}(i_2, i_4, i_5, i_7) = U(i_2, i_4, i_7)$  for all  $i_5 \in I_5$  and  $\tilde{V}(i_2, i_4, i_5, i_7) = V(i_2, i_5)$  for all  $i_4 \in I_4$  and  $i_7 \in I_7$ . Then, the product  $U \cdot V$  is to be understood as the pointwise product of  $\tilde{U}$  and  $\tilde{V}$ , i.e.

$$(U \cdot V) : I_\kappa \rightarrow \mathbb{R}, \quad (U \cdot V)(i_{\kappa_1}, \dots, i_{\kappa_k}) = \tilde{U}(i_{\kappa_1}, \dots, i_{\kappa_k}) \tilde{V}(i_{\kappa_1}, \dots, i_{\kappa_k}).$$

Besides multiplication of tensors, a second operation will be important in this article. Let  $\mu = \{\mu_1, \dots, \mu_m\}$  and  $\nu = \{\nu_1, \dots, \nu_n\}$  be subsets of  $\{1, \dots, d\}$ , let again  $\kappa = \{\kappa_1, \dots, \kappa_k\} = \mu \cup \nu$ , and let  $\lambda = \{\lambda_1, \dots, \lambda_m\} = (\mu \cup \nu) \setminus (\mu \cap \nu)$  be the indices which occur either in  $\mu$  or in  $\nu$ . Then, with the convention that  $\mathbb{R}^\emptyset = \mathbb{R}$ , we define

$$\langle \cdot, \cdot \rangle : \mathbb{R}^{N_{\mu_1} \times \dots \times N_{\mu_m}} \times \mathbb{R}^{N_{\nu_1} \times \dots \times N_{\nu_n}} \rightarrow \mathbb{R}^{N_{\lambda_1} \times \dots \times N_{\lambda_l}}$$

$$\langle U, V \rangle (i_{\lambda_1}, \dots, i_{\lambda_l}) = \sum_{i_{\kappa_1} \in I_{\kappa_1}}^{(\kappa_1 \in \mu \cap \nu)} \dots \sum_{i_{\kappa_k} \in I_{\kappa_k}}^{(\kappa_k \in \mu \cap \nu)} (U \cdot V)(i_{\kappa_1}, \dots, i_{\kappa_k}) \tag{2.1}$$

where the symbol  $\sum_{i_{\kappa_j} \in I_{\kappa_j}}^{(\kappa_j \in \mu \cap \nu)}$  means that the sum is taken only if  $\kappa_j \in \mu \cap \nu$ . Hence, after multiplying the two tensors we take the sum over all common directions of the two tensors. For example, in case of the tensors  $U = U(i_2, i_3, i_5, i_9)$  and  $V = V(i_2, i_4, i_5)$ , (2.1) simply reduces to

$$\begin{aligned} \langle U, V \rangle (i_3, i_4, i_9) &= \sum_{i_2 \in I_{\kappa_2}} \sum_{i_5 \in I_{\kappa_5}} U(i_2, i_3, i_5, i_9) V(i_2, i_4, i_5) \\ &= \sum_{i_2=1}^{N_{\kappa_2}} \sum_{i_5=1}^{N_{\kappa_5}} U(i_2, i_3, i_5, i_9) V(i_2, i_4, i_5). \end{aligned}$$

The result is a tensor which depends on all directions which are not common directions of  $U$  and  $V$ . Hence, if  $U$  and  $V$  depend on exactly the same directions (i.e.  $\mu \cap \nu = \mu = \nu$ ), then  $\langle U, V \rangle \in \mathbb{R}^\emptyset = \mathbb{R}$  is a real number, and it can be verified that  $\langle \cdot, \cdot \rangle$  defines a scalar product in this case. If both  $U = U(i_j)$  and  $V = V(i_j)$  are one-dimensional tensors with respect to the same direction, then this scalar product coincides with the Euclidean scalar product of the two vectors  $(U(1), \dots, U(N_j))$  and  $(V(1), \dots, V(N_j))$ . This is why we have chosen the symbol  $\langle \cdot, \cdot \rangle$ .

The numerical evaluation of  $\langle U, V \rangle$  can be simplified if  $U$  and  $V$  are products of tensors. If, for example,  $U = U(i_2, i_5) = U^{[2]}(i_2)U^{[5]}(i_5)$  and  $V = V(i_2, i_5) = V^{[2]}(i_2)V^{[5]}(i_5)$ , then

$$\begin{aligned} \langle U, V \rangle &= \sum_{i_2 \in I_{k_2}} \sum_{i_5 \in I_{k_5}} U^{[2]}(i_2)V^{[2]}(i_2)U^{[5]}(i_5)V^{[5]}(i_5) \\ &= \langle U^{[2]}, V^{[2]} \rangle \cdot \langle U^{[5]}, V^{[5]} \rangle. \end{aligned} \tag{2.2}$$

This will be frequently used in the proofs below.

The following definitions provide the framework of a hierarchical tensor structure.

**Definition 2.1** (dimension tree) A dimension tree  $T$  is a binary tree which represents a recursive decomposition of the set  $\{1, \dots, d\}$ . Every node  $\mathcal{N} \in T$  is a selection of directions,  $\mathcal{N} \subseteq \{1, \dots, d\}$ , where  $\mathcal{N}_0 = \text{root}(T) = \{1, \dots, d\}$  is called the root of the tree. Apart from the root, every node  $\mathcal{N} \in T$  is linked by an edge with exactly one father  $\tilde{\mathcal{N}} \supset \mathcal{N}$ . Conversely, every node is either linked with two successors (child nodes)  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ , or it does not have any successors at all. The successors of  $\mathcal{N}$  have the properties that

$$\mathcal{N}_1 \cup \mathcal{N}_2 = \mathcal{N} \quad \text{and} \quad \mathcal{N}_1 \cap \mathcal{N}_2 = \emptyset.$$

$\mathcal{N}$  is an interior node if

$$\mathcal{N} \in \mathcal{I}(T) := \{\tilde{\mathcal{N}} \in T : \tilde{\mathcal{N}} \text{ has successors}\},$$

and  $\mathcal{N}$  is called a leaf of the dimension tree  $T$  if

$$\mathcal{N} \in \mathcal{L}(T) := \{\tilde{\mathcal{N}} \in T : \tilde{\mathcal{N}} \text{ has no successors}\} = \{\{1\}, \dots, \{d\}\} = T \setminus \mathcal{I}(T).$$

Figure 1 shows an example of a dimension tree with inner nodes  $\mathcal{N}_0, \mathcal{N}_1$  and leaves  $\mathcal{N}_2, \mathcal{N}_{11}, \mathcal{N}_{12}$ .

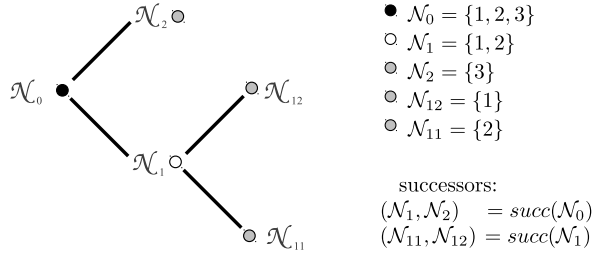
**Definition 2.2** (hierarchical Tucker format) Let  $T$  be a dimension tree with root  $\mathcal{N}_0$  and let  $(r_{\mathcal{N}})_{\mathcal{N} \in T}$  be a family of positive integers with  $r_{\mathcal{N}_0} = 1$ .  $Y$  is a tensor in the hierarchical Tucker format if and only if there exist **transfer tensors**

$$\mathcal{B}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \quad \text{for all } \mathcal{N} \in \mathcal{I}(T) \text{ with } (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$$

and univariate functions

$$U_i^\ell : I_\omega \rightarrow \mathbb{R} \quad \text{for all } \ell = \{\omega\} \in \mathcal{L}(T), \omega \in \{1, \dots, d\}, i = 1, \dots, r_\ell$$

**Fig. 1** Dimension tree  $T$



such that  $Y$  can be represented by

$$Y = \sum_{j=1}^{r_{\mathcal{N}_{01}}} \sum_{l=1}^{r_{\mathcal{N}_{02}}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_{01}} \cdot U_l^{\mathcal{N}_{02}}, \quad (\mathcal{N}_{01}, \mathcal{N}_{02}) = succ(\mathcal{N}_0)$$

with  $\mathcal{N}$ -frames

$$U_i^{\mathcal{N}} : I_{\mathcal{N}} \rightarrow \mathbb{R} \quad \text{for all } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}, i = 1, \dots, r_{\mathcal{N}}$$

defined by the recursion

$$U_i^{\mathcal{N}} = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \quad \text{for all } \mathcal{N} \in \mathcal{I}(T), (\mathcal{N}_1, \mathcal{N}_2) = succ(\mathcal{N}).$$

The family  $(r_{\mathcal{N}})_{\mathcal{N} \in T}$  is called the **hierarchical representation rank** of  $Y$ . The set of all tensors in the hierarchical Tucker format with hierarchical representation rank  $(r_{\mathcal{N}})_{\mathcal{N} \in T}$  is denoted by  $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ .

*Remark*

1. Since the transfer tensors in Definition 2.2 are always three-dimensional objects, we do not interpret these particular tensors as multivariate functions.
2. Every tensor  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  can be reconstructed from the transfer tensors  $(\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}$  and the functions  $(U^\ell)_{\ell \in \mathcal{L}(T)}$  via the above recursion. The  $\mathcal{N}$ -frames at the interior nodes can be considered as auxiliary variables which are only used to reconstruct the full tensor  $Y$  if necessary.
3. Definitions 2.1 and 2.2 have been adapted from Definitions 3.3 and 3.6 in [2]. At this point, it is not yet assumed that the  $\mathcal{N}$ -frames are linearly independent. For the approximation of time-dependent problems, however, we will later have to assume that the  $\mathcal{N}$ -frames at every node  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  are an orthonormal basis, and that the above representation is not redundant, cf. Assumptions 1 and 2 below.
4. A special case of the hierarchical Tucker format is the so-called tensor train format which has been investigated, e.g., in [5].

At this point, a comparison with the standard (non-hierarchical) Tucker format is helpful. For the hierarchical Tucker format, an equation similar to (1.1) can be

obtained, but due to the recursive definition of the  $\mathcal{N}$ -frames, the corresponding formula is much more complicated. The crucial difference is the fact that the number of terms in the linear combination can be considerably lower. In fact, the total number of degrees of freedom in the hierarchical Tucker format is at most

$$\sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} r_{\mathcal{N}} r_{\mathcal{N}_1} r_{\mathcal{N}_2} + \sum_{k=1}^d r_{\{k\}} N_k$$

because every  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  is defined by the  $r_\ell$  basis functions at the leaves  $\ell = \{k\}$  with  $k = 1, \dots, d$  and by the entries of all transfer tensors, cf. Lemma 3.7 in [2].

**Lemma 1** (orthonormality of transfer tensors) *Let  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be a tensor represented by  $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)})$ . Then, the following statements are equivalent:*

- (a) *The  $\mathcal{N}$ -frames  $(U^{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$  are orthonormal with respect to mapping (2.1), i.e.*

$$\langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}} \rangle = \delta_{i,j}$$

*holds for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and  $i, j \in \{1, \dots, r_{\mathcal{N}}\}$ .*

- (b) *For all  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ ,  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$  and  $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$  the equation*

$$\sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j, l}^{\mathcal{N}} \mathcal{B}_{i_2, j, l}^{\mathcal{N}} = \delta_{i_1, i_2} \tag{2.3}$$

*holds and the  $\ell$ -frames  $(U^\ell)_{\ell \in \mathcal{L}(T)}$  are orthonormal with respect to mapping (2.1), i.e.*

$$\langle U_i^\ell, U_j^\ell \rangle = \delta_{i,j}$$

*holds for all  $\ell \in \mathcal{L}(T)$ ,  $i, j \in \{1, \dots, r_\ell\}$ .*

*Proof* Let  $\mathcal{N} \in \mathcal{I}(T)$  be a node in the interior of the dimension tree,  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$  and  $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$ . If (a) holds, then it follows from Definition 2.2 that

$$\begin{aligned} \langle U_{i_1}^{\mathcal{N}}, U_{i_2}^{\mathcal{N}} \rangle &= \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j_1, l_1}^{\mathcal{N}} \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}} \langle U_{j_1}^{\mathcal{N}_1} \cdot U_{l_1}^{\mathcal{N}_2}, U_{j_2}^{\mathcal{N}_1} \cdot U_{l_2}^{\mathcal{N}_2} \rangle \\ &= \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j_1, l_1}^{\mathcal{N}} \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}} \langle U_{j_1}^{\mathcal{N}_1}, U_{j_2}^{\mathcal{N}_1} \rangle \langle U_{l_1}^{\mathcal{N}_2}, U_{l_2}^{\mathcal{N}_2} \rangle \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j, l}^{\mathcal{N}} \mathcal{B}_{i_2, j, l}^{\mathcal{N}} \end{aligned} \tag{2.4}$$

which proves (a)  $\Rightarrow$  (b). Now assume that (b) holds, and that  $\mathcal{N}_1, \mathcal{N}_2 \in \mathcal{L}(T)$ . Then, Eq. (2.4) yields that

$$\langle U_{i_1}^{\mathcal{N}}, U_{i_2}^{\mathcal{N}} \rangle = \delta_{i_1, i_2}$$

for the father  $\mathcal{N}$  of  $\mathcal{N}_1$  and  $\mathcal{N}_2$ , and by induction over the levels of the tree we obtain (b)  $\Rightarrow$  (a). □

**Assumption 1** For every tensor  $Y$  in the hierarchical Tucker format we assume the  $\mathcal{N}$ -frames  $(U^{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$  to be orthonormal with respect to mapping (2.1), i.e.

$$\langle U_i^{\mathcal{N}}, U_j^{\mathcal{N}} \rangle = \delta_{i,j} \quad \text{for all } \mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}, i, j \in \{1, \dots, r_{\mathcal{N}}\}.$$

Since for any  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  we can find a representation with orthonormal  $\mathcal{N}$ -frames, Assumption 1 is not a restriction; see [2, Algorithm 3] for an orthonormalization algorithm. It will be shown in Lemma 4 that when the tensor  $Y$  approximates the time-dependent solution of a high-dimensional differential equation, the differential equations for the  $\mathcal{N}$ -frames preserve the orthonormality for all times.

An important consequence of Assumption 1 is that for  $\mathcal{N} \in \mathcal{I}(T)$  with  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ , we have

$$\langle U_k^{\mathcal{N}_1}, U_i^{\mathcal{N}_2} \rangle = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} \langle U_k^{\mathcal{N}_1}, U_j^{\mathcal{N}_1} \rangle U_l^{\mathcal{N}_2} = \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,k,l}^{\mathcal{N}} U_l^{\mathcal{N}_2}$$

for any  $k \in \{1, \dots, r_{\mathcal{N}_1}\}$  and hence

$$\sum_{k=1}^{r_{\mathcal{N}_1}} U_k^{\mathcal{N}_1} \langle U_k^{\mathcal{N}_1}, U_i^{\mathcal{N}_2} \rangle = \sum_{k=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,k,l}^{\mathcal{N}} U_k^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} = U_i^{\mathcal{N}_2}.$$

Substituting this into the recursive representation of  $Y$  yields

$$\sum_{k=1}^{r_{\mathcal{N}_1}} U_k^{\mathcal{N}_1} \langle U_k^{\mathcal{N}_1}, Y \rangle = Y \tag{2.5}$$

for all  $\mathcal{N}_1 \in T \setminus \{\mathcal{N}_0\}$ . These and similar formulas will be frequently used henceforth.

**Lemma 2** (minimal representation rank) *Let  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be a tensor in the hierarchical Tucker format represented by  $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)})$ . Assume that the tensors*

$$\langle Y, U_1^{\mathcal{N}} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$$

*are linearly independent for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ . Then,  $Y$  has minimal representation rank, which means that there exists no other family of positive integers  $(r_{\mathcal{N}}^*)_{\mathcal{N} \in T}$  and no node  $\mathcal{N}^* \in T$  with*



$$r_{\mathcal{N}}^* \leq r_{\mathcal{N}} \quad \text{for all } \mathcal{N} \in T \text{ and } r_{\mathcal{N}^*}^* < r_{\mathcal{N}^*}, \tag{2.6}$$

such that  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}}^*)_{\mathcal{N} \in T})$ .

*Proof* Suppose that  $Y$  can also be represented by  $((\tilde{\mathcal{B}}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\tilde{U}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{L}(T)})$  and that the hierarchical representation rank  $(r_{\mathcal{N}}^*)_{\mathcal{N} \in T}$  of this representation has the properties (2.6). According to (2.5) we have

$$Y = \sum_{i^*=1}^{r_{\mathcal{N}^*}^*} \tilde{U}_{i^*}^{\mathcal{N}^*} \langle Y, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle = \sum_{i=1}^{r_{\mathcal{N}^*}^*} U_i^{\mathcal{N}^*} \langle Y, U_i^{\mathcal{N}^*} \rangle.$$

Substituting the first representation into the second yields

$$Y = \sum_{i=1}^{r_{\mathcal{N}^*}^*} U_i^{\mathcal{N}^*} \sum_{i^*=1}^{r_{\mathcal{N}^*}^*} \langle U_i^{\mathcal{N}^*}, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle \langle Y, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle$$

and hence

$$\langle Y, U_i^{\mathcal{N}^*} \rangle = \sum_{i^*=1}^{r_{\mathcal{N}^*}^*} \langle U_i^{\mathcal{N}^*}, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle \langle Y, \tilde{U}_{i^*}^{\mathcal{N}^*} \rangle.$$

This is a contradiction to the assumption that  $\langle Y, U_1^{\mathcal{N}^*} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}^*}^*}^{\mathcal{N}^*} \rangle$  are linearly independent.  $\square$

*Remark* It can be shown that the converse assertion is also true.

**Assumption 2** For every tensor  $Y$  in the hierarchical Tucker format and for every  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  we assume

$$\langle Y, U_1^{\mathcal{N}} \rangle, \dots, \langle Y, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$$

to be linearly independent.

As shown in Sect. 4, under Assumption 2 the set  $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  is a smooth manifold, which is essential in our approach. Furthermore we will use the term ‘**hierarchical rank**’ instead of ‘hierarchical representation rank’ and denote the  $\mathcal{N}$ -frames as **basis functions**. Assumption 2 and Lemma 2 motivate the following definition.

**Definition 2.3** (Single-hole operator) Let  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be a tensor in the hierarchical Tucker format, which is represented by the tuple  $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^{\ell})_{\ell \in \mathcal{L}(T)})$ ,  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and  $i \in \{1, \dots, r_{\mathcal{N}}\}$ . The single-hole operator  $\mathcal{S}$  is defined as

$$\mathcal{S}(Y, \mathcal{N}, i) := \langle Y, U_i^{\mathcal{N}} \rangle.$$

This is a generalization of the single-hole functions defined in Sect. II.3.3 in [11].

*Remark* With the single-hole operator (2.5) reads

$$Y = \sum_{i=1}^{r_{\mathcal{N}}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_i^{\mathcal{N}} \tag{2.7}$$

for all  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}), \mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ .

**Definition 2.4** For  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}), \mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and  $X \in \mathbb{R}^{N_1 \times \dots \times N_d}$  the orthogonal projection of  $X$  onto the space spanned by  $U_1^{\mathcal{N}}, \dots, U_{r_{\mathcal{N}}}^{\mathcal{N}}$  is denoted by

$$\mathcal{P}_{\mathcal{N}} X = \sum_{i=1}^{r_{\mathcal{N}}} \langle X, U_i^{\mathcal{N}} \rangle U_i^{\mathcal{N}}.$$

The orthogonal projection onto the orthogonal complement is given by  $\mathcal{P}_{\mathcal{N}}^{\perp} X = X - \mathcal{P}_{\mathcal{N}} X$ .

### 3 Solving high-dimensional initial value problems via hierarchical tensor approximation

In this section we show how the hierarchical Tucker format can be used to approximate the time-dependent solution  $Y_{ex}(t) \in \mathbb{R}^{N_1 \times \dots \times N_d}$  of a high-dimensional ordinary differential equation

$$\dot{Y}_{ex} = F(Y_{ex}), \quad t \geq 0, \quad F : \mathbb{R}^{N_1 \times \dots \times N_d} \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d} \tag{3.1}$$

with a given initial value  $Y_{ex}(0) \in \mathbb{R}^{N_1 \times \dots \times N_d}$  and a function  $F$  mapping tensors to tensors. Such problems arise, e.g., when the method of lines is applied to a high-dimensional partial differential equation. If the partial differential equation is linear, then the function  $F$  is linear, too. The exact solution  $Y_{ex}(t)$  of (3.1) is a time-dependent,  $d$ -dimensional tensor with  $\prod_{\mu=1}^d N_{\mu}$  entries. Hence, classical numerical schemes for ordinary differential equations cannot be applied if  $d \gg 3$  due to the huge number of unknowns. Therefore, we are looking for an approximation  $Y \approx Y_{ex}$  which lies on the manifold

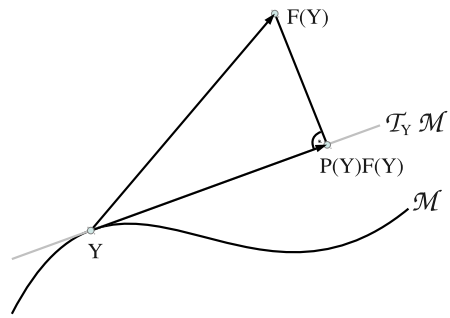
$$\mathcal{M} := \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$$

at any time. For this purpose, the variational principle of Dirac-Frenkel is applied (cf. [11]): Starting with an approximation of the initial value on the manifold, the right-hand side  $F(Y)$  is projected onto  $\mathcal{T}_Y \mathcal{M}$ , the tangent space of  $\mathcal{M}$  at  $Y(t)$ , for every  $t \geq 0$ ; see Fig. 2 for an illustration.

With  $P(Y)$  denoting the orthogonal projector on the tangent space  $\mathcal{T}_Y \mathcal{M}$  the differential equation for  $Y(t) \in \mathcal{M}$  takes the form

$$\dot{Y} = P(Y)F(Y) \in \mathcal{T}_Y \mathcal{M}, \quad t \geq 0, \quad Y(0) \in \mathcal{M} \tag{3.2}$$

**Fig. 2** Orthogonal projection on the tangent space



or equivalently

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0 \quad \forall \delta Y \in \mathcal{T}_Y \mathcal{M}.$$

Even in applications where the original problem (3.1) is linear, the new differential equation (3.2) is **nonlinear**. The great advantage of (3.2) is the fact that for the propagation of  $Y(t)$  only the transfer tensors  $\mathcal{B}^{\mathcal{N}}(t)$ ,  $\mathcal{N} \in \mathcal{I}(T)$  and the basis functions  $U_i^\ell(t)$ ,  $\ell \in \mathcal{L}(T)$  for the time-dependent representation of  $Y(t)$  have to be computed. In order to make use of this advantage, however, (3.2) has to be replaced by differential equations for  $\mathcal{B}^{\mathcal{N}}(t)$  and  $U_i^\ell(t)$  instead of  $Y(t)$ . This is the goal of this section.

Since every point  $Y$  on the manifold is determined by

$$((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)}),$$

the transfer tensors and the basis functions on the leaves of the dimension tree can be considered as the parameters of the representation. Let  $\Psi$  be the mapping which maps the parameters to the corresponding tensor, i.e.

$$\begin{aligned} \Psi : & \prod_{\substack{\mathcal{N} \in \mathcal{I}(T) \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}} \times \prod_{\{j\} \in \mathcal{L}(T)} \mathbb{R}^{N_j \times r_{\{j\}}} \rightarrow \mathbb{R}^{N_1 \times \dots \times N_d} \\ & \Psi((\tilde{\mathcal{B}}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\tilde{U}^\ell)_{\ell \in \mathcal{L}(T)}) = \tilde{Y}. \end{aligned}$$

$\tilde{Y}$  is constructed according to Definition 2.2. Then, each element  $\delta Y$  of the tangent space  $\mathcal{T}_Y \mathcal{M}$  can be written as

$$\begin{aligned} \delta Y = & \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \frac{\partial \Psi}{\partial \tilde{\mathcal{B}}_{1,j,l}^{\mathcal{N}_0}} \bigg|_{\substack{\tilde{\mathcal{B}}^{\mathcal{N}} = \mathcal{B}^{\mathcal{N}} \forall \mathcal{N} \in \mathcal{I}(T) \\ \tilde{U}^\ell = U^\ell \forall \ell \in \mathcal{L}(T)}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \\ & + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{i=1}^{r_{\mathcal{N}}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \frac{\partial \Psi}{\partial \tilde{\mathcal{B}}_{i,j,l}^{\mathcal{N}}} \bigg|_{\substack{\tilde{\mathcal{B}}^{\mathcal{N}} = \mathcal{B}^{\mathcal{N}} \forall \mathcal{N} \in \mathcal{I}(T) \\ \tilde{U}^\ell = U^\ell \forall \ell \in \mathcal{L}(T)}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \\ & + \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_\ell} \frac{\partial \Psi}{\partial \tilde{U}_i^\ell} \bigg|_{\substack{\tilde{\mathcal{B}}^{\mathcal{N}} = \mathcal{B}^{\mathcal{N}} \forall \mathcal{N} \in \mathcal{I}(T) \\ \tilde{U}^\ell = U^\ell \forall \ell \in \mathcal{L}(T)}} \delta U_i^\ell, \end{aligned} \tag{3.3}$$

where the tuple  $((\delta\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^\ell)_{\ell \in \mathcal{L}(T)})$  represents all possible variations of the transfer tensors and the basis functions on the leaves of the dimension tree. Since  $\Psi$  is linear in each argument, we find with

$$\frac{\partial \Psi}{\partial \tilde{\mathcal{B}}_{1,j,l}^{\mathcal{N}_0}} \Big|_{\substack{\tilde{\mathcal{B}}^{\mathcal{N}} = \mathcal{B}^{\mathcal{N}} \forall \mathcal{N} \in \mathcal{I}(T) \\ \tilde{U}^\ell = U^\ell \forall \ell \in \mathcal{L}(T)}} = U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \quad (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N}_0), \tag{3.4}$$

$$\frac{\partial \Psi}{\partial \tilde{\mathcal{B}}_{i,j,l}^{\mathcal{N}}} \Big|_{\substack{\tilde{\mathcal{B}}^{\mathcal{N}} = \mathcal{B}^{\mathcal{N}} \forall \mathcal{N} \in \mathcal{I}(T) \\ \tilde{U}^\ell = U^\ell \forall \ell \in \mathcal{L}(T)}} = \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \quad (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N}) \tag{3.5}$$

and

$$\frac{\partial \Psi}{\partial \tilde{U}_i^\ell} \Big|_{\substack{\tilde{\mathcal{B}}^{\mathcal{N}} = \mathcal{B}^{\mathcal{N}} \forall \mathcal{N} \in \mathcal{I}(T) \\ \tilde{U}^\ell = U^\ell \forall \ell \in \mathcal{L}(T)}} = \mathcal{S}(Y, \ell, i) \tag{3.6}$$

expressions for every partial derivative in (3.3). Therefore, the tangent space can be written as

$$\begin{aligned} \mathcal{T}_Y \mathcal{M} = \{ \delta Y \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{there exists a tuple } ((\delta\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^\ell)_{\ell \in \mathcal{L}(T)}) \\ \text{such that (3.3), (3.4), (3.5) and (3.6) hold} \}. \end{aligned}$$

*Remark* Since Assumption 1 means no restriction for the set  $\mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ , the orthogonality conditions means no restriction for the tangent space as well. Even the restriction of mapping  $\Psi$  to transfer tensors and basis functions which fulfill Assumptions 2 does not influence the tangent space since the restricted set is a dense and open subset of the domain, cf. Sect. 3.2 in [14].

We are now ready to formulate the differential equations for the transfer tensors and the basis functions at the leaves. In view of the recursive definition of  $\mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$ , it does not come at a surprise that these differential equations are defined recursively, too. The same result has also been obtained with other techniques in [12] and [14].

**Theorem 3** (equations of motion) *Let  $Y \in \mathcal{H}$ -Tucker $((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be a tensor in the hierarchical Tucker format, let  $(\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}$  the corresponding transfer tensors, and let  $(U^\ell)_{\ell \in \mathcal{L}(T)}$  the corresponding basis functions on the leaves of the dimension tree. Furthermore we define for every  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  the symmetric stiffness matrix*

$$M^{\mathcal{N}} = (m_{ij}^{\mathcal{N}})_{i,j=1}^{r_{\mathcal{N}}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}}} \quad \text{with entries } m_{ij}^{\mathcal{N}} = \langle \mathcal{S}(Y, \mathcal{N}, i), \mathcal{S}(Y, \mathcal{N}, j) \rangle \tag{3.7}$$

and its inverse

$$W^{\mathcal{N}} = (w_{ij}^{\mathcal{N}})_{i,j=1}^{r_{\mathcal{N}}} = (M^{\mathcal{N}})^{-1}. \tag{3.8}$$

Then, the differential equation

$$\dot{Y} = P(Y)F(Y), \quad t \geq 0$$

is equivalent to

$$\begin{aligned} \dot{Y} &= \sum_{j=1}^{r_{\mathcal{N}_01}} \sum_{l=1}^{r_{\mathcal{N}_02}} \dot{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_01} \cdot U_l^{\mathcal{N}_02} \\ &+ \sum_{j=1}^{r_{\mathcal{N}_01}} \sum_{l=1}^{r_{\mathcal{N}_02}} B_{1,j,l}^{\mathcal{N}_0} \dot{U}_j^{\mathcal{N}_01} \cdot U_l^{\mathcal{N}_02} + \sum_{j=1}^{r_{\mathcal{N}_01}} \sum_{l=1}^{r_{\mathcal{N}_02}} B_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_01} \cdot \dot{U}_l^{\mathcal{N}_02} \end{aligned} \tag{3.9}$$

with  $(\mathcal{N}_01, \mathcal{N}_02) = \text{succ}(\mathcal{N}_0)$ . The time derivatives of the basis functions on an interior node of the dimension tree satisfy the recursion

$$\begin{aligned} \dot{U}_i^{\mathcal{N}} &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \dot{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &+ \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} B_{i,j,l}^{\mathcal{N}} \dot{U}_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} B_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot \dot{U}_l^{\mathcal{N}_2} \end{aligned} \tag{3.10}$$

for all  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$  with  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ . On the leaves of the dimension tree we have

$$\dot{U}_i^\ell = \sum_{i'=1}^{r_\ell} w_{i',i}^\ell \mathcal{P}_\ell^\perp \langle F(Y), \mathcal{S}(Y, \ell, i') \rangle, \quad \ell \in \mathcal{L}(T), \quad i = 1, \dots, r_\ell. \tag{3.11}$$

The transfer tensors evolve according to the differential equations

$$\dot{B}_{1,j,l}^{\mathcal{N}_0} = \langle F(Y), U_j^{\mathcal{N}_01} \cdot U_l^{\mathcal{N}_02} \rangle$$

and

$$\dot{B}_{i,j,l}^{\mathcal{N}} = \sum_{i'=1}^{r_{\mathcal{N}}} w_{i',i}^{\mathcal{N}} \langle F(Y), \mathcal{S}(Y, \mathcal{N}, i') \cdot \mathcal{P}_{\mathcal{N}}^\perp(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle.$$

We remark that the crucial terms are  $\dot{U}_i^\ell$  for  $\ell \in \mathcal{L}(T)$  and  $\dot{B}_{i,j,l}^{\mathcal{N}}$  for  $\mathcal{N} \in \mathcal{I}(T)$ . If these objects are known, all  $\dot{U}_i^{\mathcal{N}}$  with  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$  can be computed recursively via (3.10). Hence, the time derivatives of the  $\mathcal{N}$ -frames at the interior nodes can be considered as auxiliary variables.

Before the proof of Theorem 3, we show that the solution of the equations of motion preserve the orthonormality of the basis functions.

**Lemma 4** (Gauge-conditions) *Let  $((\mathcal{B}^{\mathcal{N}}(t))_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell(t))_{\ell \in \mathcal{L}(T)})$  be the solution of the equations of motion of Theorem 3, and let  $((\dot{\mathcal{B}}^{\mathcal{N}}(t))_{\mathcal{N} \in \mathcal{I}(T)}, (\dot{U}^\ell(t))_{\ell \in \mathcal{L}(T)})$  be the corresponding time derivatives. Then the **Gauge-condition***

$$\langle \dot{U}_i^{\mathcal{N}}(t), U_j^{\mathcal{N}}(t) \rangle = 0$$

is satisfied for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ ,  $i, j \in \{1, \dots, r_{\mathcal{N}}\}$  and  $t \geq 0$ . Furthermore, the orthonormality of the basis functions

$$\langle U_i^{\mathcal{N}}(t), U_j^{\mathcal{N}}(t) \rangle = \delta_{i,j} \tag{3.12}$$

is preserved for all  $t \geq 0$  if (3.12) holds for  $t = 0$ .

*Proof of Lemma 4* For all  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ ,  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$  and  $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$  the equations

$$\begin{aligned} & \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{i_1, j, l}^{\mathcal{N}} \mathcal{B}_{i_2, j, l}^{\mathcal{N}} \\ &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_2, j, l}^{\mathcal{N}} \sum_{i=1}^{r_{\mathcal{N}}} w_{i_1, i}^{\mathcal{N}} \langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle \\ &= \sum_{i=1}^{r_{\mathcal{N}}} w_{i_1, i}^{\mathcal{N}} \left\langle F(Y), \mathcal{S}(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp} \left( \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_2, j, l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right) \right\rangle \\ &= 0 \end{aligned} \tag{3.13}$$

hold. Furthermore, we have

$$\langle \dot{U}_{i_1}^{\ell}, U_{i_2}^{\ell} \rangle = \left\langle \sum_{i=1}^{r_{\ell}} w_{i_1, i}^{\ell} \mathcal{P}_{\ell}^{\perp} \langle F(Y), \mathcal{S}(Y, \ell, i) \rangle, U_{i_2}^{\ell} \right\rangle = 0$$

for all  $\ell \in \mathcal{L}(T)$ ,  $i_1, i_2 \in \{1, \dots, r_{\ell}\}$ . Now suppose that

$$\langle \dot{U}_{j_1}^{\mathcal{N}_1}, U_{j_2}^{\mathcal{N}_1} \rangle = 0 \quad \text{and} \quad \langle \dot{U}_{l_1}^{\mathcal{N}_2}, U_{l_2}^{\mathcal{N}_2} \rangle = 0$$

for an interior node  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ ,  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$  for all  $j_1, j_2 \in \{1, \dots, r_{\mathcal{N}_1}\}$  and  $l_1, l_2 \in \{1, \dots, r_{\mathcal{N}_2}\}$ . Then with (3.13) we deduce

$$\begin{aligned} & \langle \dot{U}_{i_1}^{\mathcal{N}}, U_{i_2}^{\mathcal{N}} \rangle \\ &= \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{i_1, j_1, l_1}^{\mathcal{N}} \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}} \langle U_{j_1}^{\mathcal{N}_1} \cdot U_{l_1}^{\mathcal{N}_2}, U_{j_2}^{\mathcal{N}_1} \cdot U_{l_2}^{\mathcal{N}_2} \rangle \\ &+ \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j_1, l_1}^{\mathcal{N}} \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}} \langle \dot{U}_{j_1}^{\mathcal{N}_1} \cdot U_{l_1}^{\mathcal{N}_2}, U_{j_2}^{\mathcal{N}_1} \cdot U_{l_2}^{\mathcal{N}_2} \rangle \\ &+ \sum_{j_1=1}^{r_{\mathcal{N}_1}} \sum_{l_1=1}^{r_{\mathcal{N}_2}} \sum_{j_2=1}^{r_{\mathcal{N}_1}} \sum_{l_2=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i_1, j_1, l_1}^{\mathcal{N}} \mathcal{B}_{i_2, j_2, l_2}^{\mathcal{N}} \langle U_{j_1}^{\mathcal{N}_1} \cdot \dot{U}_{l_1}^{\mathcal{N}_2}, U_{j_2}^{\mathcal{N}_1} \cdot U_{l_2}^{\mathcal{N}_2} \rangle \\ &= 0 \end{aligned}$$

for all  $i_1, i_2 \in \{1, \dots, r_{\mathcal{N}}\}$ . Hence, the Gauge-conditions are satisfied, and the equations

$$\begin{aligned} \langle U_i^{\mathcal{N}}(t), U_j^{\mathcal{N}}(t) \rangle &= \langle U_i^{\mathcal{N}}(0), U_j^{\mathcal{N}}(0) \rangle + \int_0^t (\dot{U}_i^{\mathcal{N}}(s), U_j^{\mathcal{N}}(s)) + \langle U_i^{\mathcal{N}}(s), \dot{U}_j^{\mathcal{N}}(s) \rangle ds \\ &= \delta_{i,j} \end{aligned}$$

show the preservation of the orthonormality. □

*Proof of Theorem 3* In order to prove the theorem it has to be shown that the variational equation

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0$$

is satisfied for every variation  $\delta Y \in \mathcal{T}_Y \mathcal{M}$ .

(i) We start with the variation

$$\delta Y = \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \quad (\mathcal{N}_1, \mathcal{N}_2) = succ(\mathcal{N}_0) \tag{3.14}$$

for an arbitrary  $\delta \mathcal{B}^{\mathcal{N}_0} \in \mathbb{R}^{r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$ ,  $j \in \{1, \dots, r_{\mathcal{N}_1}\}$ ,  $l \in \{1, \dots, r_{\mathcal{N}_2}\}$ . Together with (3.9) this yields

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \dot{\mathcal{B}}_{1,j',l'}^{\mathcal{N}_0} U_{j'}^{\mathcal{N}_1} \cdot U_{l'}^{\mathcal{N}_2}, \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &\quad + \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j',l'}^{\mathcal{N}_0} \dot{U}_{j'}^{\mathcal{N}_1} \cdot U_{l'}^{\mathcal{N}_2}, \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &\quad + \left\langle \sum_{j'=1}^{r_{\mathcal{N}_1}} \sum_{l'=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j',l'}^{\mathcal{N}_0} U_{j'}^{\mathcal{N}_1} \cdot \dot{U}_{l'}^{\mathcal{N}_2}, \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &= \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \dot{\mathcal{B}}_{1,j,l}^{\mathcal{N}_0} \\ &= \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \langle P(Y) F(Y), U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &= \langle F(Y), \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &= \langle F(Y), \delta Y \rangle. \end{aligned}$$

This proves that

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0 \quad \text{for all } \delta Y = \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}.$$

(ii) Next, we assume  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ ,  $(\mathcal{N}_1, \mathcal{N}_2) = succ(\mathcal{N})$  and choose the variation

$$\delta Y = \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \tag{3.15}$$

for an arbitrary  $\delta\mathcal{B}^{\mathcal{N}} \in \mathbb{R}^{r_{\mathcal{N}} \times r_{\mathcal{N}_1} \times r_{\mathcal{N}_2}}$ ,  $i \in \{1, \dots, r_{\mathcal{N}}\}$ ,  $j \in \{1, \dots, r_{\mathcal{N}_1}\}$ ,  $l \in \{1, \dots, r_{\mathcal{N}_2}\}$ . Deriving both sides of (2.7) with respect to  $t$  gives that

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \left\langle \sum_{i'=1}^{r_{\mathcal{N}}} \dot{S}(Y, \mathcal{N}, i') \cdot U_{i'}^{\mathcal{N}}, \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} S(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &\quad + \left\langle \sum_{i'=1}^{r_{\mathcal{N}}} S(Y, \mathcal{N}, i') \cdot \dot{U}_{i'}^{\mathcal{N}}, \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} S(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \right\rangle \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle \dot{S}(Y, \mathcal{N}, i'), S(Y, \mathcal{N}, i) \rangle \langle U_{i'}^{\mathcal{N}}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &\quad + \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle S(Y, \mathcal{N}, i'), S(Y, \mathcal{N}, i) \rangle \langle \dot{U}_{i'}^{\mathcal{N}}, U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle P(Y)F(Y), S(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} m_{i',i}^{\mathcal{N}} \dot{\mathcal{B}}_{i',j,l}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle P(Y)F(Y), S(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \sum_{i''=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i'') \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle \left( \sum_{i'=1}^{r_{\mathcal{N}}} m_{i',i}^{\mathcal{N}} w_{i',i''}^{\mathcal{N}} \right) \end{aligned}$$

and since  $S(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \in \mathcal{T}_Y\mathcal{M}$  we obtain

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle. \end{aligned} \tag{3.16}$$

On the other hand

$$\begin{aligned} \langle F(Y), \delta Y \rangle &= \langle F(Y), \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} S(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \rangle \\ &= \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle \\ &\quad + \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i) \cdot U_{i'}^{\mathcal{N}} \rangle \mathcal{B}_{i',j,l}^{\mathcal{N}} \\ &\quad + \delta\mathcal{B}_{i,j,l}^{\mathcal{N}} \langle F(Y), S(Y, \mathcal{N}, i) \cdot \mathcal{P}_{\mathcal{N}}^{\perp}(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle \end{aligned} \tag{3.17}$$



holds. Since (3.16) is equal to (3.17), it follows that

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0 \quad \text{for all } \delta Y = \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \cdot U_j^{\mathcal{N}_1} \cdot U_j^{\mathcal{N}_2}.$$

(iii) Finally, let  $\ell \in \mathcal{L}(T)$  and consider the variation

$$\delta Y = \mathcal{S}(Y, \ell, i) \cdot \delta U_i^\ell \tag{3.18}$$

for an arbitrary  $\delta U_i^\ell : I_\ell \rightarrow \mathbb{R}, i \in \{1, \dots, r_\ell\}$ . Similar as before, the time derivative of (2.7) yields

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \left\langle \sum_{i'=1}^{r_\ell} \dot{\mathcal{S}}(Y, \ell, i') \cdot U_{i'}^\ell, \mathcal{S}(Y, \ell, i) \cdot \delta U_i^\ell \right\rangle \\ &\quad + \left\langle \sum_{i'=1}^{r_\ell} \mathcal{S}(Y, \ell, i') \cdot \dot{U}_{i'}^\ell, \mathcal{S}(Y, \ell, i) \cdot \delta U_i^\ell \right\rangle \\ &= \sum_{i'=1}^{r_\ell} \langle \dot{\mathcal{S}}(Y, \ell, i'), \mathcal{S}(Y, \ell, i) \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\ &\quad + \sum_{i'=1}^{r_\ell} \langle \mathcal{S}(Y, \ell, i'), \mathcal{S}(Y, \ell, i) \rangle \langle \dot{U}_{i'}^\ell, \delta U_i^\ell \rangle \\ &= \sum_{i'=1}^{r_\ell} \langle P(Y)F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\ &\quad + \sum_{i'=1}^{r_\ell} m_{i',i}^\ell \left\langle \sum_{i''=1}^{r_\ell} w_{i',i''}^\ell \mathcal{P}_\ell^\perp \langle F(Y), \mathcal{S}(Y, \ell, i'') \rangle, \delta U_i^\ell \right\rangle \\ &= \sum_{i'=1}^{r_\ell} \langle P(Y)F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\ &\quad + \sum_{i''=1}^{r_\ell} \langle \mathcal{P}_\ell^\perp F(Y), \mathcal{S}(Y, \ell, i'') \cdot \delta U_i^\ell \rangle \left( \sum_{i'=1}^{r_\ell} m_{i',i}^\ell w_{i',i''}^\ell \right) \end{aligned}$$

and with  $\mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \in \mathcal{T}_Y \mathcal{M}$  we get

$$\begin{aligned} \langle \dot{Y}, \delta Y \rangle &= \sum_{i'=1}^{r_\ell} \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\ &\quad + \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell^\perp \delta U_i^\ell \rangle. \end{aligned} \tag{3.19}$$

On the other hand

$$\begin{aligned} \langle F(Y), \delta Y \rangle &= \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \delta U_i^\ell \rangle \\ &= \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell \delta U_i^\ell \rangle \end{aligned}$$

$$\begin{aligned}
 & + \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell^\perp \delta U_i^\ell \rangle \\
 & = \sum_{i'=1}^{r_\ell} \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot U_{i'}^\ell \rangle \langle U_{i'}^\ell, \delta U_i^\ell \rangle \\
 & + \langle F(Y), \mathcal{S}(Y, \ell, i) \cdot \mathcal{P}_\ell^\perp \delta U_i^\ell \rangle \tag{3.20}
 \end{aligned}$$

holds. Since (3.19) is equal to (3.20), it follows that

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0 \quad \text{for all } \delta Y = \mathcal{S}(Y, \ell, i) \cdot \delta U_i^\ell.$$

The superposition of the variations (3.14), (3.15) and (3.18) span the whole tangent space and therefore

$$\langle F(Y) - \dot{Y}, \delta Y \rangle = 0 \quad \text{for all variations } \delta Y \in \mathcal{T}_Y \mathcal{M}. \quad \square$$

**Corollary 5** *Let  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$  be an arbitrary tensor, let  $Y \in \mathcal{M}$  and let  $P(Y)B$  be the orthogonal projection of  $B$  onto the tangent space  $\mathcal{T}_Y \mathcal{M}$  defined by*

$$\langle P(Y)B - B, \delta Y \rangle = 0 \quad \text{for all } \delta Y \in \mathcal{T}_Y \mathcal{M}.$$

*Then, formally replacing  $\dot{Y}$  by  $P(Y)B$  and  $F(Y)$  by  $B$  in Theorem 3 yields a recursive representation of  $P(Y)B$ .*

### 4 Matrix representation of tensors

Since the solution  $Y_{ex}$  of the original differential equation (3.1) does, in general, not evolve on the approximation manifold  $\mathcal{M} = \mathcal{H}\text{-Tucker}(r_{\mathcal{N}})_{\mathcal{N} \in T}$ , the solution  $Y$  of the projected problem only yields an approximation  $Y \approx Y_{ex}$ . For the error analysis of this approximation, the **matricization** of a tensor is a useful concept. Such a matricization turns a tensor into a matrix with the same entries, just as the command `reshape(...)` in MATLAB. For any  $\mathcal{N} \in T$  the set  $\mathcal{N}_C := \{1, \dots, d\} \setminus \mathcal{N}$  is called the **complement node** of  $\mathcal{N}$ . Then, the matricization

$$Y^{(\mathcal{N}, \mathcal{N}_C)} \in \mathbb{R}^{\pi(\mathcal{N}) \times \pi(\mathcal{N}_C)}, \quad \pi(\mathcal{N}) = \prod_{k \in \mathcal{N}} N_k, \quad \pi(\mathcal{N}_C) = \prod_{k \in \mathcal{N}_C} N_k$$

of a tensor  $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$  rearranges the elements of the tensor in such a way that all the different directions in  $\mathcal{N}$  range over the rows whereas all the directions in  $\mathcal{N}_C$  range over the columns of the matricization. There are many different ways to do this, but it does not really matter which mapping is used. The only condition we impose is that  $Y^{(\mathcal{N}, \mathcal{N}_C)} = (Y^{(\mathcal{N}_C, \mathcal{N})})^T$ . For the inverse mapping the notation

$$Y = \left( Y^{(\mathcal{N}, \mathcal{N}_C)} \right)_{(\mathcal{N}, \mathcal{N}_C)}$$

will be used. A special case of a matricization is the **vectorization**

$$Y^{(\mathcal{N}_0, \{\})} \in \mathbb{R}^{N_1 \dots N_d}$$

which reshapes a tensor to a vector. The corresponding inverse mapping is denoted by

$$Y = (Y^{(\mathcal{N}_0, \emptyset)})_{(\mathcal{N}_0, \emptyset)}.$$

For  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  and  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$ , let the **single-hole matrix**

$$\mathbb{S}^{\mathcal{N}} := ((\mathcal{S}(Y, \mathcal{N}, 1))^{(\mathcal{N}_C, \emptyset)} \mid \dots \mid (\mathcal{S}(Y, \mathcal{N}, r_{\mathcal{N}}))^{(\mathcal{N}_C, \emptyset)}) \in \mathbb{R}^{\pi(\mathcal{N}_C) \times r_{\mathcal{N}}}$$

be the matrix whose  $j$ -th column is the vectorization of  $\mathcal{S}(Y, \mathcal{N}, j)$ . Similarly, the **orthonormal basis matrix** is defined by

$$\mathbb{U}^{\mathcal{N}} := ((U_1^{\mathcal{N}})^{(\mathcal{N}, \emptyset)} \mid \dots \mid (U_{r_{\mathcal{N}}}^{\mathcal{N}})^{(\mathcal{N}, \emptyset)}) \in \mathbb{R}^{\pi(\mathcal{N}) \times r_{\mathcal{N}}}.$$

Assumptions 1 implies that  $(\mathbb{U}^{\mathcal{N}})^T \mathbb{U}^{\mathcal{N}} = I$ , and it follows from (2.7) that

$$Y^{(\mathcal{N}, \mathcal{N}_C)} = \mathbb{U}^{\mathcal{N}} (\mathbb{S}^{\mathcal{N}})^T, \quad \mathbb{S}^{\mathcal{N}} = (Y^{(\mathcal{N}, \mathcal{N}_C)})^T \mathbb{U}^{\mathcal{N}}.$$

Hence, the singular values of the matricization  $Y^{(\mathcal{N}, \mathcal{N}_C)}$  coincide with the singular values of the single-hole matrix  $\mathbb{S}^{\mathcal{N}}$ . The projection onto the space spanned by the columns of  $\mathbb{U}^{\mathcal{N}}$  and on the orthogonal complement will be denoted by

$$P^{\mathcal{N}} := \mathbb{U}^{\mathcal{N}} (\mathbb{U}^{\mathcal{N}})^T \quad \text{and} \quad (P^{\mathcal{N}})^{\perp} := I - \mathbb{U}^{\mathcal{N}} (\mathbb{U}^{\mathcal{N}})^T,$$

respectively. In addition we define for  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$  the binary matrix operator

$$\begin{aligned} \mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2} &\rightarrow \mathbb{R}^{\pi(\mathcal{N}) \times (r_{\mathcal{N}_1} r_{\mathcal{N}_2})} \\ \mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2} &= (v_{11} | v_{12} | \dots | v_{1r_{\mathcal{N}_2}} | v_{21} | \dots | v_{2r_{\mathcal{N}_2}} | \dots | \dots | v_{r_{\mathcal{N}_1} r_{\mathcal{N}_2}}) \\ v_{j k} &= (U_j^{\mathcal{N}_1} \cdot U_k^{\mathcal{N}_2})^{(\mathcal{N}, \emptyset)}. \end{aligned}$$

After these preparations, the results of [14] can now be adapted to show that  $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  is indeed a manifold.

**Lemma 6** *The set*

$$\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) = \{X \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{rank}(X^{(\mathcal{N}, \mathcal{N}_C)}) = r_{\mathcal{N}} \forall \mathcal{N} \in T \setminus \{\mathcal{N}_0\}\}$$

is a smooth submanifold of  $\mathbb{R}^{N_1 \times \dots \times N_d}$ .

*Proof* In [14, Theorem 4.11] it has been shown that

$$\widehat{\mathcal{M}} := \{X \in \mathbb{R}^{N_1 \times \dots \times N_d} : \text{rank}(X^{(\mathcal{N}, \mathcal{N}_C)}) = r_{\mathcal{N}} \text{ for all } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}\}$$

is a smooth submanifold of  $\mathbb{R}^{N_1 \times \dots \times N_d}$ . Hence, it only has to be shown that  $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) = \widehat{\mathcal{M}}$ . Let  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be represented by

$(\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)}$ . By Assumption 2, the sequence  $\langle Y, U_1^{\mathcal{N}}, \dots, \langle Y, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$  is linearly independent for all  $T \setminus \{\mathcal{N}_0\}$ , which means that

$$\text{rank}(\mathbb{S}^{\mathcal{N}}) = \text{rank}(Y^{\mathcal{N}, \mathcal{N}^c}) = r_{\mathcal{N}} \quad \text{for all } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}$$

and therefore  $\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \subseteq \widehat{\mathcal{M}}$ . Conversely, for every  $\widehat{Y} \in \widehat{\mathcal{M}}$  Proposition 3.6 in [14] and orthogonalization with Algorithm 3 in [2] ensures the existence of  $((\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (U^\ell)_{\ell \in \mathcal{L}(T)})$  such that the equation

$$\widehat{Y} = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \quad (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N}_0)$$

and the recursion

$$U_i^{\mathcal{N}} = \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}$$

hold for all  $\mathcal{N} \in \mathcal{I}(T), i \in \{1, \dots, r_{\mathcal{N}}\}$  and  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ . The rank of the single-hole matrix  $\mathbb{S}^{\mathcal{N}} = (X^{\mathcal{N}, \mathcal{N}^c})^T \mathbb{U}^{\mathcal{N}}$  equals  $r_{\mathcal{N}}$  for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ , which means that the sequence  $\langle X, U_1^{\mathcal{N}}, \dots, \langle X, U_{r_{\mathcal{N}}}^{\mathcal{N}} \rangle$  is linearly independent for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and therefore  $\widehat{\mathcal{M}} \subseteq \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$ .  $\square$

One of the main ingredients for the error analysis is a compact representation of the projection of a tensor on the tangent space of the manifold, i.e. an explicit representation which is not based on a recursion. Such a compact representation can be obtained in matrix-vector notation. As a preparatory step, we reformulate the results of Corollary 5 in matrix-vector notation.

**Corollary 7** *Let  $Y \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be a tensor in the hierarchical Tucker format represented by the transfer tensors  $(\mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}$  and by the basis functions  $(U^\ell)_{\ell \in \mathcal{L}(T)}$  on the leaves of the dimension tree. If  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$  is an arbitrary tensor, then under the conditions of Corollary 5 there exists a tuple  $((\delta \mathcal{B}^{\mathcal{N}})_{\mathcal{N} \in \mathcal{I}(T)}, (\delta U^\ell)_{\ell \in \mathcal{L}(T)})$  such that for  $(\mathcal{N}_{01}, \mathcal{N}_{02}) = \text{succ}(\mathcal{N}_0)$*

$$\begin{aligned} P(Y)B &= \sum_{j=1}^{r_{\mathcal{N}_{01}}} \sum_{l=1}^{r_{\mathcal{N}_{02}}} \delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_{01}} \cdot U_l^{\mathcal{N}_{02}} \\ &+ \sum_{j=1}^{r_{\mathcal{N}_{01}}} \sum_{l=1}^{r_{\mathcal{N}_{02}}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} \delta U_j^{\mathcal{N}_{01}} \cdot U_l^{\mathcal{N}_{02}} + \sum_{j=1}^{r_{\mathcal{N}_{01}}} \sum_{l=1}^{r_{\mathcal{N}_{02}}} \mathcal{B}_{1,j,l}^{\mathcal{N}_0} U_j^{\mathcal{N}_{01}} \cdot \delta U_l^{\mathcal{N}_{02}} \end{aligned}$$

with

$$\delta \mathcal{B}_{1,j,l}^{\mathcal{N}_0} = ((U_j^{\mathcal{N}_{01}} \cdot U_l^{\mathcal{N}_{02}})^{(\mathcal{N}_0, \{\})})^T B^{(\mathcal{N}_0, \{\})}$$

For all  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$  with  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ ,  $\delta U^{\mathcal{N}}$  is defined by the recursion

$$\begin{aligned} \delta U_i^{\mathcal{N}} &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \\ &+ \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} (\delta U^{\mathcal{N}_1})_j \cdot U_l^{\mathcal{N}_2} + \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} \mathcal{B}_{i,j,l}^{\mathcal{N}} U_j^{\mathcal{N}_1} \cdot \delta U_l^{\mathcal{N}_2} \end{aligned}$$

with

$$\begin{aligned} (\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \cdots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) &= ((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \emptyset)})^T (P^{\mathcal{N}})^\perp B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}^+})^T \\ \delta \mathbb{U}^\ell &= (P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^{\ell^+})^T. \end{aligned}$$

Here and below,  $M^+$  is the pseudoinverse of a matrix  $M$ , i.e.  $M^+ = (M^T M)^{-1} M^T$  if  $M$  has full column rank.

*Proof* The equation for  $\delta \mathcal{B}^{\mathcal{N}_0}$  can be deduced directly from Corollary 5. For  $\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\}$  and  $(\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})$ , Corollary 5 states that

$$\begin{aligned} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} &= \sum_{i'=1}^{r_{\mathcal{N}}} \langle B, \mathcal{S}(Y, \mathcal{N}, i') \cdot \mathcal{P}_{\mathcal{N}}^\perp(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}) \rangle w_{i,i'}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \langle \mathcal{P}_{\mathcal{N}}^\perp(U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}), \langle B, \mathcal{S}(Y, \mathcal{N}, i') \rangle \rangle w_{i,i'}^{\mathcal{N}} \\ &= \sum_{i'=1}^{r_{\mathcal{N}}} \langle U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2}, \mathcal{P}_{\mathcal{N}}^\perp \langle B, \mathcal{S}(Y, \mathcal{N}, i') \rangle \rangle w_{i,i'}^{\mathcal{N}} \end{aligned}$$

for all  $j \in \{1, \dots, r_{\mathcal{N}_1}\}, l \in \{1, \dots, r_{\mathcal{N}_2}\}$ . Since  $W^{\mathcal{N}}$  defined in (3.8) satisfies

$$(\mathbb{S}^{\mathcal{N}}) W^{\mathcal{N}} = (\mathbb{S}^{\mathcal{N}}) ((\mathbb{S}^{\mathcal{N}})^T \mathbb{S}^{\mathcal{N}})^{-1} = (\mathbb{S}^{\mathcal{N}^+})^T \tag{4.1}$$

we obtain

$$(\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \cdots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) = ((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \emptyset)})^T (P^{\mathcal{N}})^\perp B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}^+})^T.$$

Last, for  $\ell \in \mathcal{L}(T)$  and  $i = 1, \dots, r_\ell$  we recall from Corollary 5 that

$$\delta U_i^\ell = \sum_{i'=1}^{r_\ell} \mathcal{P}_\ell^\perp \langle B, \mathcal{S}(Y, \ell, i') \rangle w_{i',i}^\ell,$$

and substituting (4.1) with  $\mathcal{N} = \ell$  yields

$$\delta \mathbb{U}^\ell = (P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^{\ell^+})^T. \tag{□}$$

With Corollary 7 we are now able to derive a compact representation for the projection of a tensor on the tangent space of the manifold.

**Corollary 8** *Under the assumptions of Corollary 7, the projection of a tensor  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$  on the tangent space  $\mathcal{T}_Y \mathcal{M}$  can be written as*

$$\begin{aligned}
 P(Y)B &= ((\mathbb{U}^{\mathcal{N}_{01}} \odot \mathbb{U}^{\mathcal{N}_{02}})(\mathbb{U}^{\mathcal{N}_{01}} \odot \mathbb{U}^{\mathcal{N}_{02}})^T B^{(\mathcal{N}_0, \{\})})_{(\mathcal{N}_0, \{\})} \\
 &+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} ((\mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2})(\mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2})^T \\
 &\times (P^{\mathcal{N}})^\perp B^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}+})^T)_{(\mathcal{N}, \mathcal{N}_C)} \\
 &+ \sum_{\ell \in \mathcal{L}(T)} ((P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^\ell \mathbb{S}^{\ell+})^T)_{(\ell, \ell_C)},
 \end{aligned}$$

where  $(\mathcal{N}_{01}, \mathcal{N}_{02}) = \text{succ}(\mathcal{N}_0)$ .

*Proof* By resolving the recursion in Corollary 5 and 7, respectively, we obtain the equation

$$P(Y)B = \sum_{j=1}^{r_{\mathcal{N}_{01}}} \sum_{l=1}^{r_{\mathcal{N}_{02}}} U_j^{\mathcal{N}_{01}} \cdot U_l^{\mathcal{N}_{02}} \delta B_{1,j,l}^{\mathcal{N}_0} \tag{4.2}$$

$$+ \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot \sum_{i=1}^{r_{\mathcal{N}}} \delta B_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \tag{4.3}$$

$$+ \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{\ell} \delta U_i^\ell \cdot \mathcal{S}(Y, \ell, i), \tag{4.4}$$

see also Eqs. (3.3), (3.4), (3.5) and (3.6) in Sect. 3. A closer inspection of the first term (4.2) shows that

$$\begin{aligned}
 &\sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \delta B_{1,j,l}^{\mathcal{N}_0} \\
 &= \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} ((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}_0, \{\})})^T B^{(\mathcal{N}_0, \{\})} \\
 &= \left( \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} (U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}_0, \{\})} ((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}_0, \{\})})^T B^{(\mathcal{N}_0, \{\})} \right)_{(\mathcal{N}_0, \{\})}
 \end{aligned}$$

$$= ((\mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2})(\mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2})^T \mathbf{B}^{(\mathcal{N}_0, \{\})})_{(\mathcal{N}_0, \{\})}.$$

For the second term (4.3), we obtain

$$\begin{aligned} & \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \\ &= \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left( \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} (U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})} \right. \\ & \quad \left. \times \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} (\mathcal{S}(Y, \mathcal{N}, i)^{(\mathcal{N}_C, \{\})})^T \right)_{(\mathcal{N}, \mathcal{N}_C)} \\ &= \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left( \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} ((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})}) \right. \\ & \quad \left. \times (\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \dots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) (\mathbb{S}^{\mathcal{N}})^T \right)_{(\mathcal{N}, \mathcal{N}_C)} \end{aligned}$$

and together with

$$(\delta \mathcal{B}_{1,j,l}^{\mathcal{N}} \dots \delta \mathcal{B}_{r_{\mathcal{N}},j,l}^{\mathcal{N}}) = ((U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2})^{(\mathcal{N}, \{\})})^T (\mathbf{P}^{\mathcal{N}})^\perp \mathbf{B}^{(\mathcal{N}, \mathcal{N}_C)} (\mathbb{S}^{\mathcal{N}+})^T$$

we obtain

$$\begin{aligned} & \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \sum_{j=1}^{r_{\mathcal{N}_1}} \sum_{l=1}^{r_{\mathcal{N}_2}} U_j^{\mathcal{N}_1} \cdot U_l^{\mathcal{N}_2} \cdot \sum_{i=1}^{r_{\mathcal{N}}} \delta \mathcal{B}_{i,j,l}^{\mathcal{N}} \mathcal{S}(Y, \mathcal{N}, i) \\ &= \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} ((\mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2})(\mathbb{U}^{\mathcal{N}_1} \odot \mathbb{U}^{\mathcal{N}_2})^T (\mathbf{P}^{\mathcal{N}})^\perp \mathbf{B}^{(\mathcal{N}, \mathcal{N}_C)} \\ & \quad \times (\mathbb{S}^{\mathcal{N}} \mathbb{S}^{\mathcal{N}+})^T)_{(\mathcal{N}, \mathcal{N}_C)}. \end{aligned}$$

Finally, the last term (4.4) can be reformulated as

$$\begin{aligned} \sum_{\ell \in \mathcal{L}(T)} \sum_{i=1}^{r_\ell} \delta U_i^\ell \cdot \mathcal{S}(Y, \ell, i) &= \sum_{\ell \in \mathcal{L}(T)} \left( \sum_{i=1}^{r_\ell} (\delta U_i^\ell)^{(\ell, \{\})} (\mathcal{S}(Y, \ell, i)^{(\ell_C, \{\})})^T \right)_{(\ell, \ell_C)} \\ &= \sum_{\ell \in \mathcal{L}(T)} (\delta \mathbb{U}^\ell (\mathbb{S}^\ell)^T)_{(\ell, \ell_C)} \end{aligned}$$

$$\begin{aligned}
 &= \sum_{\ell \in \mathcal{L}(T)} ((P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^{\ell^+})^T (\mathbb{S}^\ell)^T)_{(\ell, \ell_C)} \\
 &= \sum_{\ell \in \mathcal{L}(T)} ((P^\ell)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^\ell \mathbb{S}^{\ell^+})^T)_{(\ell, \ell_C)}.
 \end{aligned}$$

This proves the assertion. □

### 5 Curvature bounds

In this section estimates for the curvature of the manifold are proven. These estimates, formulated in Lemma 10 below, will play a crucial role in the error analysis of the variational approximation in Sect. 6. Similar results have been obtained for the dynamical low-rank approximation of matrices or tensors, respectively, in [7, 8]. Our general strategy follows the ideas developed there, but extending the results of [7, 8] to the hierarchical Tucker format is a considerable challenge due to the recursive definition of the equations of motion.

In the following  $\sigma_i(M)$  denotes the  $i$ -th singular value (in descending order) of a matrix  $M \in \mathbb{R}^{n \times m}$  with  $n, m \geq i$ . Moreover, we will use the Frobenius norm  $\|Y\|_F = \sqrt{\langle Y, Y \rangle}$  of tensors  $Y \in \mathbb{R}^{N_1 \times \dots \times N_d}$ . By construction the Frobenius norm of a tensor equals the matrix Frobenius norm of a corresponding matricization, i.e.  $\|Y\|_F = \|Y^{(\mathcal{N}, \mathcal{N}_C)}\|_F$  for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ .

**Lemma 9** *If  $Y, \tilde{Y} \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  are tensors in the hierarchical Tucker format fulfilling*

$$\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho > 0 \quad \text{and} \quad \|Y - \tilde{Y}\|_F = \delta$$

for some  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ , then

$$\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}) \geq \rho - \delta.$$

If in addition  $\rho > \delta$ , then the pseudoinverse  $\tilde{\mathbb{S}}^{\mathcal{N}+}$  of the single-hole-matrix  $\tilde{\mathbb{S}}^{\mathcal{N}}$  of  $\tilde{Y}$  is bounded by

$$\|\tilde{\mathbb{S}}^{\mathcal{N}+}\|_2 \leq \frac{1}{\rho - \delta}.$$

*Proof* The estimates

$$\begin{aligned}
 \delta &= \|\tilde{Y} - Y\|_F = \|\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)} - Y^{(\mathcal{N}, \mathcal{N}_C)}\|_F \\
 &\geq \|\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)} - Y^{(\mathcal{N}, \mathcal{N}_C)}\|_2 \\
 &\geq |\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}) - \sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)})| \tag{5.1}
 \end{aligned}$$

hold for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ . The inequality (5.1) is shown with Theorem 7.4.51 in [6]. Then

$$\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}) \geq \sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)}) - |\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_C)}) - \sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_C)})| \geq \rho - \delta.$$



For  $\rho > \delta$  the pseudoinverse can be bounded by

$$\|\tilde{\mathbb{S}}^{\mathcal{N}^+}\|_2 = (\sigma_{r_{\mathcal{N}}}(\tilde{\mathbb{S}}^{\mathcal{N}}))^{-1} = (\sigma_{r_{\mathcal{N}}}(\tilde{Y}^{(\mathcal{N}, \mathcal{N}_c)}))^{-1} \leq (\rho - \delta)^{-1}. \quad \square$$

**Definition 5.1** An ordered tuple  $(Y, \tilde{Y}) \in (\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}))^2$  is called **path-connected** with respect to ODE (5.2), if there exists a solution for the boundary value problem

$$\begin{aligned} \dot{X}(\tau) &= \sum_{k=0}^{\infty} (P(X(\tau)) - P(Y))^k P(Y)(\tilde{Y} - Y) \\ X(0) &= Y, \quad X(1) = \tilde{Y} \end{aligned} \tag{5.2}$$

on the interval  $[0, 1]$ .

**Lemma 10** (curvature bounds) *Let  $(Y, \tilde{Y}) \in (\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}))^2$  tensors in the hierarchical Tucker format, which are path-connected with respect to ODE (5.2). Let  $(\rho_{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$  and  $(c_{\mathcal{N}})_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}}$  be families of positive real values such that*

$$\sigma_{r_{\mathcal{N}}}(Y^{(\mathcal{N}, \mathcal{N}_c)}) \geq \rho_{\mathcal{N}} > 0 \quad \text{and} \quad \delta := \|Y - \tilde{Y}\|_F \leq c_{\mathcal{N}} \rho_{\mathcal{N}} \tag{5.3}$$

for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$ . Furthermore let  $\tilde{c} > 0$  be a constant fulfilling the inequalities

$$\sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} 8c_{\mathcal{N}}(\tilde{c}(1 - \tilde{c} - c_{\mathcal{N}}))^{-1} \leq 1 \quad \text{and} \quad \tilde{c} + c_{\mathcal{N}} < 1$$

and define

$$C_{Y, \tilde{Y}} := \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} 8(\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}}))^{-1}.$$

Then the bounds

$$\|(P(Y) - P(\tilde{Y}))B\|_F \leq C_{Y, \tilde{Y}} \|Y - \tilde{Y}\|_F \|B\|_F, \tag{5.4}$$

$$\|P^\perp(Y)(\tilde{Y} - Y)\|_F \leq \frac{C_{Y, \tilde{Y}}}{1 - \tilde{c}} \|Y - \tilde{Y}\|_F^2 \tag{5.5}$$

hold for every tensor  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ .

A similar result has been proven independently in [12].

*Proof* Since the proof is rather long and technical, it is divided into several parts.

*Part I:* For the moment we assume that there are

$$X(\tau) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}) \quad \text{and} \quad Z(\tau) \perp \mathcal{T}_Y \mathcal{M}$$

such that

$$Y + \tau(\tilde{Y} - Y) = X(\tau) + Z(\tau) \tag{5.6}$$

for all  $0 \leq \tau \leq 1$ . Below we will **prove** that this assumption is indeed true. In the following  $\Delta$  denotes the projection of the difference from  $Y$  to  $\tilde{Y}$  on the tangent space  $\mathcal{T}_Y\mathcal{M}$ :

$$\Delta := P(Y)(\tilde{Y} - Y) \in \mathcal{T}_Y\mathcal{M}.$$

Obviously  $\|\Delta\|_F \leq \|\tilde{Y} - Y\|_F = \delta$  holds. The projection of (5.6) on the tangent space leads to the equation

$$P(Y)(X(\tau) - Y) = \tau \Delta, \tag{5.7}$$

and differentiating with respect to  $\tau$  gives

$$P(Y)(\dot{X}(\tau)) = \Delta.$$

The derivative  $\dot{X}(\tau)$  belongs to the tangent space of  $X(\tau)$ , therefore

$$P(X(\tau))\dot{X}(\tau) = \dot{X}(\tau)$$

is satisfied, which leads to

$$\begin{aligned} \dot{X}(\tau) &= P(X(\tau))\dot{X}(\tau) + \underbrace{\Delta - P(Y)\dot{X}(\tau)}_{=0} \\ &= \Delta + (P(X(\tau)) - P(Y))\dot{X}(\tau). \end{aligned} \tag{5.8}$$

In part 6 of the proof it will be shown that

$$\|P(X(\tau)) - P(Y)\|_F \leq \tilde{c} < 1 \tag{5.9}$$

for all  $0 \leq \tau \leq 1$ . Together with (5.8) this yields

$$\|\dot{X}(\tau)\|_F \leq \|\Delta\|_F + \|P(X(\tau)) - P(Y)\|_F \|\dot{X}(\tau)\|_F \leq \delta + \tilde{c} \|\dot{X}(\tau)\|_F$$

and hence

$$\|\dot{X}(\tau)\|_F \leq \delta(1 - \tilde{c})^{-1}. \tag{5.10}$$

Estimate (5.9) guarantees for all  $\tau \in [0, 1]$  the convergence of the Neumann series for  $P(X(\tau)) - P(Y)$ . Hence, the operator  $I - [P(X(\tau)) - P(Y)]$  is invertible. Defining the function

$$G(X(\tau)) := \sum_{k=0}^{\infty} (P(X(\tau)) - P(Y))^k \Delta = [I - (P(X(\tau)) - P(Y))]^{-1} \Delta$$

leads to the equations

$$\dot{X}(\tau) = \Delta + (P(X(\tau)) - P(Y))\dot{X}(\tau) = G(X(\tau)). \tag{5.11}$$

Then, the existence of a solution  $X(\tau)$  of (5.11), (5.8) respectively, with initial value  $X(0) = Y$  follows from the theory of ordinary differential equations and (5.10). Together with  $\tilde{c} < 1$  the estimates

$$\begin{aligned} & \|P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - P(Y)\Delta)\|_F \\ &= \|[I - P(X(\tau))]P(Y)(\dot{X}(\tau) - \Delta)\|_F \\ &= \|[P(Y) - P(X(\tau))](P(Y)\dot{X}(\tau) - \Delta)\|_F \\ &= \|[P(Y) - P(X(\tau))][P(X(\tau)) + P^\perp(X(\tau))](P(Y)\dot{X}(\tau) - \Delta)\|_F \\ &= \|[P(Y) - P(X(\tau))]P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - \Delta)\|_F \\ &\leq \|P(X(\tau)) - P(Y)\|_F \|P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - \Delta)\|_F \\ &\leq \tilde{c} \|P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - \Delta)\|_F \end{aligned}$$

show

$$P^\perp(X(\tau))(P(Y)\dot{X}(\tau) - \Delta) = 0.$$

Moreover the projection of (5.8) on the tangent space of  $X(\tau)$  leads to

$$P(X(\tau))(P(Y)\dot{X}(\tau) - \Delta) = 0$$

and hence

$$P(Y)\dot{X}(\tau) = \Delta. \tag{5.12}$$

Since integrating (5.12) leads to (5.7), setting

$$Z(\tau) = \tau P^\perp(Y)(\tilde{Y} - Y) - P^\perp(Y)X(\tau)$$

shows the existence of the decomposition stated above.

*Part 2:* In this part of the proof, we prove a number of auxiliary inequalities for later use. Since  $X(0) = Y$  by construction, the fundamental theorem of calculus and (5.10) reveal the estimate

$$\|X(\tau) - Y\|_F \leq \tau \sup_{s \in [0, \tau]} \|\dot{X}(s)\|_F \leq \delta(1 - \tilde{c})^{-1}$$

because  $\tau \leq 1$ . Let  $\mathbb{U}^{\mathcal{N}}(\tau)$  and  $\mathbb{S}^{\mathcal{N}}(\tau)$  be the orthonormal basis matrix and the single-hole-matrix of  $X(\tau)$ , i.e.  $(X(\tau))^{\langle \mathcal{N}, \mathcal{N} \rangle} = \mathbb{U}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^T$ . Under Assumption (5.3), Lemma 9 can be applied and provides the bound

$$\begin{aligned} \|\mathbb{S}^{\mathcal{N}}(\tau)^+\|_2 &\leq \frac{1}{\rho_{\mathcal{N}} - (1 - \tilde{c})^{-1}\delta} \leq \frac{1}{\rho_{\mathcal{N}} - (1 - \tilde{c})^{-1}c_{\mathcal{N}}\rho_{\mathcal{N}}} \\ &= \frac{1 - \tilde{c}}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \end{aligned} \tag{5.13}$$

for all  $0 \leq \tau \leq 1$  for the pseudoinverse  $\mathbb{S}^{\mathcal{N}}(\tau)$ . Next, a bound for  $\|\dot{\mathbb{U}}^{\mathcal{N}}\mathbb{S}^{\mathcal{N}}(\tau)^T\|_F$  is shown. Since  $P^{\mathcal{N}}(\tau)^\perp = I - \mathbb{U}^{\mathcal{N}}(\tau)\mathbb{U}^{\mathcal{N}}(\tau)^T$  by definition, it follows from

$$\mathbb{U}^{\mathcal{N}}(\tau)^T\mathbb{U}^{\mathcal{N}}(\tau) = I \quad \text{and} \quad \mathbb{U}^{\mathcal{N}}(\tau)^T\dot{\mathbb{U}}^{\mathcal{N}}(\tau) = 0 \tag{5.14}$$

that

$$\begin{aligned} \dot{\mathbb{U}}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^T &= P^{\mathcal{N}}(\tau)^\perp(\dot{\mathbb{U}}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^T + \mathbb{U}^{\mathcal{N}}(\tau)\dot{\mathbb{S}}^{\mathcal{N}}(\tau)^T) \\ &= P^{\mathcal{N}}(\tau)^\perp\dot{X}^{(\mathcal{N},\mathcal{N}_c)}(\tau). \end{aligned}$$

With  $\|P^{\mathcal{N}}(\tau)^\perp\|_2 \leq 1$  and (5.10), we thus obtain

$$\|\dot{\mathbb{U}}^{\mathcal{N}}\mathbb{S}^{\mathcal{N}}(\tau)^T\|_F \leq \|P^{\mathcal{N}}(\tau)^\perp\|_2 \cdot \|\dot{X}^{(\mathcal{N},\mathcal{N}_c)}(\tau)\|_F \leq \delta(1 - \tilde{c})^{-1}.$$

Combining this with (5.13) yields

$$\begin{aligned} \|\dot{\mathbb{U}}^{\mathcal{N}}(\tau)\|_F &\leq \|\dot{\mathbb{U}}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^T\|_F\|\mathbb{S}^{\mathcal{N}}(\tau)^+\|_2 \leq \delta(1 - \tilde{c})^{-1} \frac{1 - \tilde{c}}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \\ &= \frac{\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}. \end{aligned}$$

With this inequality, the derivative of  $P^{\mathcal{N}}(\tau)^\perp = I - \mathbb{U}^{\mathcal{N}}(\tau)\mathbb{U}^{\mathcal{N}}(\tau)^T$  can be bounded by

$$\|\dot{P}^{\mathcal{N}}(\tau)^\perp\|_F \leq \frac{2\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}.$$

A bound for the derivative of the single-hole-matrix  $\mathbb{S}^{\mathcal{N}}(\tau)$  of  $X(\tau)$  can be derived from the orthonormality (5.14) and (5.10) via

$$\begin{aligned} \|\dot{\mathbb{S}}^{\mathcal{N}}(\tau)^T\|_F &= \|\mathbb{U}^{\mathcal{N}}(\tau)^T\dot{\mathbb{U}}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^T + \mathbb{U}^{\mathcal{N}}(\tau)^T\mathbb{U}^{\mathcal{N}}(\tau)\dot{\mathbb{S}}^{\mathcal{N}}(\tau)^T\|_F \\ &= \|\mathbb{U}^{\mathcal{N}}(\tau)^T\dot{X}^{(\mathcal{N},\mathcal{N}_c)}(\tau)\|_F \\ &\leq \delta(1 - \tilde{c})^{-1}. \end{aligned} \tag{5.15}$$

Next, we want to estimate the derivative of  $\mathbb{S}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^+$ , which can be written as

$$\begin{aligned} \frac{d}{d\tau}[\mathbb{S}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^+] &= \frac{d}{d\tau}[\mathbb{S}^{\mathcal{N}}(\tau)(\mathbb{S}^{\mathcal{N}}(\tau)^T\mathbb{S}^{\mathcal{N}}(\tau))^{-1}\mathbb{S}^{\mathcal{N}}(\tau)^T] \\ &= T_1(\tau) + T_2(\tau) + T_1^T(\tau) \end{aligned}$$

with

$$\begin{aligned} T_1(\tau) &= \dot{\mathbb{S}}^{\mathcal{N}}(\tau)\mathbb{S}^{\mathcal{N}}(\tau)^+ \\ T_2(\tau) &= \mathbb{S}^{\mathcal{N}}(\tau)\frac{d}{d\tau}[(\mathbb{S}^{\mathcal{N}}(\tau)^T\mathbb{S}^{\mathcal{N}}(\tau))^{-1}]\mathbb{S}^{\mathcal{N}}(\tau)^T. \end{aligned}$$

It follows from (5.13) and (5.15) that

$$\begin{aligned} \|T_1\|_F &\leq \|\mathbb{S}^{\mathcal{N}}(\tau)^+\|_2 \cdot \|\dot{\mathbb{S}}^{\mathcal{N}}(\tau)\|_F \leq \frac{1 - \tilde{c}}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \delta(1 - \tilde{c})^{-1} \\ &= \frac{\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}. \end{aligned}$$

In order to bound  $T_2$  we substitute

$$\begin{aligned} &\frac{d}{d\tau} [(\mathbb{S}^{\mathcal{N}}(\tau)^T \mathbb{S}^{\mathcal{N}}(\tau))^{-1}] \\ &= -(\mathbb{S}^{\mathcal{N}}(\tau)^T \mathbb{S}^{\mathcal{N}}(\tau))^{-1} \frac{d}{d\tau} [\mathbb{S}^{\mathcal{N}}(\tau)^T \mathbb{S}^{\mathcal{N}}(\tau)] (\mathbb{S}^{\mathcal{N}}(\tau)^T \mathbb{S}^{\mathcal{N}}(\tau))^{-1} \end{aligned}$$

and obtain

$$\begin{aligned} T_2(\tau) &= -(\mathbb{S}^{\mathcal{N}}(\tau)^+)^T \frac{d}{d\tau} [\mathbb{S}^{\mathcal{N}}(\tau)^T \mathbb{S}^{\mathcal{N}}(\tau)] \mathbb{S}^{\mathcal{N}}(\tau)^+ \\ &= -T_1^T(\tau) \mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+ - (\mathbb{S}^{\mathcal{N}}(\tau)^+)^T \mathbb{S}^{\mathcal{N}}(\tau)^T T_1(\tau). \end{aligned}$$

Since  $\|\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+\|_2 = 1$ , this gives the bound  $\|T_2(\tau)\|_F \leq 2\|T_1(\tau)\|_F$  and in total

$$\left\| \frac{d}{d\tau} [(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T] \right\|_F \leq 4\|T_1(\tau)\|_F \leq \frac{4\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})}.$$

*Part 3:* Since  $(Y, \tilde{Y}) \in (\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}))^2$  is path-connected with respect to ODE (5.2), we have  $X(1) = \tilde{Y}$ . As a first step towards an error bound for  $\|(P(\tilde{Y}) - P(Y))B\|_F$  we use

$$[P(\tilde{Y}) - P(Y)]B = [P(X(1)) - P(X(0))]B = \int_0^1 \frac{d}{d\tau} [P(X(\tau))B] d\tau \tag{5.16}$$

and replace  $P(X(\tau))B$  by the representation from Corollary 8. This yields

$$\begin{aligned} &\|(P(\tilde{Y}) - P(Y))B\|_F \\ &\leq \max_{\tau \in [0,1]} \left\| \frac{d}{d\tau} [(\mathbb{U}^{\mathcal{N}_{01}}(\tau) \odot \mathbb{U}^{\mathcal{N}_{02}}(\tau)) (\mathbb{U}^{\mathcal{N}_{01}}(\tau) \odot \mathbb{U}^{\mathcal{N}_{02}}(\tau))^T B^{(\mathcal{N}_0, \cdot)}] \right\|_F \tag{5.17} \\ &\quad + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \max_{\tau \in [0,1]} \left\| \frac{d}{d\tau} [\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) P^{\mathcal{N}}(\tau)^\perp B^{(\mathcal{N}, \mathcal{N}_C)} \right. \\ &\quad \left. \times (\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}^+}(\tau))^T] \right\|_F \tag{5.18} \end{aligned}$$

$$+ \sum_{\ell \in \mathcal{L}(T)} \max_{\tau \in [0,1]} \left\| \frac{d}{d\tau} [P^\ell(\tau)^\perp B^{(\ell, \ell_C)} (\mathbb{S}^\ell(\tau) \mathbb{S}^{\ell^+}(\tau))^T] \right\|_F, \tag{5.19}$$

where

$$\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) := (\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau))(\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau))^T$$

and  $(\mathcal{N}_{01}, \mathcal{N}_{02}) = succ(\mathcal{N}_0)$ .

*Part 4:* Now for each of the three terms (5.17), (5.18), (5.19), a suitable bound has to be derived. Assuming  $\mathcal{N}$  to be an interior node with successors  $(\mathcal{N}_1, \mathcal{N}_2) = succ(\mathcal{N})$  allows us to bound

$$\begin{aligned} \left\| \frac{d}{d\tau} \mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) \right\|_2 &= \left\| \frac{d}{d\tau} [(\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau))(\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau))^T] \right\|_2 \\ &\leq 2 \left\| \frac{d}{d\tau} [\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau)] \right\|_2 \\ &\leq 2 \|\dot{\mathbb{U}}^{\mathcal{N}_1}(\tau) \odot \mathbb{U}^{\mathcal{N}_2}(\tau)\|_2 + 2 \|\mathbb{U}^{\mathcal{N}_1}(\tau) \odot \dot{\mathbb{U}}^{\mathcal{N}_2}(\tau)\|_2 \\ &\leq 2 \|\dot{\mathbb{U}}^{\mathcal{N}_1}(\tau)\|_2 + 2 \|\dot{\mathbb{U}}^{\mathcal{N}_2}(\tau)\|_2 \\ &\leq \frac{2\delta}{\rho_{\mathcal{N}_1}(1 - \tilde{c} - c_{\mathcal{N}_1})} + \frac{2\delta}{\rho_{\mathcal{N}_2}(1 - \tilde{c} - c_{\mathcal{N}_2})}. \end{aligned}$$

Hence, the inequality

$$\begin{aligned} &\left\| \frac{d}{d\tau} [(\mathbb{U}^{\mathcal{N}_{01}}(\tau) \odot \mathbb{U}^{\mathcal{N}_{02}}(\tau))(\mathbb{U}^{\mathcal{N}_{01}}(\tau) \odot \mathbb{U}^{\mathcal{N}_{02}}(\tau))^T \mathbf{B}^{(\mathcal{N}_0, \emptyset)}] \right\|_F \\ &\leq \frac{2\delta}{\rho_{\mathcal{N}_{01}}(1 - \tilde{c} - c_{\mathcal{N}_{01}})} \|\mathbf{B}\|_F + \frac{2\delta}{\rho_{\mathcal{N}_{02}}(1 - \tilde{c} - c_{\mathcal{N}_{02}})} \|\mathbf{B}\|_F \end{aligned} \tag{5.20}$$

holds for all tensors  $\mathbf{B} \in \mathbb{R}^{\mathcal{N}_1 \times \dots \times \mathcal{N}_d}$ .

Next, we consider (5.18). Assuming  $\mathcal{N} \neq \mathcal{N}_0$  allows us to deduce the estimate

$$\begin{aligned} &\left\| \frac{d}{d\tau} [\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau) P^{\mathcal{N}}(\tau)^\perp \mathbf{B}^{(\mathcal{N}, \mathcal{N}_c)}(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T] \right\|_F \\ &\leq \left\| \frac{d}{d\tau} [\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau)] \right\|_2 \|\mathbf{B}\|_F \|(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T\|_2 \\ &\quad + \|\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau)\|_2 \left\| \frac{d}{d\tau} [P^{\mathcal{N}}(\tau)^\perp] \right\|_F \|\mathbf{B}\|_F \|(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T\|_2 \\ &\quad + \|\mathbb{W}^{(\mathcal{N}_1, \mathcal{N}_2)}(\tau)\|_2 \|\mathbf{B}\|_F \left\| \frac{d}{d\tau} [(\mathbb{S}^{\mathcal{N}}(\tau) \mathbb{S}^{\mathcal{N}}(\tau)^+)^T] \right\|_2 \\ &\leq \left( \frac{2\delta}{\rho_{\mathcal{N}_1}(1 - \tilde{c} - c_{\mathcal{N}_1})} + \frac{2\delta}{\rho_{\mathcal{N}_2}(1 - \tilde{c} - c_{\mathcal{N}_2})} + \frac{6\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \right) \|\mathbf{B}\|_F. \end{aligned} \tag{5.21}$$

Finally, let  $\ell$  be a leaf of the dimension tree. Then, for the term in (5.19) we obtain the estimate

$$\begin{aligned} & \left\| \frac{d}{d\tau} [P^\ell(\tau)^\perp B^{(\ell, \ell c)} (\mathbb{S}^\ell(\tau) \mathbb{S}^\ell(\tau)^+)] \right\|_F \\ & \leq \left\| \frac{d}{d\tau} [P^\ell(\tau)^\perp] \right\|_F \|B\|_F \|(\mathbb{S}^\ell(\tau) \mathbb{S}^\ell(\tau)^+)^T\|_2 \\ & \quad + \|B\|_F \left\| \frac{d}{d\tau} [(\mathbb{S}^\ell(\tau) \mathbb{S}^\ell(\tau)^+)^T] \right\|_2 \\ & \leq \frac{6\delta}{\rho_\ell(1 - \tilde{c} - c_\ell)} \|B\|_F. \end{aligned} \tag{5.22}$$

Part 5: Substituting (5.20), (5.21), (5.22) into (5.17), (5.18), (5.19) yields the first assertion (5.4):

$$\begin{aligned} & \|(P(\tilde{Y}) - P(Y))B\|_F \\ & \leq \frac{2\delta}{\rho_{\mathcal{N}_{01}}(1 - \tilde{c} - c_{\mathcal{N}_{01}})} \|B\|_F + \frac{2\delta}{\rho_{\mathcal{N}_{02}}(1 - \tilde{c} - c_{\mathcal{N}_{02}})} \|B\|_F \\ & \quad + \sum_{\substack{\mathcal{N} \in \mathcal{I}(T) \setminus \{\mathcal{N}_0\} \\ (\mathcal{N}_1, \mathcal{N}_2) = \text{succ}(\mathcal{N})}} \left( \frac{2\delta}{\rho_{\mathcal{N}_1}(1 - \tilde{c} - c_{\mathcal{N}_1})} + \frac{2\delta}{\rho_{\mathcal{N}_2}(1 - \tilde{c} - c_{\mathcal{N}_2})} \right. \\ & \quad \left. + \frac{6\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \right) \|B\|_F \\ & \quad + \sum_{\ell \in \mathcal{L}(T)} \frac{6\delta}{\rho_\ell(1 - \tilde{c} - c_\ell)} \|B\|_F \\ & = \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \|B\|_F, \end{aligned}$$

where  $(\mathcal{N}_{01}, \mathcal{N}_{02}) = \text{succ}(\mathcal{N}_0)$ . The proof of the second assertion (5.5) is much shorter. With  $P(X(\tau))\dot{X}(\tau) = \dot{X}(\tau)$  we obtain

$$P^\perp(Y)(\tilde{Y} - Y) = \int_0^1 (I - P(Y))\dot{X}(\tau) d\tau = \int_0^1 (P(X(\tau)) - P(Y))\dot{X}(\tau) d\tau,$$

and with (5.4) and (5.10) the error bound

$$\begin{aligned} \|P^\perp(Y)(\tilde{Y} - Y)\|_F & \leq \max_{\tau \in [0,1]} \|P(X(\tau)) - P(Y)\|_F \cdot \|\dot{X}(\tau)\|_F \\ & \leq \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \delta(1 - \tilde{c})^{-1} \end{aligned}$$

$$= \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta^2}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})(1 - \tilde{c})}$$

follows.

*Part 6:* Finally, we can prove inequality (5.9) which had been used in part 1 of the proof. Assume (5.9) does not hold. Since  $\mathcal{M}$  is a smooth manifold, there exist  $0 < \tau^* < 1$  and a tensor  $B^* \in \mathbb{R}^{N_1 \times \dots \times N_d}$  such that

$$\begin{aligned} \|P(X(s)) - P(Y)\|_F &\leq \tilde{c} < 1 \quad \text{and} \\ \|(P(X(\tau^*)) - P(Y))B^*\|_F &= \tilde{c}\|B^*\|_F \end{aligned} \tag{5.23}$$

hold for all  $s \in [0, \tau^*]$ . Under these conditions all the estimates of the proof remain valid for  $\tau \in [0, \tau^*]$ . However, in (5.16) we take the integral over the interval  $[0, \tau^*]$  instead of  $[0, 1]$  and obtain with similar arguments the estimate

$$\|(P(X(\tau^*)) - P(Y))B\|_F \leq \tau^* \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \|B\|_F$$

for any tensor  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ . This results in a contradiction to (5.23) since

$$\begin{aligned} &\|(P(X(\tau^*)) - P(Y))B^*\|_F \\ &\leq \tau^* \sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} \frac{8\delta}{\rho_{\mathcal{N}}(1 - \tilde{c} - c_{\mathcal{N}})} \|B^*\|_F \leq \tau^* \tilde{c} \|B^*\|_F < \tilde{c} \|B^*\|_F. \end{aligned} \quad \square$$

In order to simplify the previous theorem, we assume  $c_{\mathcal{N}} = c$  for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and maximize this constant under the constraints of Lemma 10, i.e.

$$\sum_{\mathcal{N} \in T \setminus \{\mathcal{N}_0\}} 8c_{\mathcal{N}}(\tilde{c}(1 - \tilde{c} - c_{\mathcal{N}}))^{-1} \leq 1 \quad \text{and} \quad \tilde{c} + c_{\mathcal{N}} < 1.$$

Since the number of nodes in the dimension tree is always  $2d - 1$ , the maximum will be attained for

$$\tilde{c} = -8(2d - 2) + \sqrt{8^2(2d - 2)^2 + 8(2d - 2)} \quad \text{and} \quad c = \frac{-\tilde{c}^2 + \tilde{c}}{8(2d - 2) + \tilde{c}}.$$

As  $\tilde{c} \rightarrow \frac{1}{2}$  for  $d \rightarrow \infty$ , we let

$$\tilde{c} = \frac{1}{2} \quad \text{and} \quad c = (64d - 62)^{-1}$$

and obtain the next corollary.

**Corollary 11** (curvature bounds) *Let  $(Y, \tilde{Y}) \in (\mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T}))^2$  tensors, which are path-connected with respect to ODE (5.2). Let  $\rho > 0$  such that*

$$\sigma_{r_{\mathcal{N}}}(Y^{\mathcal{N}, \mathcal{N}_C}) \geq \rho > 0 \quad \text{and} \quad \|Y - \tilde{Y}\|_F \leq (64d - 62)^{-1} \rho$$



is satisfied for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and  $\beta := (32d - 31)\rho^{-1}$ . Then

$$\|(P(Y) - P(\tilde{Y}))B\|_F \leq \beta \|Y - \tilde{Y}\|_F \|B\|_F \tag{5.24}$$

$$\|P^\perp(Y)(\tilde{Y} - Y)\|_F \leq 2\beta \|Y - \tilde{Y}\|_F^2 \tag{5.25}$$

for any tensor  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$ .

*Proof* Inserting  $c = c_{\mathcal{N}} = (64d - 62)^{-1}$  for all  $\mathcal{N} \in T \setminus \{\mathcal{N}_0\}$  and  $\tilde{c} = \frac{1}{2}$  in the bounds (5.4) and (5.5) of Lemma 10 proves the assertion.  $\square$

### 6 Error analysis

We are finally ready to analyze the accuracy of the variational approximation  $Y(t) \in \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  to the solution of (3.1). The error bound presented in Theorem 12 is the main result of this article. Very similar results have independently and simultaneously been obtained in [12].

Let  $Y_{ex}(t) \in \mathbb{R}^{N_1 \times \dots \times N_d}$  be the solution of the differential equation (3.1), i.e.  $\dot{Y}_{ex} = F(Y_{ex})$  for all  $t \in [0, t_{\text{end}}]$  with initial value  $Y_{ex}(0) \in \mathbb{R}^{N_1 \times \dots \times N_d}$ . Let  $Y(t) \in \mathcal{M} = \mathcal{H}\text{-Tucker}((r_{\mathcal{N}})_{\mathcal{N} \in T})$  be the variational approximation defined by (3.2), i.e.  $\dot{Y} = P(Y)F(Y)$ . For simplicity, it is assumed that  $Y_{ex}(0) \in \mathcal{M}$  and that  $Y(0) = Y_{ex}(0)$ . (Otherwise, an additional error term for the initial error has to be included in Theorem 12.) Moreover, let  $X(t) \in \mathcal{M}$  be the best approximation of  $Y_{ex}(t)$  on  $\mathcal{M}$ , i.e. for every  $t \in [0, T]$  we have

$$\|X(t) - Y_{ex}(t)\|_F \leq \|\widehat{X} - Y_{ex}(t)\|_F \quad \forall \widehat{X} \in \mathcal{M}.$$

Since by definition and by the triangle inequality we have

$$\|X(t) - Y_{ex}(t)\|_F \leq \|Y(t) - Y_{ex}(t)\|_F \leq \|Y(t) - X(t)\|_F + \|X(t) - Y_{ex}(t)\|_F,$$

it makes sense to prove a result which bounds the error  $\|Y(t) - X(t)\|_F$  in terms of  $\|X(t) - Y_{ex}(t)\|_F$ . The following assumptions are made for all  $t \in [0, t_{\text{end}}]$ .

- (A1)  $X(t)$  is continuously differentiable.
- (A2) There is a constant  $\mu > 0$  such that

$$\|F(Y_{ex}(t))\|_F \leq \mu, \quad \|F(X(t))\|_F \leq \mu \quad \text{and} \quad \|F(Y(t))\|_F \leq \mu.$$

- (A3) There exist constants  $L > 0$  and  $\lambda \in \mathbb{R}$  such that the Lipschitz conditions

$$\begin{aligned} \|F(Y_{ex}(t)) - F(X(t))\|_F &\leq L \|Y_{ex}(t) - X(t)\|_F \quad \text{and} \\ \langle F(Y(t)) - F(X(t)), Y(t) - X(t) \rangle &\leq \lambda \|Y(t) - X(t)\|_F^2 \end{aligned}$$

are fulfilled.

(A4) The singular values of the matricizations of the best approximation are bounded from below by

$$\sigma_{r_{\mathcal{N}}}(X^{(\mathcal{N}, \mathcal{N}_c)}(t)) \geq \rho > 0 \quad \text{for all } \mathcal{N} \in T \setminus \{\mathcal{N}_0\}.$$

(A5) The distance from the best approximation  $X(t)$  to the exact solution  $Y_{ex}(t)$  is bounded by

$$\|X(t) - Y_{ex}(t)\|_F \leq \frac{1}{2\beta},$$

where  $\beta = (32d - 31)\rho^{-1}$ .

(A6) The tuple  $(Y(t), X(t))$  is path-connected with respect to ODE (5.2).

Similar assumptions have been made in [8] where the approximation error of the non-hierarchical Tucker format has been analyzed.

**Theorem 12** *Under the above assumptions the difference between the variational approximation and the best approximation is bounded by*

$$\|Y(t) - X(t)\|_F \leq (L + 2\mu\beta) \int_0^t \|Y_{ex}(s) - X(s)\|_F e^{(5\mu\beta + \lambda)(t-s)} ds \tag{6.1}$$

as long as the right-hand side of (6.1) is bounded by  $\frac{1}{2\beta}$ .

*Proof* Because of  $X(t)$  being the best approximation, the deviation  $Y_{ex}(t) - X(t)$  is orthogonal to the tangent space  $\mathcal{T}_X\mathcal{M}$ ,

$$P(X(t))(Y_{ex}(t) - X(t)) = 0$$

and differentiating yields

$$\begin{aligned} 0 &= \frac{d}{dt} [P(X(t))(Y_{ex}(t) - X(t))] \\ &= [P'(X(t)) \cdot (Y_{ex}(t) - X(t))] \dot{X}(t) + P(X(t))(\dot{Y}_{ex}(t) - \dot{X}(t)). \end{aligned}$$

With  $\dot{X}(t) = P(X(t))\dot{X}(t)$  we obtain

$$\begin{aligned} \dot{X}(t) &= P(X(t))\dot{Y}_{ex}(t) + D(t) \\ \text{with } D(t) &= P'(X(t)) \cdot (Y_{ex}(t) - X(t))\dot{X}(t). \end{aligned} \tag{6.2}$$

Next, we derive an estimate for  $D(t)$ . For an arbitrary tensor  $B \in \mathbb{R}^{N_1 \times \dots \times N_d}$  the bound (5.24) yields

$$\begin{aligned} \|(P'(X(t)) \cdot B)\dot{X}(t)\|_F &= \left\| \frac{d}{dt} [P(X(t))B] \right\|_F \\ &= \lim_{h \rightarrow 0} \frac{1}{h} \|[P(X(t+h)) - P(X(t))]B\|_F \end{aligned}$$

$$\begin{aligned} &\leq \lim_{h \rightarrow 0} \frac{1}{h} \beta \|X(t+h) - X(t)\|_F \|B\|_F \\ &= \beta \|\dot{X}(t)\|_F \|B\|_F, \end{aligned}$$

and for  $B := Y_{ex}(t) - X(t)$  this yields

$$\|D(t)\|_F \leq \beta \|\dot{X}(t)\|_F \|Y_{ex}(t) - X(t)\|_F. \tag{6.3}$$

In the following we let  $\delta(t) := \|Y_{ex}(t) - X(t)\|_F$  and  $\varepsilon(t) := \|Y(t) - X(t)\|_F$ . Inserting (6.2) into (6.3) leads to

$$\|D(t)\|_F \leq \beta \delta(t) (\|P(X(t))\dot{Y}_{ex}(t)\|_F + \|D(t)\|_F).$$

With Assumption (A5), i.e.  $\|\delta(t)\|_F \leq \frac{1}{2\beta}$ , we obtain the estimate

$$\|D(t)\|_F \leq \beta \delta(t) \|P(X(t))\dot{Y}_{ex}(t)\|_F + \frac{1}{2} \|D(t)\|_F,$$

and with Assumption (A2)

$$\|D(t)\|_F \leq 2\beta \delta(t) \|P(X(t))\dot{Y}_{ex}(t)\|_F \leq 2\beta \delta(t) \mu.$$

Subtracting (6.2) from  $\dot{Y}(t) = P(Y(t))F(Y(t))$  gives

$$\begin{aligned} \dot{Y}(t) - \dot{X}(t) &= P(Y(t))F(Y(t)) - P(X(t))F(Y_{ex}(t)) - D(t) \\ &= [P(Y(t)) - P(X(t))]F(X(t)) \\ &\quad + P(X(t))[F(X(t)) - F(Y_{ex}(t))] + [F(Y(t)) - F(X(t))] \\ &\quad - P^\perp(Y(t))[F(Y(t)) - F(X(t))] - D(t). \end{aligned} \tag{6.4}$$

Next, we want to find an estimate for the inner product  $\langle \dot{Y}(t) - \dot{X}(t), Y(t) - X(t) \rangle$ . For this purpose we consider the single terms on the right hand side of (6.4) and, via Corollary 11, obtain the following estimates:

1.  $\langle [P(Y(t)) - P(X(t))]F(X(t)), Y(t) - X(t) \rangle$   
 $\leq \| [P(Y(t)) - P(X(t))]F(X(t)) \|_F \|Y(t) - X(t)\|_F$   
 $\leq \beta \|Y(t) - X(t)\|_F^2 \|F(X(t))\|_F$   
 $\leq \mu \beta \varepsilon^2(t),$
2.  $\langle P(X(t))[F(X(t)) - F(Y_{ex}(t))], Y(t) - X(t) \rangle$   
 $\leq \|F(X(t)) - F(Y_{ex}(t))\|_F \|Y(t) - X(t)\|_F$   
 $\leq L \|X(t) - Y_{ex}(t)\|_F \|Y(t) - X(t)\|_F$   
 $= L \delta(t) \varepsilon(t),$

3.  $\langle F(Y(t)) - F(X(t)), Y(t) - X(t) \rangle \leq \lambda \|Y(t) - X(t)\|_F^2 = \lambda \varepsilon^2(t),$
4.  $\begin{aligned} \langle P^\perp(Y(t))[F(Y(t)) - F(X(t))], Y(t) - X(t) \rangle \\ = \langle F(Y(t)) - F(X(t)), P^\perp(Y(t))(Y(t) - X(t)) \rangle \\ \leq \|F(Y(t)) - F(X(t))\|_F \|P^\perp(Y(t))(Y(t) - X(t))\|_F \\ \leq (\|F(Y(t))\|_F + \|F(X(t))\|_F) 2\beta \|Y(t) - X(t)\|_F^2 \\ \leq 4\mu\beta\varepsilon^2(t), \end{aligned}$
5.  $\langle D(t), Y(t) - X(t) \rangle \leq \|D(t)\|_F \|Y(t) - X(t)\|_F \leq 2\beta\mu\delta(t)\varepsilon(t).$

With these inequalities, we obtain

$$\begin{aligned} \varepsilon(t)\dot{\varepsilon}(t) &= \|Y(t) - X(t)\|_F \frac{d}{dt} \|Y(t) - X(t)\|_F \\ &= \frac{1}{2} \frac{d}{dt} \|Y(t) - X(t)\|_F^2 \\ &= \langle \dot{Y}(t) - \dot{X}(t), Y(t) - X(t) \rangle \\ &\leq \mu\beta\varepsilon^2(t) + L\delta(t)\varepsilon(t) + \lambda\varepsilon^2(t) + 4\mu\beta\varepsilon^2(t) + 2\beta\delta(t)\mu\varepsilon(t) \end{aligned}$$

and hence

$$\dot{\varepsilon}(t) \leq (5\mu\beta + \lambda)\varepsilon(t) + (L + 2\mu\beta)\delta(t).$$

Finally Gronwall's inequality leads to the assertion

$$\varepsilon(t) \leq (L + 2\mu\beta) \int_0^t \delta(s) e^{(5\mu\beta + \lambda)(t-s)} ds. \quad \square$$

**Acknowledgements** The authors thank the anonymous referees for their helpful remarks and suggestions.

## References

1. De Lathauwer, L., De Moor, B., Vandewalle, J.: A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.* **21**(4), 1253–1278 (2000)
2. Grasedyck, L.: Hierarchical singular value decomposition of tensors. *SIAM J. Matrix Anal. Appl.* **31**, 2029–2054 (2010)
3. Hackbusch, W.: *Tensor Spaces and Numerical Tensor Calculus*. Springer Series in Computational Mathematics, vol. 42. Springer, Berlin (2012)
4. Hackbusch, W., Kühn, S.: A new scheme for the tensor representation. *J. Fourier Anal. Appl.* **15**, 706–722 (2009)
5. Holtz, S., Rohwedder, T., Schneider, R.: On manifolds of tensors of fixed tt-rank. *Numer. Math.* **120**(4), 701–731 (2012)
6. Horn, R.A., Johnson, C.R.: *Matrix Analysis*. Cambridge University Press, Cambridge (1985), 1. publ. edition
7. Koch, O., Lubich, C.: Dynamical low-rank approximation. *SIAM J. Matrix Anal. Appl.* **29**, 434–454 (2007)

8. Koch, O., Lubich, C.: Dynamical tensor approximation. *SIAM J. Matrix Anal. Appl.* **31**, 2360–2375 (2010)
9. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. *SIAM Rev.* **51**, 455–500 (2009)
10. Kressner, D., Tobler, C.: *htucker*—a Matlab toolbox for tensors in hierarchical Tucker format. Technical report, ETH, Zurich (2012)
11. Lubich, C.: From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis. Zurich Lectures in Advanced Mathematics. European Mathematical Society, Zürich (2008)
12. Lubich, C., Rohwedder, T., Schneider, R., Vandereycken, B.: Dynamical approximation by hierarchical Tucker and tensor-train tensors. *SIAM J. Matrix Anal. Appl.* **34**, 470–494 (2013)
13. Manthe, U.: A multilayer multiconfigurational time-dependent Hartree approach for quantum dynamics on general potential energy surfaces. *J. Chem. Phys.* **128**, 164116 (2008)
14. Uschmajew, A., Vandereycken, B.: The geometry of algorithms using hierarchical tensors. *Linear Algebra Appl.* **439**, 133–166 (2013)
15. Vendrell, O., Meyer, H.-D.: Multilayer multi-configuration time-dependent Hartree method: implementation and applications to a Henon-Heiles Hamiltonian and to pyrazine. *J. Chem. Phys.* **134**(4), 044135 (2011)
16. Wang, H., Thoss, M.: Multilayer formulation of the multiconfiguration time-dependent Hartree theory. *J. Chem. Phys.* **119**, 1289–1299 (2003)