

## DYNAMICAL APPROXIMATION BY HIERARCHICAL TUCKER AND TENSOR-TRAIN TENSORS\*

CHRISTIAN LUBICH<sup>†</sup>, THORSTEN ROHWEDDER<sup>‡</sup>, REINHOLD SCHNEIDER<sup>‡</sup>, AND  
BART VANDEREYCKEN<sup>§</sup>

**Abstract.** We extend results on the dynamical low-rank approximation for the treatment of time-dependent matrices and tensors (Koch and Lubich; see [*SIAM J. Matrix Anal. Appl.*, 29 (2007), pp. 434–454], [*SIAM J. Matrix Anal. Appl.*, 31 (2010), pp. 2360–2375]) to the recently proposed hierarchical Tucker (HT) tensor format (Hackbusch and Kühn; see [*J. Fourier Anal. Appl.*, 15 (2009), pp. 706–722]) and the tensor train (TT) format (Oseledets; see [*SIAM J. Sci. Comput.*, 33 (2011), pp. 2295–2317]), which are closely related to tensor decomposition methods used in quantum physics and chemistry. In this dynamical approximation approach, the time derivative of the tensor to be approximated is projected onto the time-dependent tangent space of the approximation manifold along the solution trajectory. This approach can be used to approximate the solutions to tensor differential equations in the HT or TT format and to compute updates in optimization algorithms within these reduced tensor formats. By deriving and analyzing the tangent space projector for the manifold of HT/TT tensors of fixed rank, we obtain curvature estimates, which allow us to obtain quasi-best approximation properties for the dynamical approximation, showing that the prospects and limitations of the ansatz are similar to those of the dynamical low rank approximation for matrices. Our results are exemplified by numerical experiments.

**Key words.** low-rank approximation, time-varying tensors, hierarchical Tucker format, tensor train format, tensor differential equations, tensor updates

**AMS subject classifications.** 15A69, 90C06, 65F99, 65L05

**DOI.** 10.1137/120885723

**1. Introduction.** In this work, we employ the recent hierarchical Tucker (HT) format of [7, 9] and tensor train (TT) format of [25] for the data-sparse approximate computation of a time-varying family of tensors in a  $d$ -dimensional tensor space  $\mathcal{V}$ , namely,

$$(1.1) \quad A(t) \in \mathcal{V} := \mathbb{R}^{n_1 \times n_2 \times \cdots \times n_d}, \quad t \in [0, \bar{t}].$$

The tensor  $A(t)$  need not be known explicitly but it may instead be given implicitly as the solution of a tensor differential equation  $\dot{A}(t) = F(A)$ , in combination with a known initial value  $A(0) \in \mathcal{V}$ . Applications where this problem arises are (spatial discretizations of) time-dependent PDEs or ODEs formulated on a high-dimensional space, such as the Fokker–Planck equation, instationary Schrödinger-type equations, or master equation approaches to stochastic systems. Another less obvious application is in efficiently computing updates to a given tensor in the HT or TT format, as they are required in optimization algorithms using these formats.

The storage of explicit representations of the solution  $A(t)$  scales exponentially with the dimension  $d$  and is therefore infeasible in most cases. Instead, given an

---

\*Received by the editors July 23, 2012; accepted for publication (in revised form) by L. Grasedyck February 25, 2013; published electronically May 9, 2013.

<http://www.siam.org/journals/simax/34-2/88572.html>

<sup>†</sup>Mathematisches Institut, Universität Tübingen, D-72076 Tübingen, Germany (lubich@na.uni-tuebingen.de).

<sup>‡</sup>Institut für Mathematik, Technische Universität Berlin, D-10623 Berlin, Germany (rohwedde@math.tu-berlin.de, schneidr@math.tu-berlin.de).

<sup>§</sup>ANCHP, MATHICSE, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland (bartv@math.princeton.edu).

embedded manifold  $\mathcal{M} \subseteq \mathcal{V}$  typically depending on much fewer parameters than the linear parametrization of  $\mathcal{V}$ , the *dynamical tensor approximation* may be utilized: Assuming  $Y(0) \in \mathcal{M}$ , an approximation  $Y(t) \in \mathcal{M}$  is determined such that its derivative at every time  $t$  is the element of the tangent space  $\mathcal{T}_{Y(t)}\mathcal{M}$  closest to  $\dot{A}(t)$ :

$$(1.2) \quad \dot{Y}(t) \in \mathcal{T}_{Y(t)}\mathcal{M} \quad \text{with} \quad \|\dot{Y}(t) - \dot{A}(t)\| = \min.$$

In terms of the orthogonal projection  $P_X$  onto the tangent space  $\mathcal{T}_X\mathcal{M}$  at  $X \in \mathcal{M}$ , the solution to (1.2) is equivalently characterized by projecting

$$(1.3) \quad \dot{Y}(t) = P_{Y(t)}\dot{A}(t), \quad Y(0) \in \mathcal{M},$$

which results in a differential equation on the approximation manifold  $\mathcal{M}$ .

After an explicit parametrization of the manifold  $\mathcal{M}$  under consideration, one obtains from (1.3) a set of nonlinear differential equations for the parameters of this parametrization, suitable for numerical integration. The above ansatz has been studied for manifolds of matrices of fixed rank  $k$  and tensors of fixed (multilinear) Tucker rank  $(k_1, \dots, k_d)$  in [15, 24, 16], respectively. In the case of a tensor differential equation,  $\dot{A}(t) = F(A(t))$  is replaced by the approximate value  $F(Y(t))$  on the right-hand side of (1.2). In the context of the time-dependent Schrödinger equation, the approach is known as the Dirac–Frenkel time-dependent variational principle [20].

Classical tensor formats stemming from data analysis (that is, the canonical decomposition and the Tucker format [8, 17]) exhibit certain structural weaknesses that make them unsuitable for the treatment of problems of the kind exemplified by (1.1) and (1.2) when the dimension  $d$  is large; see, e.g., the introduction in [11]. This motivated the development of recent tensor formats such as the HT format, in which a recursive, hierarchical construction of Tucker type is employed for tensor representation, and the TT format, which can be interpreted as a special case of HT in which the recursive formulation can be avoided (see section 2). The usefulness of these formats is currently being investigated for a variety of high-dimensional problems; see, e.g., [3, 11, 13, 18]. Formats of this type have been used successfully in the quantum physics and chemistry communities in the last decade [21, 23, 27, 32].

Analogous to the Tucker format, both HT and TT allow the definition of a rank vector  $\mathbf{k}$  (to be specified in section 2), and the manifold of tensors of fixed HT or TT rank can be shown to be an embedded manifold [12, 30], to which the above dynamical tensor approximation approach may then be applied. In addition to approximating time-varying tensors and the solutions of tensor ODEs, it may also be used if, given a HT tensor  $A$  and a search direction  $\Delta A$ , one wants to compute an approximation to  $A + \Delta A$  in the HT format. Approximately computing the dynamical approximation to  $A + t\Delta A$  with one or few Euler steps provides an alternative to truncation steps by higher order SVD [7], or by local optimization methods like alternating least squares methods [11]. The dynamical approximation may thus be used as a retraction, as is required in optimization methods on manifolds like CG and Newton; see [1, 31].

In this paper, we generalize the approximation results from [15, 16] to the case where an HT manifold is used in the ansatz (1.2), with the TT format included as a special case. Our motivations for this are twofold:

(i) While the general procedure in this work is essentially analogous to that of [15, 16], extending the central ingredient of the proof—obtaining curvature estimates for  $\mathcal{M}$  via the projector onto the tangent space—is not straightforward due to the more involved, hierarchical construction in the HT setting.

(ii) Although the HT/TT formats were only recently introduced into the numerical analysis community, they have—similar to the parallels between the Tucker format and the formulation of the multiconfiguration time-dependent Hartree (MCTDH) method in quantum chemistry [5, 4, 22]—a considerable significance in various fields concerned with quantum computations. In particular, TT tensors correspond to matrix product states (MPS) used to describe many-body quantum physics systems, and are the central quantity in the successful density matrix renormalization group (DMRG) algorithm. MPS and DMRG have seen a boost of interest in the last few years, and the treatment of time-dependent equations by the use of MPS is a recent field of interest (see the recent review [27] and the references therein). In particular, the geometric approach underlying this work has lately been used for the simulation of real- and imaginary-time dynamics for infinite one-dimensional quantum lattices by MPS [10]. In parallel to the advances with the HT format, developments using tensor trees and networks have recently been made independently in the quantum physics community; e.g. [21, 23]. Moreover, the multilayer formulation of the MCTDH method [32] can be interpreted as an instance of the dynamical tensor approximation approach for HT. Aside from quantum physics and chemistry, applications of the approach may also enable the numerical treatment of other high-dimensional problems; see, e.g., [6, 14] for approaches to other time-dependent equations using TT. With this work, we complement these practically motivated approaches with theoretical results to hopefully foster a unified understanding of the method and to point to the prospects and limitations of the ansatz from a mathematical perspective.

The rest of the paper is organized as follows: We recall some facts on the HT representation for tensors and on the manifold of tensors of fixed HT rank in sections 2 and 3, respectively. The central technical work is performed in section 4, where to enable and analyze the ansatz (1.3), the orthogonal projection  $P_X$  onto the tangent space  $\mathcal{T}_X\mathcal{M}$  at given  $X \in \mathcal{M}$  is analyzed and curvature estimates for the fixed-rank HT manifold are obtained. Section 5 then states approximation results for the approach (1.2)–(1.3) as well as for the ODE case where  $\dot{A}$  is given as a function  $F(A)$ . Section 6 briefly discusses the use of dynamical approximation for computing updates in the HT format in optimization algorithms. Section 7 describes the implementation of the tangent space projection, which is the key ingredient in the dynamical approximation approach, and features some numerical examples that are designed to display the properties of the approach as derived in the theoretical analysis.

**2. The HT and TT tensor formats.** In this section, we review some basic concepts related to tensors that are needed for the treatment of dynamical approximations on  $\mathcal{V}$ . In particular, we give a short introduction into the construction of the HT format (which includes the TT format as a special case).

**2.1. Prerequisites.** We regard a tensor  $X \in \mathcal{V}$  as a  $d$ -dimensional array where the entries are indexed as

$$X(x_1, x_2, \dots, x_d), \quad x_i = 1, 2, \dots, n_i, \quad i = 1, 2, \dots, d.$$

Denote any splitting of the set  $\tau_r := \{1, \dots, d\}$  into two disjoint subsets by

$$(2.1) \quad \tau = \{i_1, i_2, \dots, i_s\} \subseteq \tau_r, \quad \tau^c := \tau_r \setminus \tau = \{j_1, j_2, \dots, j_{d-s}\}.$$

Matrices will always be denoted in boldface, for example,  $\mathbf{X} \in \mathbb{R}^{n_1 \times n_2}$ . The *mode*  $\tau$ -*unfolding*  $\mathbf{X}^{(\tau)}$  of  $X$  is the matrix obtained by grouping the indices  $x_{i_1}, x_{i_2}, \dots, x_{i_s}$  of  $X(x_1, x_2, \dots, x_d)$  into row indices and the remaining indices into column indices,

each in reverse lexicographical ordering (this is the same ordering as in [17, 19] but different from [30]). In other words,

$$(2.2) \quad \mathbf{X}^{(\tau)} \in \mathbb{R}^{(n_{i_1} n_{i_2} \cdots n_{i_s}) \times (n_{j_1} n_{j_2} \cdots n_{j_{d-s}})} \quad \text{such that} \\ \mathbf{X}_{(x_{i_1}, x_{i_2}, \dots, x_{i_s}), (x_{j_1}, x_{j_2}, \dots, x_{j_{d-s}})}^{(\tau)} := X(x_1, x_2, \dots, x_d).$$

These unfoldings can also be used to define the *mode- $\tau$  rank* of  $X$  as

$$(2.3) \quad k_\tau = \text{rank}(\mathbf{X}^{(\tau)}).$$

For splittings where  $\tau = \{i\}$  is a singleton, one obtains the special case of the mode- $i$ -unfolding  $\mathbf{X}^{(i)} \in \mathbb{R}^{n_i \times (n_1 \cdots n_{i-1} n_{i+1} \cdots n_d)}$  of  $X$ . Throughout the paper, sets that are singletons will often just be indexed by  $i$  instead of by  $\{i\}$  as in  $X^{(i)} = X^{\{\{i\}\}}$  and  $k_i = k_{\{i\}}$ . Collecting the corresponding mode- $i$ -separation ranks  $k_i$  for  $i \in \{1, \dots, d\}$  in the vector

$$\text{rank}_{\text{ML}}(X) := (k_1, k_2, \dots, k_d)$$

gives the *multilinear rank* of  $X$ , common in the context of Tucker decomposition [17]. Taking all indices  $\tau_r := \{1, \dots, d\}$  of the tensor  $X$  as column indices yields the vectorization

$$\text{vec}(X) \in \mathbb{R}^{n_1 n_2 \cdots n_d}$$

of  $X$ , a column vector containing all the entries of  $X$  in reverse lexicographical ordering.

We equip the tensor space  $\mathcal{V}$  with the inner product

$$\langle X, Y \rangle := \sum_{x_1=1}^{n_1} \cdots \sum_{x_d=1}^{n_d} X(x_1, x_2, \dots, x_d) \cdot Y(x_1, x_2, \dots, x_d) = \text{tr}((\mathbf{X}^{(\tau)})^T \mathbf{Y}^{(\tau)})$$

and corresponding norm

$$\|X\| := \left( \sum_{x_1=1}^{n_1} \cdots \sum_{x_d=1}^{n_d} |X(x_1, x_2, \dots, x_d)|^2 \right)^{\frac{1}{2}} = \|\mathbf{X}^{(\tau)}\|_{\text{Frob}}.$$

We will also come across the spectral norm  $\|\mathbf{A}\|_2$  of matrices  $\mathbf{A}$ ; in particular, we will make use of the inequality

$$(2.4) \quad \|\mathbf{ABC}\| \leq \|\mathbf{A}\|_2 \|\mathbf{B}\| \|\mathbf{C}\|_2$$

which holds for any matrices  $\mathbf{A}, \mathbf{B}, \mathbf{C}$  of suitable size.

Given two matrices  $\mathbf{A}_1 \in \mathbb{R}^{p_1 \times m_1}$ ,  $\mathbf{A}_2 \in \mathbb{R}^{p_2 \times m_2}$ , recall that their Kronecker product may be defined pointwise by

$$\mathbf{A}_1 \otimes \mathbf{A}_2 \in \mathbb{R}^{(p_1 p_2) \times (m_1 m_2)}, \quad (\mathbf{A}_1 \otimes \mathbf{A}_2)_{(x_1, x_2), (y_1, y_2)} := (\mathbf{A}_1)_{(x_1, y_1)} \cdot (\mathbf{A}_2)_{(x_2, y_2)},$$

where  $(x_1, y_1)$  is a multi-index, enumerated lexicographically.

To denote in a succinct fashion the HT decomposition to be introduced below, we will follow the notation chosen in [19]: For an order-3 tensor  $B \in \mathbb{R}^{m_1 \times m_2 \times m_3}$ , its

indexwise multiplication by matrices  $\mathbf{A}_1 \in \mathbb{R}^{p_1 \times m_1}$ ,  $\mathbf{A}_2 \in \mathbb{R}^{p_2 \times m_2}$ ,  $\mathbf{A}_3 \in \mathbb{R}^{p_3 \times m_3}$ , that is, the three-way product  $C = (\mathbf{A}_1, \mathbf{A}_2, \mathbf{A}_3) \circ B \in \mathbb{R}^{p_1 \times p_2 \times p_3}$ , defined elementwise as

$$C(y_1, y_2, y_3) := \sum_{x_1=1}^{m_1} \sum_{x_2=1}^{m_2} \sum_{x_3=1}^{m_3} \mathbf{A}_1(x_1, y_1) \mathbf{A}_2(x_2, y_2) \mathbf{A}_3(x_3, y_3) B(x_1, x_2, x_3),$$

can be expressed succinctly as

$$(2.5) \quad (\mathbf{C}^{(3)})^T = (\mathbf{A}_2 \otimes \mathbf{A}_1) (\mathbf{B}^{(3)})^T \mathbf{A}_3^T,$$

returning  $(\mathbf{C}^{(3)})^T$  instead of the tensor  $C$  itself. Also, we have that (see, e.g., [17])

$$(2.6) \quad \mathbf{C}^{(1)} = \mathbf{A}_1 \mathbf{B}^{(1)} (\mathbf{A}_3 \otimes \mathbf{A}_2)^T, \quad \mathbf{C}^{(2)} = \mathbf{A}_2 \mathbf{B}^{(2)} (\mathbf{A}_3 \otimes \mathbf{A}_1)^T.$$

**2.2. The HT and TT formats.** We recall the definition of the hierarchical Tucker (HT) format and the set of tensors of fixed hierarchical rank as introduced in [7, 9]. Aside from some necessary concepts from [7, 9], we also introduce notation similar to that of [19, 18, 30] that is better suited to our analysis.

The basic idea of the hierarchical Tucker decomposition is that a tensor  $X \in \mathcal{V}$  can be represented in the HT format if it allows for a recursive construction out of lower-dimensional subspaces. This recursion is completely defined by a dimension tree.

**DEFINITION 2.1.** *For fixed dimension  $d$ , a nontrivial binary tree is called a dimension tree  $T$  if the following holds:*

- (i) *The node  $\tau_r = \{1, 2, \dots, d\}$  is the root of  $T$ .*
- (ii) *All nodes of  $T$  are nonempty subsets of the root.*
- (iii) *Every node  $\tau \in T$  that is not a singleton has two sons,  $\tau_1$  and  $\tau_2$ , such that*

$$\tau = \tau_1 \cup \tau_2, \quad \mu < \nu \quad \text{for all } \mu \in \tau_1, \nu \in \tau_2.$$

*The leaves of  $T$  are the singletons  $\{i\}$  with  $i \in \{1, 2, \dots, d\}$ .*

*The set of leaves is denoted by  $L$ ; the collection of all inner nodes by  $I := T \setminus L$ .*

For convenience, we require the splittings to be ordered in condition (iii). This is, however, not a restriction since one can always permute the indices of  $X$  such that the indices of the new tensor  $\tilde{X}$  satisfy this condition. During the evaluation of  $\tilde{X}$ , this permutation can be trivially undone to recover  $X$ . Also note that although the HT decomposition is introduced based on a binary tree, the results in this paper also hold via straightforward modifications for decomposition trees with more than two sons in each node. Therefore, the present work includes the matrix case and also the Tucker tensor case as special cases.

The problem of determining the “best” dimension tree for a given tensor—and the related question of the “best” permutation of its indices—is highly nontrivial. In the present context, we assume that the permutation and the dimension tree have been chosen. Typical choices for such dimension trees are those obtained by a *balanced splitting* of the index set, as used, e.g., by Grasedyck [7], or a *front-to-back splitting* which (with minor adaptations; see Remark 1) leads to the TT format due to Oseledets and Tyrtyshnikov [25, 26]. The corresponding dimension trees are visualized for  $d = 5$  in Figure 2.1.

Note that, irrespective of the particular tree, we have

$$(2.7) \quad |I| = d - 1, \quad |L| = d.$$

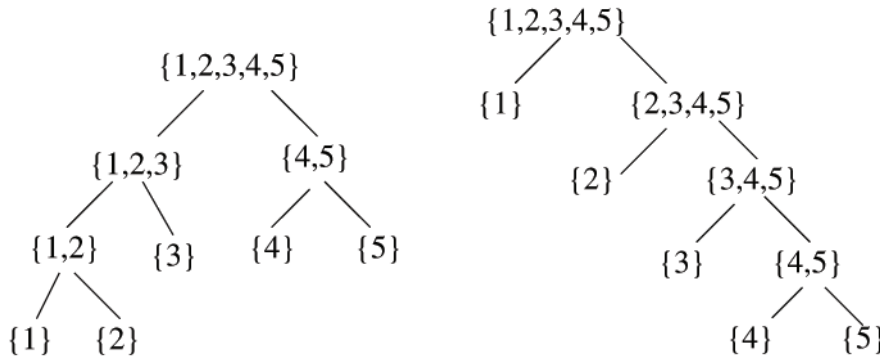


FIG. 2.1. The balanced and front-to-back dimension trees for the HT and TT formats, respectively.

Let

$$n_\tau := \prod_{\mu \in \tau} n_\mu.$$

Then we can introduce the set of HT tensors of a given hierarchical rank as follows.

DEFINITION 2.2. Let  $T$  be a dimension tree and let  $\mathbf{k} = (k_\tau)_{\tau \in T}$  be positive integers with  $k_\tau = 1$ . Then,  $X \in \mathcal{V}$  is an HT tensor of  $T$ -rank  $\mathbf{k}$ —or, a  $(T, \mathbf{k})$ -tensor—if it can be constructed in the following way:

- (i) To each leaf node  $\tau \in L$ , assign a leaf matrix  $\mathbf{U}_\tau \in \mathbb{R}^{n_\tau \times k_\tau}$  of full rank,

$$\text{rank}(\mathbf{U}_\tau) = k_\tau.$$

The collection of all leaf matrices is denoted by  $\mathbf{U}_L := (\mathbf{U}_\tau)_{\tau \in L}$ .

- (ii) To each inner node  $\tau \in I$ , assign a transfer tensor  $B_\tau \in \mathbb{R}^{k_{\tau_1} \times k_{\tau_2} \times k_\tau}$  of full multilinear rank

$$\text{rank}(\mathbf{B}_\tau^{(1)}) = k_{\tau_1}, \quad \text{rank}(\mathbf{B}_\tau^{(2)}) = k_{\tau_2}, \quad \text{rank}(\mathbf{B}_\tau^{(3)}) = k_\tau.$$

For notational simplicity, we define the corresponding transfer matrix as

$$\mathbf{B}_\tau := (\mathbf{B}_\tau^{(3)})^T.$$

The collection of all transfer matrices is denoted by  $\mathbf{B}_I := (\mathbf{B}_\tau)_{\tau \in I}$ .

- (iii) To each inner node  $\tau \in I$ , recursively construct the inner frame matrices

$$(2.8) \quad \mathbf{U}_\tau = (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \mathbf{B}_\tau \in \mathbb{R}^{n_\tau \times k_\tau} \quad \text{for } \tau \in I$$

such that  $X$  is recovered as  $\mathbf{X}^{(\tau_r)} = \mathbf{U}_{\tau_r}$ .

The relation (2.8) can equivalently be expressed elementwise as

$$(\mathbf{U}_\tau)_{(l_1, l_2), i} = \sum_{j_1=1}^{k_{\tau_1}} \sum_{j_2=1}^{k_{\tau_2}} B_\tau(j_1, j_2, i) (\mathbf{U}_{\tau_1})_{l_1, j_1} (\mathbf{U}_{\tau_2})_{l_2, j_2},$$

with  $(l_1, l_2)$  a multi-index for the rows of  $U_\tau$ . We prefer the more succinct version (2.8).

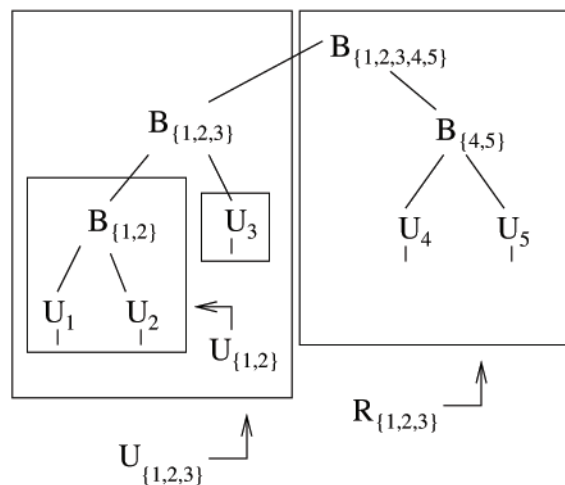


FIG. 2.2. Graphical representation of the leaf matrices  $\mathbf{U}_L$  and the transfer matrices  $\mathbf{B}_I$  associated to the balanced dimension tree from Figure 2.1. The inner matrices  $\{\mathbf{U}_\tau : \tau \in I\}$  can be constructed recursively from (2.8), as, for instance, the matrix  $\mathbf{U}_{\{1,2,3\}}$  from the matrices  $\mathbf{U}_{\{1,2\}}$ ,  $\mathbf{U}_3$ , and  $\mathbf{B}_{\{1,2,3\}}$ .

In Figure 2.2, we depicted the leaf and transfer matrices  $\mathbf{U}_L$  and  $\mathbf{B}_I$  associated to the balanced dimension tree from Figure 2.1. Observe that  $\tau_1$  corresponds to the left child of  $\tau$ , while  $\tau_2$  is the right one. In addition, we also indicated with arrows the inner frame matrices from (2.8); for example,  $\mathbf{U}_{\{1,2,3\}}$  is constructed from the matrices  $\mathbf{U}_{\{1,2\}}$  and  $\mathbf{U}_3$  and the transfer tensor  $\mathbf{B}_{\{1,2,3\}}$ . Although we will not use this feature much, note that the lines in Figure 2.2 have actual interpretations from tensor networks: lines interconnecting tensors are summations over mutual indices, while open lines represent the free indices of the resulting tensor.

Given a dimension tree  $T$ , every tensor  $X \in \mathcal{V}$  can be represented as a  $(T, \mathbf{k})$ -tensor for some  $\mathbf{k} = (k_\tau)_{\tau \in T}$  when  $k_\tau \leq n_\tau$  for all  $\tau \in T$ . An exact HT decomposition with minimal  $\mathbf{k}$  can be computed by Algorithms 1 or 2 in [7], which amounts to the hierarchical application of several SVDs. Due to the full rank requirements on the matrices  $\mathbf{U}_L$  and  $\mathbf{B}_I$  in our definition, this minimal rank vector  $\mathbf{k}$  is unique and coincides elementwise with the mode- $\tau$  ranks of  $X$  from (2.3). Hence, there exists a unique HT rank of  $X$ , defined as

$$(2.9) \quad \text{rank}_{\text{HT}}(X) := \mathbf{k} = (k_\tau)_{\tau \in T}, \quad \text{where } k_\tau = \text{rank}(\mathbf{X}^{(\tau)}).$$

The benefit of representing tensors in the HT format is that only the leaf matrices  $\mathbf{U}_L$  and transfer matrices  $\mathbf{B}_I$  need to be stored. The total number of entries is therefore

$$D = \sum_{\tau \in L} n_\tau k_\tau + \sum_{\tau \in I} k_\tau k_{\tau_1} k_{\tau_2} \leq dNK + (d-2)K^3 + K^2,$$

where  $N := \max\{n_\tau : \tau \in L\}$  and  $K := \max\{k_\tau : \tau \in T\}$ . This is a considerable reduction compared to the full tensor case with  $N^d$  entries provided  $K \ll N$ .

The representation of an HT tensor by  $\mathbf{U}_L$  and  $\mathbf{B}_I$  is not unique; more precisely, two sets of matrices  $(\mathbf{U}_L, \mathbf{B}_I)$  and  $(\tilde{\mathbf{U}}_L, \tilde{\mathbf{B}}_I)$  are both representations of the same tensor  $X \in \mathcal{V}$  (with the construction from Definition 2.2) if and only if there is a collection  $(\mathbf{A}_\tau)_{\tau \in T}$  of invertible matrices  $\mathbf{A}_\tau \in \mathbb{R}^{k_\tau \times k_\tau}$  and  $\mathbf{A}_{\tau_r} = 1$  such that

$$(2.10) \quad \tilde{\mathbf{U}}_\tau = \mathbf{U}_\tau \mathbf{A}_\tau^{-1} \quad \text{for all } \tau \in L,$$

$$(2.11) \quad \tilde{\mathbf{B}}_\tau = (\mathbf{A}_{\tau_2} \otimes \mathbf{A}_{\tau_1}) \mathbf{B}_\tau \mathbf{A}_\tau^{-1} \quad \text{for all } \tau \in I;$$

see [30, Prop. 3.9] for a proof.

*Remark 1 (TT format).* The above definition of the HT format includes the TT format of [25] in the following way: Given the dimension  $d$  and  $n_1, \dots, n_d$ , choose the corresponding “front-to-back splitting” dimension tree  $T_{TT}$  introduced above (see Figure 2.1) and define  $\mathbf{k}$  by letting  $k_i = n_i$  for  $i = 1, \dots, d - 1$  and fixing the complementary ranks  $r_{i-1} := k_{\tau_i}$ ,  $\tau_i = \{i, \dots, d\}$ . The set of TT tensors of TT-rank  $(r_1, \dots, r_{d-1})$  then corresponds to all  $(T_{TT}, \mathbf{k})$ -HT tensors. Because the corresponding leaf matrices  $\mathbf{U}_i$  of size  $n_i \times n_i$  are invertible, it is then a priori possible to avoid the redundancy (2.10) at the leaves  $i = 1, \dots, d - 1$  by fixing  $\mathbf{U}_i = \mathbf{I}_{n_i}$ . With this convention,  $(T_{TT}, \mathbf{k})$ -HT tensors can be characterized using only  $d - 1$  transfer tensors  $\mathbf{B}_{\tau_i} \in \mathbb{R}^{n_i \times r_i \times r_{i-1}}$  and the leaf matrix  $\mathbf{U}_d \in \mathbb{R}^{n_d \times r_{d-1}}$  resulting in the canonical TT format as introduced by [25].

For numerical reasons, it is usually advisable to work with orthonormal matrices.

**DEFINITION 2.3.** *An HT decomposition is called an orthogonal HT decomposition if all leaf and inner frame matrices are orthonormal except at the root,*

$$\mathbf{U}_\tau^T \mathbf{U}_\tau = \mathbf{I}_{k_\tau} \quad \text{for all } \tau \in T \setminus \{\tau_r\}.$$

*Equivalently, all leaf and transfer matrices are orthonormal except at the root,*

$$\mathbf{U}_\tau^T \mathbf{U}_\tau = \mathbf{I}_{k_\tau} \quad \text{for all } \tau \in L \quad \text{and} \quad \mathbf{B}_\tau^T \mathbf{B}_\tau = \mathbf{I}_{k_\tau} \quad \text{for all } \tau \in I \setminus \{\tau_r\}.$$

By a recursive orthogonalization starting at the leaves, any given HT tensor can be represented by an orthogonal HT decomposition with the same hierarchical  $T$ -rank. Note that for TT, with the left-to-right splitting from above, this yields a right-orthogonal decomposition as used, for instance, in [12, 11].

**3. The manifold of HT tensors of fixed rank.** We will now restrict our attention to tensors with fixed dimension tree  $T$  and fixed HT rank  $\mathbf{k} = (k_\tau)_{\tau \in T}$ ,

$$\mathcal{M} := \mathcal{M}(T, \mathbf{k}) = \{X \in \mathcal{V} : X \text{ is an HT tensor of hierarchical } T\text{-rank } \mathbf{k}\}.$$

The set  $\mathcal{M}$  is a smooth, embedded submanifold in  $\mathcal{V}$ . These and other properties of  $\mathcal{M}$  have recently been investigated quite exhaustively in [30]. We therefore refer to this work and only collect the main facts needed for our treatment of dynamical equations on  $\mathcal{M}$ .

**3.1. Parametrization of  $\mathcal{M}$  and its tangent space  $\mathcal{T}_X \mathcal{M}$ .** The elements in  $\mathcal{M}$  can be parameterized by matrices from the following *coefficient space*:

$$\mathcal{C} := \prod_{\tau \in T} \mathcal{C}_\tau$$



with

$$\begin{aligned} \mathcal{C}_\tau &:= \{\mathbf{U}_\tau \in \mathbb{R}^{n_\tau \times k_\tau} : \text{rank}(\mathbf{U}_\tau) = k_\tau\} && \text{for } \tau \in L, \\ \mathcal{C}_\tau &:= \{\mathbf{B}_\tau \in \mathbb{R}^{(k_{\tau_1} k_{\tau_2}) \times k_\tau} : \text{rank}(\mathbf{B}_\tau^{(1)}) = k_{\tau_1}, \text{rank}(\mathbf{B}_\tau^{(2)}) = k_{\tau_2}, \\ &\quad \text{rank}(\mathbf{B}_\tau^{(3)}) = k_\tau\} && \text{for } \tau \in I. \end{aligned}$$

As in (ii) of Definition 2.2, we conveniently abbreviate the mode-3 unfolding of a transfer tensor  $B_\tau \in \mathbb{R}^{k_{\tau_1} \times k_{\tau_2} \times k_\tau}$  by  $\mathbf{B}_\tau^T = \mathbf{B}_\tau^{(3)}$ .

Throughout the paper, we denote elements of  $\mathcal{C}$  in a short-hand way as

$$(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{C} \quad \text{with} \quad \mathbf{U}_L := (\mathbf{U}_\tau)_{\tau \in L} \quad \text{and} \quad \mathbf{B}_I := (\mathbf{B}_\tau)_{\tau \in I}.$$

Given these coefficients  $(\mathbf{U}_L, \mathbf{B}_I)$ , we can construct a unique HT tensor in  $\mathcal{M}$ , denoted as  $X(\mathbf{U}_L, \mathbf{B}_I)$ . According to Definition 2.2, this reconstruction mapping

$$(3.1) \quad \varphi : \mathcal{C} \rightarrow \mathcal{M} \subset \mathcal{V}, \quad (\mathbf{U}_L, \mathbf{B}_I) \mapsto X(\mathbf{U}_L, \mathbf{B}_I)$$

is multilinear—linear in each  $\mathbf{U}_\tau$  and each  $\mathbf{B}_\tau$ —and onto  $\mathcal{M}$ .

Note that by (2.10–2.11), the mapping  $\varphi$  is not invertible, but fortunately this will never pose any problems. For the tangent space  $\mathcal{T}_X \mathcal{M}$ , however, we follow the approach of [15, 16, 20, 12, 30] and work with unique representations for tangent vectors, given a specific  $(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{C}$ .

**THEOREM 3.1.** *Let  $(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{C}$  be an orthogonal HT decomposition of  $X = X(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{M}$ . Then the gauge space*

$$(3.2) \quad \mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)} := \bigtimes_{\tau \in T} \mathcal{G}_\tau,$$

with

$$(3.3) \quad \mathcal{G}_\tau = \{\delta \mathbf{B}_\tau \in \mathbb{R}^{(k_{\tau_1} k_{\tau_2}) \times k_\tau}\} \quad \text{for } \tau = \tau_r,$$

$$(3.4) \quad \mathcal{G}_\tau = \{\delta \mathbf{B}_\tau \in \mathbb{R}^{(k_{\tau_1} k_{\tau_2}) \times k_\tau} : \delta \mathbf{B}_\tau^T \mathbf{B}_\tau = \mathbf{0}\} \quad \text{for all } \tau \in I \setminus \{\tau_r\},$$

$$(3.5) \quad \mathcal{G}_\tau = \{\delta \mathbf{U}_\tau \in \mathbb{R}^{n_\tau \times k_\tau} : \delta \mathbf{U}_\tau^T \mathbf{U}_\tau = \mathbf{0}\} \quad \text{for all } \tau \in L,$$

is isomorphic to the tangent space  $\mathcal{T}_X \mathcal{M}$ .

Although a formal proof of this result was probably given for the first time in [30], let us mention that the gauge conditions (3.3)–(3.5) were already used in the quantum chemists community [32]; see also [20, p. 45] for a mathematical formulation. Note also that in the case of nonorthogonal HT decompositions, the gauge condition (3.4) has to be adapted accordingly; see [30, eq. (4.8)].

Elements of  $\mathcal{T}_X \mathcal{M}$  can be computed by differentiating  $\varphi$  of (3.1). By virtue of Theorem 3.1, the restriction of the domain of the differential  $D\varphi$  to the gauge space,

$$D\varphi(X)|_{\mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)}} : \mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)} \rightarrow \mathcal{T}_X \mathcal{M},$$

is a bijection at  $X(\mathbf{U}_L, \mathbf{B}_I)$  for *fixed*  $(\mathbf{U}_L, \mathbf{B}_I)$ . In the equations above, we have again used a similar short-hand notation,

$$(\delta \mathbf{U}_L, \delta \mathbf{B}_I) \in \mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)} \quad \text{with} \quad \delta \mathbf{U}_L := (\delta \mathbf{U}_\tau)_{\tau \in L} \quad \text{and} \quad \delta \mathbf{B}_I := (\delta \mathbf{B}_\tau)_{\tau \in I}.$$

Since  $\varphi$  is multilinear, its differential is conceptually straightforward to compute. Recursively differentiating (2.8) from leaves to root, one obtains matrices

$$(3.6) \quad (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \delta \mathbf{B}_\tau + (\delta \mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \mathbf{B}_\tau + (\mathbf{U}_{\tau_2} \otimes \delta \mathbf{U}_{\tau_1}) \mathbf{B}_\tau =: \delta \mathbf{U}_\tau \in \mathbb{R}^{n_\tau \times k_\tau}$$

such that  $\delta X \in \mathcal{T}_X \mathcal{M} \subset \mathcal{V}$  is recovered as  $\delta \mathbf{X}^{(\tau_r)} = \delta \mathbf{U}_{\tau_r}$ . The relation (3.6) shows that tangent tensors are highly structured and can be parameterized by the matrices  $(\delta \mathbf{U}_L, \delta \mathbf{B}_I)$ . Similar to the notation of an HT decomposition  $X(\mathbf{U}_L, \mathbf{B}_I)$ , we denote this as  $\delta X(\delta \mathbf{U}_L, \delta \mathbf{B}_I; \mathbf{U}_L, \mathbf{B}_I)$ .

In case  $(\delta \mathbf{U}_L, \delta \mathbf{B}_I) \in \mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)}$ , the matrices  $\delta \mathbf{U}_\tau$  from (3.6) additionally satisfy

$$(3.7) \quad \delta \mathbf{U}_\tau^T \mathbf{U}_\tau = \mathbf{0} \quad \text{for all } \tau \in T \setminus \{\tau_r\};$$

see [30, Cor. 4.13]. Observe that this extends the gauge condition (3.5) to the inner nodes, except the root.

In order to simplify notation, we will abbreviate  $\mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)}$  by  $\mathcal{G}$  and the tangent tensor  $\delta X(\delta \mathbf{U}_L, \delta \mathbf{B}_I; \mathbf{U}_L, \mathbf{B}_I)$  by  $\delta X(\delta \mathbf{U}_L, \delta \mathbf{B}_I)$  from now on, dropping the notational dependence on  $(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{C}$ .

**3.2. Dynamical HT approximation as projection.** We now turn to the main problem of this paper: The dynamical approximation  $Y(t) \in \mathcal{M}$  of a time-varying tensor  $A(t) \in \mathcal{V}$ .

The minimization condition in (1.2) can be formulated in terms of the orthogonal projector onto the tangent space, that is, for  $X \in \mathcal{M}$ , the operator

$$(3.8) \quad P_X: \mathcal{V} \rightarrow \mathcal{T}_X \mathcal{M}, \quad Z \mapsto P_X(Z)$$

satisfying

$$(3.9) \quad \langle P_X(Z) - Z, \delta V \rangle = 0 \quad \text{for all } \delta V \in \mathcal{T}_X \mathcal{M}.$$

With this operator, we can state the differential equations for  $Y(t)$  as

$$(3.10) \quad \dot{Y}(t) = P_{Y(t)} \dot{A}(t).$$

The next proposition shows how  $P_X$  can be computed. Although this result was already shown by construction in [30, section 6.2.2], our proof—which is the direct analogue of that in [15, 16]—is simpler. We postpone the discussion of how to actually implement  $P_X$  to section 7.

**PROPOSITION 3.2.** *Suppose  $X = X(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{M}$  and let  $Z \in \mathcal{V}$  be given. For all  $\tau = \{i_1, i_2, \dots, i_s\} \in T$ , recursively define symmetric and positive definite matrices  $\mathbf{M}_\tau \in \mathbb{R}^{k_\tau \times k_\tau}$  via*

$$(3.11) \quad \begin{aligned} \mathbf{M}_{\tau_r} &= 1, \\ \mathbf{M}_{\tau_1} &= \mathbf{B}_\tau^{(1)} (\mathbf{M}_\tau \otimes \mathbf{I}_{k_{\tau_2}}) (\mathbf{B}_\tau^{(1)})^T \end{aligned} \quad \text{for all } \tau \in I,$$

$$(3.12) \quad \mathbf{M}_{\tau_2} = \mathbf{B}_\tau^{(2)} (\mathbf{M}_\tau \otimes \mathbf{I}_{k_{\tau_1}}) (\mathbf{B}_\tau^{(2)})^T \quad \text{for all } \tau \in I,$$

and tensors  $Z_\tau \in \mathbb{R}^{n_{i_1} \times n_{i_2} \times \dots \times n_{i_s} \times k_\tau}$  via

$$(3.13) \quad \begin{aligned} \mathbf{Z}_{\tau_r} &= \mathbf{Z}^{(\tau_r)}, \\ \mathbf{Z}_{\tau_1} &= \mathbf{Z}_\tau^{(\tau_1)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_2}) (\mathbf{B}_\tau^{(1)})^T \mathbf{M}_{\tau_1}^{-1} \end{aligned} \quad \text{for all } \tau \in I,$$

$$(3.14) \quad \mathbf{Z}_{\tau_2} = \mathbf{Z}_\tau^{(\tau_2)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_1}) (\mathbf{B}_\tau^{(2)})^T \mathbf{M}_{\tau_2}^{-1} \quad \text{for all } \tau \in I.$$

Then, the projection  $P_X(Z) = \delta X(\delta \mathbf{U}_L, \delta \mathbf{B}_I)$  with  $(\delta \mathbf{U}_L, \delta \mathbf{B}_I) \in \mathcal{G}$  satisfies

$$(3.15) \quad \delta \mathbf{B}_\tau = (\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}_\tau^{(\tau)} \quad \text{for } \tau = \tau_r,$$

$$(3.16) \quad \delta \mathbf{B}_\tau = (\mathbf{I} - \mathbf{B}_\tau \mathbf{B}_\tau^T) (\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}_\tau^{(\tau)} \quad \text{for all } \tau \in I \setminus \{\tau_r\},$$

$$(3.17) \quad \delta \mathbf{U}_\tau = (\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T) \mathbf{Z}_\tau^{(\tau)} \quad \text{for all } \tau \in L.$$

*Proof.* (a) First, observe that because the  $B_\tau$  all have full multilinear rank, the  $\mathbf{M}_\tau$  are indeed symmetric and positive definite as stated.

Define next the images of the mappings (3.6):

$$\begin{aligned} \mathcal{U}_\tau &= \{ (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \delta \mathbf{B}_\tau + (\delta \mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \mathbf{B}_\tau + (\mathbf{U}_{\tau_2} \otimes \delta \mathbf{U}_{\tau_1}) \mathbf{B}_\tau \\ &\quad : \delta \mathbf{B}_\tau \in \mathcal{G}_\tau, \delta \mathbf{U}_{\tau_1} \in \mathcal{U}_{\tau_1}, \delta \mathbf{U}_{\tau_2} \in \mathcal{U}_{\tau_2} \} \quad \text{for all } \tau \in I, \\ \mathcal{U}_\tau &= \mathcal{G}_\tau \quad \text{for all } \tau \in L. \end{aligned}$$

Then, given the (unknown) parametrization  $(\delta \mathbf{U}_L, \delta \mathbf{B}_I)$  of the projection  $P_X(Z)$ , define matrices  $\delta \mathbf{U}_\tau \in \mathcal{U}_\tau$  for each  $\tau \in T$  as the result of evaluating (3.6). Below, we shall show that these  $\delta \mathbf{U}_\tau$  each satisfy their own Galerkin condition,

$$(3.18) \quad \langle \delta \mathbf{U}_\tau - \mathbf{Z}_\tau^{(\tau)}, \delta \mathbf{V} \mathbf{M}_\tau \rangle = 0 \quad \text{for all } \delta \mathbf{V} \in \mathcal{U}_\tau \text{ and for all } \tau \in T.$$

The statements about  $(\delta \mathbf{U}_L, \delta \mathbf{B}_I)$  then follow from these conditions.

(b) Let  $\tau = \tau_r$ . Since  $\delta X \in \mathcal{T}_X \mathcal{M}$ , we have by definition that  $\delta \mathbf{U}_{\tau_r} = \delta \mathbf{X}^{(\tau_r)}$ . Hence, the Galerkin condition (3.9) is equivalent to (3.18) since  $\mathbf{M}_{\tau_r} = 1$ .

To prove that (3.18) holds for the other nodes in the tree, we show that it holds for the two children of an arbitrary node  $\tau$  that is not a leaf. First, choose  $\delta \mathbf{V} = (\mathbf{U}_{\tau_2} \otimes \delta \mathbf{V}_{\tau_1}) \mathbf{B}_\tau \in \mathcal{U}_\tau$  with arbitrary  $\delta \mathbf{V}_{\tau_1} \in \mathcal{U}_{\tau_1}$  and observe that by (2.5) and the symmetry of  $\mathbf{M}_\tau$  we have

$$\mathbf{M}_\tau \delta \mathbf{V}^T = ((\delta \mathbf{V}_{\tau_1}, \mathbf{U}_{\tau_2}, \mathbf{M}_\tau) \circ B_\tau)^{(3)}.$$

Then, after unfolding in the first mode and exploiting (3.7), we get

$$\begin{aligned} \langle \delta \mathbf{U}_\tau, \delta \mathbf{V} \mathbf{M}_\tau \rangle &= \langle \delta \mathbf{U}_{\tau_1} \mathbf{B}_\tau^{(1)} (\mathbf{I}_{k_\tau} \otimes \mathbf{U}_{\tau_2})^T, \delta \mathbf{V}_{\tau_1} \mathbf{B}_\tau^{(1)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_2})^T \rangle \\ &= \langle \delta \mathbf{U}_{\tau_1} \mathbf{B}_\tau^{(1)} (\mathbf{M}_\tau \otimes \mathbf{I}_{k_{\tau_2}}) (\mathbf{B}_\tau^{(1)})^T, \delta \mathbf{V}_{\tau_1} \rangle, \\ \langle \mathbf{Z}_\tau^{(\tau)}, \delta \mathbf{V} \mathbf{M}_\tau \rangle &= \langle \mathbf{Z}_\tau^{(\tau_1)}, \delta \mathbf{V}_{\tau_1} \mathbf{B}_\tau^{(1)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_2})^T \rangle \\ &= \langle \mathbf{Z}_\tau^{(\tau_1)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_2}) (\mathbf{B}_\tau^{(1)})^T, \delta \mathbf{V}_{\tau_1} \rangle. \end{aligned}$$

Plugging in definition (3.11) for  $\mathbf{M}_{\tau_1}$ , condition (3.9) implies

$$\langle \delta \mathbf{U}_{\tau_1} - \mathbf{Z}_\tau^{(\tau_1)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_2}) (\mathbf{B}_\tau^{(1)})^T \mathbf{M}_{\tau_1}^{-1}, \delta \mathbf{V}_{\tau_1} \mathbf{M}_{\tau_1} \rangle = 0 \quad \text{for all } \delta \mathbf{V}_{\tau_1} \in \mathcal{U}_{\tau_1},$$

from which we obtain (3.18) for the node  $\tau_1$  using definition (3.13) for  $\mathbf{Z}_\tau^{(\tau_1)}$ .

A similar argument using  $\delta \mathbf{V} = (\delta \mathbf{V}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \mathbf{B}_\tau \in \mathcal{U}_\tau$  with arbitrary  $\delta \mathbf{V}_{\tau_2} \in \mathcal{U}_{\tau_2}$  and definition (3.12) for  $\mathbf{M}_{\tau_2}$  yields

$$\langle \delta \mathbf{U}_{\tau_2} - \mathbf{Z}_\tau^{(\tau_2)} (\mathbf{M}_\tau \otimes \mathbf{U}_{\tau_1}) (\mathbf{B}_\tau^{(2)})^T \mathbf{M}_{\tau_2}^{-1}, \delta \mathbf{V}_{\tau_2} \mathbf{M}_{\tau_2} \rangle = 0 \quad \text{for all } \delta \mathbf{V}_{\tau_2} \in \mathcal{U}_{\tau_2}.$$

The Galerkin condition (3.18) for node  $\tau_2$  is then obtained using definition (3.14).

(c) As the last step, we show the  $(\delta\mathbf{B}_I, \delta\mathbf{U}_L)$ . Let  $\tau \in I$ . Choosing  $\delta\mathbf{V} = (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \delta\mathbf{C}_\tau \in \mathcal{U}_\tau$  with arbitrary  $\delta\mathbf{C}_\tau \in \mathcal{G}_\tau$  and using (3.7), we get

$$\begin{aligned} \langle \delta\mathbf{U}_\tau, \delta\mathbf{V} \mathbf{M}_\tau \rangle &= \langle \delta\mathbf{B}_\tau, \delta\mathbf{C}_\tau \mathbf{M}_\tau \rangle, \\ \langle \mathbf{Z}_\tau^{(\tau)}, \delta\mathbf{V} \mathbf{M}_\tau \rangle &= \langle \mathbf{Z}_\tau^{(\tau)}, (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \delta\mathbf{C}_\tau \mathbf{M}_\tau \rangle = \langle (\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}_\tau^{(\tau)}, \delta\mathbf{C}_\tau \mathbf{M}_\tau \rangle. \end{aligned}$$

Since this holds for every  $\delta\mathbf{C}_\tau \in \mathcal{G}_\tau$ , condition (3.18) gives immediately (3.15) when  $\tau = \tau_r$ . In the case when  $\tau \neq \tau_r$ , observe that because  $\mathbf{M}_\tau$  is nonsingular we have that

$$\begin{aligned} \{\delta\mathbf{C}_\tau \mathbf{M}_\tau : \delta\mathbf{C}_\tau \in \mathcal{G}_\tau\} &= \{\delta\mathbf{C}_\tau \mathbf{M}_\tau : \delta\mathbf{C}_\tau^T \mathbf{B}_\tau = \mathbf{0}, \delta\mathbf{C}_\tau \in \mathbb{R}^{(k_{\tau_1} k_{\tau_2}) \times k_\tau}\} \\ &= \{\delta\mathbf{C}_\tau \mathbf{M}_\tau : (\delta\mathbf{C}_\tau \mathbf{M}_\tau)^T \mathbf{B}_\tau = \mathbf{0}, \delta\mathbf{C}_\tau \in \mathbb{R}^{(k_{\tau_1} k_{\tau_2}) \times k_\tau}\} = \mathcal{G}_\tau. \end{aligned}$$

Hence after imposing the gauge condition  $\mathbf{B}_\tau^T \delta\mathbf{B}_\tau = \mathbf{0}$ , we obtain (3.16) from (3.18).

Let  $\tau \in L$ . Then, by a similar argument as above, (3.18) implies

$$\langle \delta\mathbf{U}_\tau - \mathbf{Z}_\tau^{(\tau)}, \delta\mathbf{V} \mathbf{M}_\tau \rangle = \langle \delta\mathbf{U}_\tau - \mathbf{Z}_\tau^{(\tau)}, \delta\mathbf{V} \rangle = 0 \quad \text{for all } \delta\mathbf{V} \in \mathcal{G}_\tau,$$

which leads to (3.17) after having imposed the gauge  $\mathbf{U}_\tau^T \delta\mathbf{U}_\tau = \mathbf{0}$ .  $\square$

We note that Proposition 3.2 turns the differential equation (3.10) on  $\mathcal{M}$  into an equivalent set of differential equations for the coefficients in  $\mathcal{C}$ : with tensors  $Z_\tau$  defined as above with  $\dot{A}$  in the role of  $Z$ , we solve the system of differential equations in  $\mathcal{C}$ ,

$$(3.19) \quad \dot{\mathbf{B}}_\tau = (\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}_\tau^{(\tau)} \quad \text{for } \tau = \tau_r,$$

$$(3.20) \quad \dot{\mathbf{B}}_\tau = (\mathbf{I} - \mathbf{B}_\tau \mathbf{B}_\tau^T) (\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}_\tau^{(\tau)} \quad \text{for all } \tau \in I \setminus \{\tau_r\},$$

$$(3.21) \quad \dot{\mathbf{U}}_\tau = (\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T) \mathbf{Z}_\tau^{(\tau)} \quad \text{for all } \tau \in L,$$

to obtain  $Y(t) = X(\mathbf{U}_L(t), \mathbf{B}_I(t))$ . This is the system of differential equations that is actually solved numerically.

**4. Tangent space projection and curvature.** In this section, we estimate the curvature of the manifold by investigating how the orthogonal projection behaves along the manifold. This will be a key tool for the approximation results in the next section.

**4.1. A nonrecursive formulation of the orthogonal projector.** The recursive nature of Proposition 3.2 complicates our analysis involving the orthogonal projector  $P_X$ . We therefore introduce an equivalent but “global” expression that will facilitate the estimates by making the isomorphism between  $\mathcal{G}$  and  $\mathcal{T}_X \mathcal{M}$  explicit. To this end, we first need the following lemma which describes the node- $\tau$  unfoldings of an HT tensor.

LEMMA 4.1. *Let  $X(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{V}$  be a  $(T, \mathbf{k})$ -tensor. Then*

$$(4.1) \quad \text{span}(\mathbf{X}^{(\tau)}) = \text{span}(\mathbf{U}_\tau) \quad \text{for all } \tau \in T,$$

where the inner frame matrices  $(\mathbf{U}_\tau)_{\tau \in I}$  are given by (2.8). In addition, there are unique matrices  $\mathbf{R}_\tau \in \mathbb{R}^{k_\tau \times n_{\tau^c}}$  of full column rank  $k_\tau$  (with  $\mathbf{R}_{\tau_r} = \mathbf{I}$ ) such that

$$(4.2) \quad \mathbf{X}^{(\tau)} = \mathbf{U}_\tau \mathbf{R}_\tau.$$

*Proof.* Equality (4.1) follows from [30, Prop. 3.4, Prop. 3.6], while (4.2) is a straightforward consequence of (4.1).  $\square$

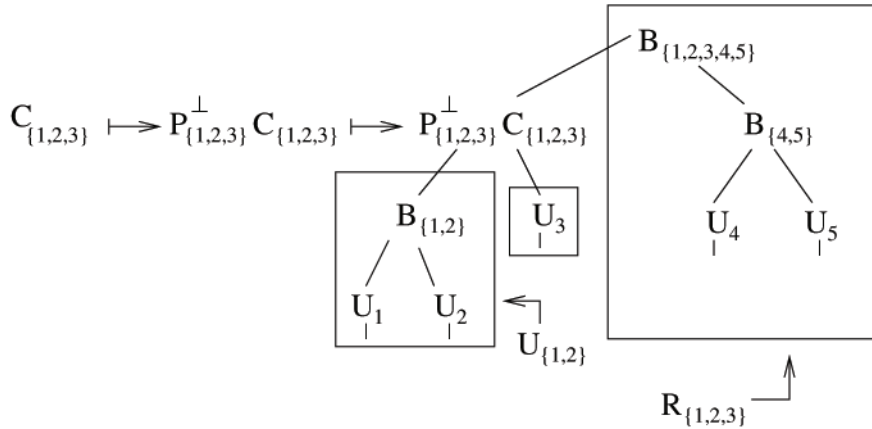


FIG. 4.1. Graphical representation of application of the embedding operator  $E_{\{1,2,3\}}$  to some tensor  $\mathbf{C}_{\{1,2,3\}} \in \mathbb{R}^{(k_1 k_2) \times k_3}$ . First, the projector  $\mathbf{P}_{\{1,2,3\}}^\perp := \mathbf{I} - \mathbf{B}_{\{1,2,3\}} \mathbf{B}_{\{1,2,3\}}^T$  is applied, then the result is embedded in the HT decomposition of  $X(\mathbf{U}_L, \mathbf{B}_I)$  of  $X$  in place of  $\mathbf{B}_{\{1,2,3\}}$  to obtain  $E_{\{1,2,3\}}(\mathbf{C}_{\{1,2,3\}}) \in \mathcal{V}$ .

The previous lemma allows us to define the following *embedding operators*.

DEFINITION 4.2. Let  $X(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{M}$  be an orthogonal HT tensor with the decomposition  $\mathbf{X}^{(\tau)} = \mathbf{U}_\tau \mathbf{R}_\tau$  as in Lemma 4.1. The node- $\tau$  embedding operators

$$\begin{aligned} E_\tau: \mathbb{R}^{(k_{\tau_1} k_{\tau_2}) \times k_\tau} &\rightarrow \mathcal{V}, \quad \mathbf{C}_\tau \mapsto E_\tau(\mathbf{C}_\tau) && \text{for } \tau \in I, \\ E_\tau: \mathbb{R}^{n_\tau \times k_\tau} &\rightarrow \mathcal{V}, \quad \mathbf{V}_\tau \mapsto E_\tau(\mathbf{V}_\tau) && \text{for } \tau \in L, \end{aligned}$$

are the linear operators defined via the unfoldings

$$(4.3) \quad (E_\tau(\mathbf{C}_\tau))^{(\tau)} = (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \mathbf{C}_\tau \quad \text{for } \tau = \tau_r,$$

$$(4.4) \quad (E_\tau(\mathbf{C}_\tau))^{(\tau)} = (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) (\mathbf{I} - \mathbf{B}_\tau \mathbf{B}_\tau^T) \mathbf{C}_\tau \mathbf{R}_\tau \quad \text{for } \tau \in I \setminus \{\tau_r\},$$

$$(4.5) \quad (E_\tau(\mathbf{V}_\tau))^{(\tau)} = (\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T) \mathbf{V}_\tau \mathbf{R}_\tau \quad \text{for } \tau \in L.$$

The definition above uses the matrices  $\mathbf{R}_\tau$  for convenience of notation only. Conceptually the action of  $E_\tau$  is very straightforward to compute. First, observe that the image of  $E_\tau(\cdot)$  is an HT tensor with the same dimension tree as  $X$  and with hierarchical rank bounded by that of  $X$ . This is easily seen from the form of the matrices in (4.2) together with the identity

$$(4.6) \quad \mathbf{X}^{(\tau)} = (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1}) \mathbf{B}_\tau \mathbf{R}_\tau \quad \text{for all } \tau \in I.$$

This shows that  $E_\tau(\mathbf{V}_\tau)$  is obtained by first applying the projector  $\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T$  to the component  $\mathbf{V}_\tau$  and then substituting the resulting matrix for  $\mathbf{U}_\tau$  in the HT decomposition  $X(\mathbf{U}_L, \mathbf{B}_I)$ . The case  $E_\tau(\mathbf{C}_\tau)$  is analogous except that one applies the projector  $\mathbf{I} - \mathbf{B}_\tau \mathbf{B}_\tau^T$  first (except for the root); see Figure 4.1 for a graphical representation.

Let  $F^+ : \mathcal{B} \rightarrow \mathcal{A}$  denote the Moore–Penrose pseudoinverse of a linear operator  $F : \mathcal{A} \rightarrow \mathcal{B}$ . Then, we immediately have from Lemma 4.1, equations (4.6) and (3.3)–(3.5), that

$$(4.7) \quad \text{range}(E_\tau) \perp \text{range}(E_\nu) \quad \text{and} \quad \text{range}(E_\tau^+) \perp \text{range}(E_\nu^+)$$

for all  $\tau \neq \nu \in T$ . We are now ready to give the alternative representation of the projector  $P_X$  of Prop. 3.2.

PROPOSITION 4.3. *Let  $X(\mathbf{U}_L, \mathbf{B}_I) \in \mathcal{M}$  be an orthogonal HT tensor with the decomposition  $\mathbf{X}^{(\tau)} = \mathbf{U}_\tau \mathbf{R}_\tau$  as in Lemma 4.1 and denote the accompanying gauge space  $\mathcal{G} := \mathcal{G}_{(\mathbf{U}_L, \mathbf{B}_I)}$ . Then, the following holds:*

(i) *The linear operator*

$$(4.8) \quad E: \mathcal{G} \rightarrow \mathcal{V}, (\delta \mathbf{U}_L, \delta \mathbf{B}_I) \mapsto \sum_{\tau \in L} E_\tau(\delta \mathbf{U}_\tau) + \sum_{\tau \in I} E_\tau(\delta \mathbf{B}_\tau)$$

*is an isomorphism between  $\mathcal{G}$  and the tangent space  $\mathcal{T}_X \mathcal{M}$ .*

(ii) *The Moore–Penrose inverses*

$$(4.9) \quad E^+ : \mathcal{V} \rightarrow \mathcal{G} \quad \text{and} \quad E_\tau^+ : \mathcal{V} \rightarrow \mathcal{G}_\tau$$

*are left inverses of  $E$  and  $E_\tau$  for each  $\tau \in T$ , respectively.*

(iii) *The orthogonal projector  $P_X : \mathcal{V} \rightarrow \mathcal{T}_X \mathcal{M}$  satisfies*

$$P_X = EE^+ = \sum_{\tau \in T} E_\tau E_\tau^+.$$

*For any  $Z \in \mathcal{V}$ , this becomes*

$$P_X(Z) = \sum_{\tau \in T} P_X^\tau(Z)$$

*with*

$$\begin{aligned} (P_X^\tau(Z))^{(\tau)} &= (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1})(\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}^{(\tau)} \mathbf{R}_\tau^+ \mathbf{R}_\tau && \text{for } \tau = \tau_r, \\ (P_X^\tau(Z))^{(\tau)} &= (\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1})(\mathbf{I} - \mathbf{B}_\tau \mathbf{B}_\tau^T)(\mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1}^T) \mathbf{Z}^{(\tau)} \mathbf{R}_\tau^+ \mathbf{R}_\tau \\ &= (\mathbf{U}_{\tau_2} \mathbf{U}_{\tau_2}^T \otimes \mathbf{U}_{\tau_1} \mathbf{U}_{\tau_1}^T - \mathbf{U}_\tau \mathbf{U}_\tau^T) \mathbf{Z}^{(\tau)} \mathbf{R}_\tau^+ \mathbf{R}_\tau && \text{for } \tau \in I \setminus \tau_r, \\ (P_X^\tau(Z))^{(\tau)} &= (\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T) \mathbf{Z}^{(\tau)} \mathbf{R}_\tau^+ \mathbf{R}_\tau && \text{for } \tau \in L. \end{aligned}$$

*Proof.* (i) Observe that the range of  $E$  lies in  $\mathcal{T}_X \mathcal{M}$ , as follows, for example, by taking in the representations (4.2) and (4.6) for  $\mathbf{X}^{(\tau)}$  the directional derivatives with respect to the components  $\mathbf{U}_\tau, \mathbf{B}_\tau$ . Second,  $E$  is injective, which, using (4.7), can be shown by proving that every  $E_\tau$  is injective as mapping from  $\mathcal{G}_\tau$  to  $\mathcal{V}$ : Indeed, using the definitions (4.5)–(4.3), this follows from the (pseudo-)invertibility of  $\mathbf{R}_\tau$  and  $(\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1})$  by  $\mathbf{R}_\tau^+$ , respectively  $(\mathbf{U}_{\tau_2} \otimes \mathbf{U}_{\tau_1})^T$ , and the fact that projectors  $\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T$  and  $\mathbf{I} - \mathbf{B}_\tau \mathbf{B}_\tau^T$  have no effect on elements from  $\mathcal{G}_\tau$ . By the isometry in Theorem 3.1, we have  $\dim \mathcal{T}_X \mathcal{M} = \dim \mathcal{G}$ , which altogether proves the claim.

(ii) From (i) we have that  $E$  and  $E_\tau$  are of full rank  $\dim(\mathcal{G})$  and  $\dim(\mathcal{G}_\tau)$ , respectively.

(iii) For every matrix  $\mathbf{A}$ ,  $\mathbf{A}\mathbf{A}^+$  projects onto the image of  $\mathbf{A}$ ; see, e.g., [29]. The second representation follows from (4.7). The formulas for  $P_X^\tau(Z)$  follow from basic manipulation of (4.5)–(4.3) together with (2.8).  $\square$

**4.2. Curvature bounds.** We first need a preparatory lemma.

LEMMA 4.4. *Let  $Y : [a, b] \rightarrow \mathcal{M}$  be a smooth curve on  $\mathcal{M}$ , and let  $(\mathbf{U}_L(a), \mathbf{B}_I(a))$  be an orthogonal HT decomposition of  $Y(a)$ . Then, there exists a unique smooth curve  $\gamma : [a, b] \rightarrow \mathcal{C}$ ,  $t \mapsto (\mathbf{U}_L(t), \mathbf{B}_I(t))$  such that  $(\mathbf{U}_L(t), \mathbf{B}_I(t))$  is an orthogonal HT*

decomposition of  $Y(t)$  and  $t \mapsto \dot{\gamma}(t) = (\dot{\mathbf{U}}_L(t), \dot{\mathbf{B}}_I(t)) \in \mathcal{G}_{(\mathbf{U}_L(t), \mathbf{B}_I(t))}$  is a smooth gauged parameterization of  $\dot{Y}(t)$ .

*Proof.* Suppose that for some  $t \geq a$ ,  $(\mathbf{U}_L(t), \mathbf{B}_I(t))$  is an orthogonal HT decomposition of  $Y(t)$ . Proposition 3.2 then gives us a gauged parameterization

$$(4.10) \quad \dot{Y}(t) = P_{Y(t)} \dot{Y}(t) = \delta X(\dot{\mathbf{U}}_L(t), \dot{\mathbf{B}}_I(t); \mathbf{U}_L(t), \mathbf{B}_I(t)),$$

where  $(\dot{\mathbf{U}}_L(t), \dot{\mathbf{B}}_I(t)) \in \mathcal{G}_{(\mathbf{U}_L(t), \mathbf{B}_I(t))}$  are constructed in a smooth way from  $\mathbf{U}_L(t), \mathbf{B}_I(t)$  and the given  $\dot{Y}(t)$ . We thus have a differential equation

$$(4.11) \quad (\dot{\mathbf{U}}_L(t), \dot{\mathbf{B}}_I(t)) = F(\mathbf{U}_L(t), \mathbf{B}_I(t), t)$$

with a smooth function  $F$ . Since the HT-rank of  $Y(t)$  is assumed constant on  $[a, b]$ , the pseudoinverses in the construction of  $F$  by Proposition 3.2 are uniformly bounded, and therefore,  $F$  is Lipschitz-continuous in a neighborhood of the solution curve as far as it exists. Standard ODE theory (the Picard–Lindelöf theorem) therefore yields that a unique solution  $(\mathbf{U}_L(t), \mathbf{B}_I(t))$  of (4.11) with the given initial value  $(\mathbf{U}_L(a), \mathbf{B}_I(a))$  exists over the whole interval  $[a, b]$ . By construction we then have (4.10) for all  $t \in [a, b]$ , so that  $Y(t)$  and  $X(\mathbf{U}_L(t), \mathbf{B}_I(t))$  have the same derivative. They also have the same initial value, and hence they are equal:  $Y(t) = X(\mathbf{U}_L(t), \mathbf{B}_I(t))$ .  $\square$

Now we can estimate the curvature of  $\mathcal{M}$ .

LEMMA 4.5. *Let  $X \in \mathcal{M}$  be such that the smallest singular value of each  $\tau$ -unfolding is uniformly bounded below by some constant  $\rho > 0$ ,*

$$\sigma_{\min}(\mathbf{X}^{(\tau)}) \geq \rho > 0 \quad \text{for all } \tau \in T.$$

*Then, there exist constants  $c$  and  $C$ , depending only on the dimension  $d$  and satisfying  $cC \leq \frac{1}{2}$ , such that*

$$(4.12) \quad \|P_{\tilde{X}}(Z) - P_X(Z)\| \leq C \rho^{-1} \|\tilde{X} - X\| \cdot \|Z\|,$$

$$(4.13) \quad \|(I - P_X)(\tilde{X} - X)\| \leq 2C \rho^{-1} \|\tilde{X} - X\|^2$$

for all  $\tilde{X} \in \mathcal{M}$  with  $\|\tilde{X} - X\| \leq c\rho$  and all  $Z \in \mathcal{V}$ . In particular,  $C \leq 48d$ .

*Proof.* (a) By Lemma 4.1, we have  $\mathbf{X}^{(\tau)} = \mathbf{U}_\tau \mathbf{R}_\tau$  for all  $\tau \in T$  since  $X \in \mathcal{M}$ . Hence, the assumption that  $\sigma_{\min}(\mathbf{X}^{(\tau)}) = \sigma_{\min}(\mathbf{R}_\tau) \geq \rho$  implies

$$\|\mathbf{R}_\tau^+\|_2 = \sigma_{\min}^{-1}(\mathbf{R}_\tau) \leq \rho^{-1} \quad \text{for all } \tau \in T.$$

By [28, Cor. I.4.31] we also have that for any  $\hat{X} \in \mathcal{M}$  with  $\hat{\mathbf{X}}^{(\tau)} = \hat{\mathbf{U}}_\tau \hat{\mathbf{R}}_\tau$  it holds that

$$|\sigma_{\min}(\hat{\mathbf{X}}^{(\tau)}) - \sigma_{\min}(\mathbf{X}^{(\tau)})| \leq \|\hat{\mathbf{X}}^{(\tau)} - \mathbf{X}^{(\tau)}\|_2 \leq \|\hat{X} - X\|.$$

When  $\|\hat{X} - X\| \leq \frac{1}{2}\rho$ , this results in  $\sigma_{\min}(\hat{\mathbf{X}}^{(\tau)}) \geq \frac{1}{2}\rho$ , which implies

$$\|\hat{\mathbf{R}}_\tau^+\|_2 \leq 2\rho^{-1} \quad \text{for all } \tau \in T.$$

(b) We decompose tensors in  $\mathcal{V}$  on a straight line connecting  $X$  and  $\tilde{X}$  as

$$(4.14) \quad X + t(\tilde{X} - X) = Y(t) + Y_\perp(t) \quad \text{with } Y(t) \in \mathcal{M} \quad \text{and} \quad Y_\perp(t) \perp \mathcal{T}_X \mathcal{M}.$$

It will be shown in (e) that this decomposition exists for  $0 \leq t \leq 1$  under the given assumptions. We denote

$$\Delta = P_X(\tilde{X} - X) \in \mathcal{T}_X \mathcal{M} \quad \text{with} \quad \|\Delta\| \leq \delta := \|\tilde{X} - X\|.$$

We then have  $P_X(Y(t) - X) = t\Delta$ , which yields

$$(4.15) \quad P_X(\dot{Y}(t)) = \Delta.$$

Since  $\dot{Y} \in \mathcal{T}_Y \mathcal{M}$ , we have  $P_Y \dot{Y} = \dot{Y}$  and therefore

$$(4.16) \quad \dot{Y}(t) = P_{Y(t)}(\Delta + (P_{Y(t)} - P_X)(\dot{Y}(t))).$$

As long as

$$(4.17) \quad \|P_{Y(t)} - P_X\|_{\text{op}} := \max_{Z \in \mathcal{V}, \|Z\|=1} \|(P_{Y(t)} - P_X)(Z)\| \leq \frac{1}{2},$$

we then get

$$\|\dot{Y}\| \leq 2\delta \quad \text{and hence} \quad \|Y(t) - X\| \leq 2\delta t.$$

(c) Let  $Y(t) = \varphi(\mathbf{U}_L(t), \mathbf{B}_I(t))$  be the smooth curve on  $0 \leq t \leq 1$  of Lemma 4.4. Then the curves  $\mathbf{U}_\tau(t)$  and  $\mathbf{R}_\tau(t)$  satisfying  $\mathbf{Y}^{(\tau)}(t) = \mathbf{U}_\tau(t)\mathbf{R}_\tau(t)$  are smooth too for all  $\tau \in T$ . From (iii) in Proposition 4.3, we have an explicit expression for  $P_{Y(t)}(Z)$ . Hence, a bound on  $\dot{P}_{Y(t)}(Z) = \frac{d}{dt}P_{Y(t)}(Z)$  may be obtained by bounding  $\dot{\mathbf{U}}_\tau$  and  $\frac{d}{dt}(\mathbf{R}_\tau^+ \mathbf{R}_\tau)$  first. If  $2\delta \leq \frac{1}{2}\rho$ , then the argument in (a) applied to  $Y(t)$  instead of  $\hat{X}$  shows that

$$\|\mathbf{R}_\tau^+(t)\|_2 \leq 2\rho^{-1}.$$

Next, by Lemma 4.4, it holds that

$$\dot{\mathbf{Y}}^{(\tau)} = \dot{\mathbf{U}}_\tau \mathbf{R}_\tau + \mathbf{U}_\tau \dot{\mathbf{R}}_\tau \quad \text{such that} \quad \dot{\mathbf{U}}_\tau^T \mathbf{U}_\tau = \mathbf{0}.$$

Hence, we get with (2.4) that

$$\|\dot{\mathbf{R}}_\tau\| = \|\mathbf{U}_\tau^T \dot{\mathbf{Y}}^{(\tau)}\| \leq 2\delta \quad \text{and} \quad \|\dot{\mathbf{U}}_\tau \mathbf{R}_\tau\| = \|(\mathbf{I} - \mathbf{U}_\tau \mathbf{U}_\tau^T) \dot{\mathbf{Y}}^{(\tau)}\| \leq 2\delta$$

and further

$$\|\dot{\mathbf{U}}_\tau\| \leq \|\dot{\mathbf{U}}_\tau \mathbf{R}_\tau\| \cdot \|\mathbf{R}_\tau^+\|_2 \leq 4\rho^{-1}\delta.$$

Working out  $\frac{d}{dt}(\mathbf{R}_\tau^T (\mathbf{R}_\tau \mathbf{R}_\tau^T)^{-1} \mathbf{R}_\tau)$  using the identity  $\frac{d}{dt} \mathbf{A}^{-1} = -\mathbf{A}^{-1} \dot{\mathbf{A}} \mathbf{A}^{-1}$ , we further obtain

$$\|\frac{d}{dt}(\mathbf{R}_\tau^+ \mathbf{R}_\tau)\| \leq 2\|\mathbf{R}_\tau^+ \dot{\mathbf{R}}_\tau (\mathbf{I} - \mathbf{R}_\tau^+ \mathbf{R}_\tau)\| \leq 2\|\mathbf{R}_\tau^+\|_2 \cdot \|\dot{\mathbf{R}}_\tau\| \leq 8\rho^{-1}\delta,$$

where we used that  $\mathbf{I} - \mathbf{R}_\tau^+ \mathbf{R}_\tau$  is an orthogonal projector. Now differentiating  $P_{Y(t)}^\tau(Z)$  and using the above estimates, we obtain

$$(4.18) \quad \|\dot{P}_Y^\tau(Z)\| \leq 34\rho^{-1}\delta \|Z\| \quad \text{for all } \tau \in I \quad \text{and} \quad \|\dot{P}_Y^\tau(Z)\| \leq 16\rho^{-1}\delta \|Z\| \quad \text{for all } \tau \in L.$$



Using (2.7), this results in

$$\|\dot{P}_{Y(t)}(Z)\| \leq \sum_{\tau \in T} \|\dot{P}_{Y(t)}^{\tau}(Z)\| \leq 50\rho^{-1}\delta d \|Z\|.$$

Expressing  $P_{Y(t)}(Z) - P_{Y(0)}(Z) = \int_0^t \dot{P}_{Y(s)}(Z) ds$  then yields

$$\|P_{Y(t)}(Z) - P_X(Z)\| \leq 50\rho^{-1}\delta dt \|Z\|.$$

The operator norm in (4.17) thus does not exceed  $\frac{1}{2}$  for  $0 \leq t \leq 1$  when

$$\delta \leq c\rho \quad \text{with} \quad c = \frac{1}{100d},$$

and at  $t = 1$  we obtain the bound (4.12) with  $C = 50d$ .

(d) Observe that

$$(I - P_X)(\tilde{X} - X) = (I - P_X) \int_0^1 \dot{Y}(s) ds = \int_0^1 (P_{Y(s)} - P_X) \dot{Y}(s) ds.$$

By the above estimates, we have  $\|\dot{Y}\| \leq 2\delta$ ,  $\|Y(s) - X\| \leq 2\delta s$ ; thus,

$$\left\| \int_0^1 (P_{Y(s)} - P_X) \dot{Y}(s) ds \right\| \leq \int_0^1 \|(P_{Y(s)} - P_X)\| \|\dot{Y}(s)\| ds \leq 2C\rho^{-1}\delta^2,$$

which yields (4.13).

(e) It remains to show that the decomposition (4.14) indeed exists up to  $t = 1$ . For this we consider (4.16) as an implicit ordinary differential equation on the manifold  $\mathcal{M}$ , which under condition (4.17) can be turned into an explicit differential equation  $\dot{Y} = F(Y)$  with a smooth vector field  $F$ . The solution  $Y(t) \in \mathcal{M}$  with initial value  $X$  exists as long as (4.17) remains satisfied; hence for  $0 \leq t \leq 1$  by the above estimates. By construction,  $Y(t)$  lies on  $\mathcal{M}$  and satisfies (4.15), which upon integration implies (4.14).  $\square$

*Remark 2.* For the TT format, the estimate for the constant  $C$  in Lemma 4.5 can be improved to  $C \leq 32d$ : According to Remark 1, the projector  $P_{Y(t)}$  for  $t \in [0, 1]$  can be written in terms of a TT-decomposition with  $\mathbf{U}_i = \mathbf{I}_{n_i}$  for all  $i = 1, \dots, d-1$ ; thus, we have in (4.18) that  $\dot{P}_{Y(t)}^{\tau} = 0$  for the leaves  $\tau = \{i\}$ ,  $i = 1, \dots, d-1$ .

**5. Approximation properties.** We give approximation results that are analogous to those of [15] and [16] for the matrix case and full Tucker tensor format case, respectively. We refer the reader to these papers for a detailed discussion of these approximation results. The proofs are the same as for the corresponding results in [15] and [16], using Lemma 4.5.

**THEOREM 5.1.** *Suppose that*

$$(5.1) \quad \|\dot{A}(t)\| \leq \mu \quad \text{for } 0 \leq t \leq \bar{t}$$

*and that a continuously differentiable best approximation  $X(t) \in \mathcal{M}$  to  $A(t)$  exists for  $0 \leq t \leq \bar{t}$ . Let  $\rho > 0$  be such that the smallest nonzero singular value of every unfolding of  $X(t)$  satisfies  $\sigma_{\min}(\mathbf{X}^{(\tau)}(t)) \geq \rho$  for all  $\tau \in T$ , and assume that the best-approximation error is bounded by  $\|X(t) - A(t)\| \leq c\rho$  for  $0 \leq t \leq \bar{t}$ , with  $c$  of*

Lemma 4.5. Then, the approximation error of the dynamical low-rank approximation (1.2) with initial value  $Y(0) = X(0)$  is bounded by

$$\|Y(t) - X(t)\| \leq 2\beta e^{\beta t} \int_0^t \|X(s) - A(s)\| ds \quad \text{with } \beta = C\mu\rho^{-1}$$

for  $t \leq \bar{t}$  and as long as the right-hand side remains bounded by  $c\rho$ . Here,  $c$  and  $C$  are the constants of Lemma 4.5.

Smaller errors over longer time intervals are obtained not only if  $X - A$ , but also if its derivative is small. We assume that  $A(t)$  is of the form

$$(5.2) \quad A(t) = X(t) + E(t), \quad 0 \leq t \leq \bar{t},$$

where  $X(t) \in \mathcal{M}$  (now this need not necessarily be the best approximation) with

$$(5.3) \quad \|\dot{X}(t)\| \leq \mu$$

and the derivative of the remainder term is bounded by

$$(5.4) \quad \|\dot{E}(t)\| \leq \varepsilon$$

with a small  $\varepsilon > 0$ .

THEOREM 5.2. In addition to the above assumptions, suppose that the smallest singular values of the  $\tau$ -unfoldings of  $X(t)$  are bounded from below by  $\rho > 0$ . Then, the approximation error of (1.2) with initial value  $Y(0) = X(0)$  is bounded by

$$\|Y(t) - X(t)\| \leq 2t\varepsilon \quad \text{for } t \leq \frac{\rho}{C\sqrt{\mu\varepsilon}},$$

provided that  $t \leq \frac{\rho}{2\varepsilon}$  and  $t \leq \bar{t}$ . The constants  $c$  and  $C$  are those of Lemma 4.5.

For the HT approximation to a solution of the tensor differential equation

$$(5.5) \quad \dot{A} = F(A),$$

condition (1.2) is replaced, at every time  $t$ , by

$$(5.6) \quad \dot{Y} \in \mathcal{T}_Y \mathcal{M} \quad \text{such that} \quad \|\dot{Y} - F(Y)\| = \min.$$

This is equivalent to the Galerkin condition

$$(5.7) \quad \langle \dot{Y} - F(Y), \delta Y \rangle = 0 \quad \text{for all } \delta Y \in \mathcal{T}_Y \mathcal{M}.$$

We formulate an extension of Theorem 5.1 to the low-rank approximation of tensor differential equations (5.5). We assume that  $F$  has a moderate bound along the approximations,

$$(5.8) \quad \|F(X(t))\| \leq \mu, \quad \|F(Y(t))\| \leq \mu \quad \text{for } 0 \leq t \leq \bar{t},$$

and satisfies a one-sided Lipschitz condition: There is a real  $\lambda$  (positive or negative or zero) such that

$$(5.9) \quad \langle F(Y) - F(X), Y - X \rangle \leq \lambda \|Y - X\|^2$$

for all tensors  $X, Y \in \mathcal{M}$ . We further assume that for the best approximation  $X(t) \in \mathcal{M}$ ,

$$(5.10) \quad \|F(X(t)) - F(A(t))\| \leq L \|X(t) - A(t)\| \quad \text{for } 0 \leq t \leq \bar{t}.$$

We then have the following extension of the quasi-optimality result of Theorem 5.1.

**THEOREM 5.3.** *Suppose that a continuously differentiable best approximation  $X(t) \in \mathcal{M}$  to a solution  $A(t)$  of (5.5) exists for  $0 \leq t \leq \bar{t}$ , and assume the bounds (5.8)–(5.10). Let  $X(t)$  be such that the smallest singular value of each  $\tau$ -unfolding is uniformly bounded below by  $\rho > 0$ , and assume that the best-approximation error is bounded by  $\|X(t) - A(t)\| \leq c\rho$  with  $c$  of Lemma 4.5, for  $0 \leq t \leq \bar{t}$ . Then, the approximation error of (5.7) with initial value  $Y(0) = X(0)$  is bounded by*

$$\|Y(t) - X(t)\| \leq (2\beta + L) e^{(5\beta + \lambda)t} \int_0^t \|X(s) - A(s)\| ds \quad \text{with } \beta = C\mu\rho^{-1}$$

for  $t \leq \bar{t}$  and as long as the right-hand side is bounded by  $c\rho$ . The constants  $c$  and  $C$  are those of Lemma 4.5.

**6. Use of the dynamical approximation in iterative methods.** In iterative methods for optimization (e.g., nonlinear CG and Newton's method) one faces in every iteration the task of updating the iterate from  $A \in \mathcal{M}$  to an approximation  $\hat{A}$  to  $A + \Delta A$  that should again lie in  $\mathcal{M}$ . This problem is usually referred to as a truncation of  $A + \Delta A$  or, if  $\Delta A$  is a tangent tensor, as a retraction [1]. A popular approach to obtaining  $\hat{A}$  is to first compute  $A + \Delta A$ , which is typically not in  $\mathcal{M}$ , and then to project this sum onto  $\mathcal{M}$ . This is a nonlinear process that can be computationally expensive since the manifold  $\mathcal{M}$  is left in the intermediate result  $A + \Delta A$ . Instead, we propose using the dynamical approximation for  $A + t\Delta A$ . This works entirely on  $\mathcal{M}$ . Here we solve numerically the differential equation

$$(6.1) \quad \dot{Y} = P_Y(\Delta A), \quad Y(0) = A,$$

or rather the equivalent differential equations (3.19)–(3.21) for the HT parameters  $(\mathbf{U}_L(t), \mathbf{B}(t))$ , and set

$$\hat{A} = Y(1) \in \mathcal{M}.$$

An accurate computation of (3.19)–(3.21) over the whole interval  $[0, 1]$  may turn out costly when the increment  $\Delta A$  is not small in norm, but for small  $\|\Delta A\|$  (near the optimum) it can be approximated accurately by just a few explicit Euler steps applied to the differential equations (3.19)–(3.21) for the parametrization  $(\mathbf{U}_L(t), \mathbf{B}(t))$ , followed by reorthogonalization of the (small) coefficient matrices. This approach does not require any decompositions of large matrices.

Consider now the situation where  $\Delta A \in \mathcal{T}_A \mathcal{M}$  is a descent direction for a cost function. In a line search algorithm one would search for an (approximate) minimum of the cost function along the numerical trajectory  $Y(t)$ . A simplification to solving (6.1) consists of taking just one (or a few) explicit Euler steps with a step size  $h$  applied to the differential equations (3.19)–(3.21) for the parametrization  $(\mathbf{U}_L(t), \mathbf{B}(t))$  with subsequent reorthogonalization. The step size  $h > 0$  can be chosen adaptively such that descent is still guaranteed, since asymptotically for  $h \rightarrow 0$  the Euler step follows the descent direction  $\Delta A$ .

**7. Implementation and numerical results.** In this section, we detail numerical aspects of the dynamical approach; in particular, we focus on the implementation of the projector  $P_X(Z)$  and how to exploit structure in  $Z$ . Then, we report on some numerical experiments showing that the dynamical approach confirms the derived theoretical properties. Throughout this section, we define  $n := \max_{\tau \in L} \{n_\tau\}$  and denote the maximal hierarchical rank of  $X$  by  $k_X$ , and when  $Z$  is in HT format the analogue by  $k_Z$ .

In the implementation, we make heavy use of the tensor toolbox [2] and the `htucker` toolbox [19]. Our code is available at <http://web.math.princeton.edu/~bartv/>. All experiments were done with MATLAB version R2012a on an Intel Core i7 2.2 GHz CPU. In all examples, the integration of (1.2) was done by `ode45` in MATLAB using `odeset('RelTol', 1e-9, 'AbsTol', 1e-9, 'NormControl', 'on')`.

**7.1. Implementation of  $P_X(Z)$  for unstructured  $Z$ .** The implementation of  $P_X$  follows directly from Proposition 3.2 and consists of two phases: First, computing the  $\mathbf{M}_\tau$  and  $Z_\tau$ , after which the  $\delta\mathbf{U}_L$  and  $\delta\mathbf{B}_I$  can be projected based on (3.15)–(3.17). We emphasize that the actual computations are performed without working out the tensor products explicitly. Instead, each expression can be formulated in terms of tensor-times-matrix products, implemented as `ttm` in the `htucker` toolbox.

The following code, for example, constructs matrices `M{t1}` and `Z_matrix{t1}` representing  $\mathbf{M}_{\tau_1}$  and  $\mathbf{Z}_{\tau_1}^{(\tau_1)}$ , respectively, for  $\tau$  not a leaf.

---

```

B1_t = matricize(X.B{t},1);
B_t_M_t = ttm(X.B{t}, M{t}, 3);
M{t1} = B1_t * matricize(B_t_M_t,1)';

Z_t_tensor = dematricize(Z_matrix{t}, [n_t1, n_t2, k_t], [1 2], 3);
Z_t1_matrix = ttm(Z_t_tensor, {U_matrix{t2}, M{t}}, [2 3], 't');
Z_matrix{t1} = matricize(Z_t1_matrix, 1) * (B1_t'/M{t1});

```

---

Here, `X` is an `htensor` object representing the HT tensor  $X$  with a dimension tree indexed by  $\tau = \mathbf{t}$ . In addition, `U_matrix{t}` represents the inner frame matrix  $\mathbf{U}_\tau$  of  $X$  which can be recursively evaluated from the leaves up by the following commands.

---

```

BUU = ttm(X.B{t}, {U_matrix{t1}, U_matrix{t2}}, [1 2]);
U_matrix{t} = matricize(BUU, [1 2], 3);

```

---

The other computations are analogous and we refer the reader to the provided source code for details.

The tensors  $Z_\tau$  are reducing in size from root to leaves. Hence, the total computational cost of applying  $P_X(Z)$  is  $O(n^d)$ . This exponential growth in  $d$  severely limits the use of  $P_X(Z)$ , which makes its application only viable for small values of  $d$ . Nevertheless, it is an order of magnitude cheaper than applying the truncation operator based on successive SVDs, which is  $O(dn^{d+1})$ ; see [7]. For numerical verification, we refer the reader to [30, section 6.2.3] where the dynamical approach is indeed shown to be faster than pointwise SVDs in the case of unstructured  $Z$ .

**7.2. Implementation of  $P_X(Z)$  for structured  $Z$ .** The exponential dependence on  $d$  in the unstructured case from above can be overcome by assuming that  $Z$  is an HT tensor. This is a reasonable assumption exploited in many tensor-based

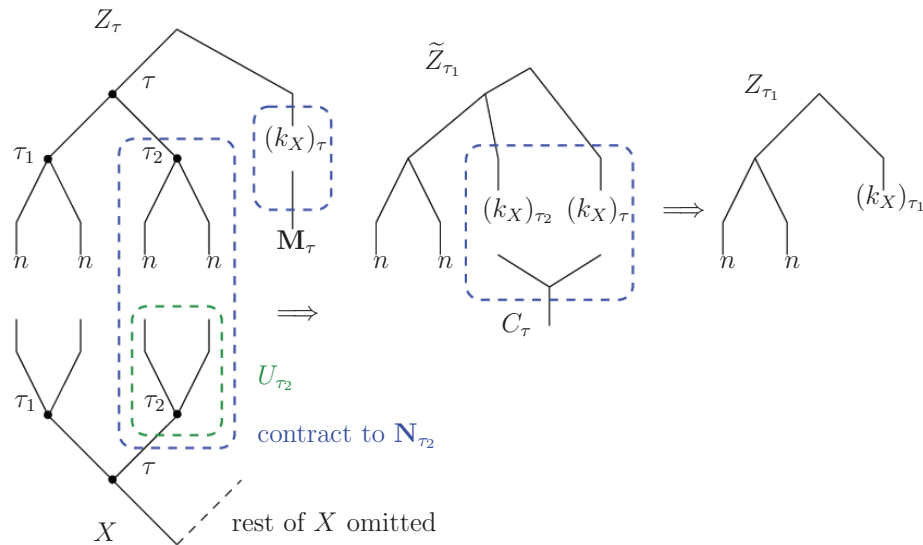


FIG. 7.1. Graphical representation of the recursion (3.13) of Proposition 3.2 in case  $Z$  is an HT tensor. Tensor  $\tilde{Z}_{\tau_1}$  is the result of the first contraction  $\tilde{\mathbf{Z}}_{\tau_1}^{(\tau_1)} = \mathbf{Z}_{\tau}^{(\tau_1)}(\mathbf{M}_{\tau} \otimes \mathbf{U}_{\tau_2})$ . The last contraction gives the desired result  $\mathbf{Z}_{\tau_1}^{(\tau_1)} = \tilde{\mathbf{Z}}_{\tau_1}^{(\tau_1)}(\mathbf{B}_{\tau}^{(1)})^T \mathbf{M}_{\tau_1}^{-1}$ .

methods like [3, 11, 13, 18]. In order to facilitate computations, we assume that  $Z$  has the same dimension tree as  $X$ , but it may have different hierarchical ranks.

The key to scalability is formulating the multiplications by  $\mathbf{U}_{\tau}$  in the recursions of Proposition 3.2 as tensor contractions of certain subtrees of  $X$  and  $Z_{\tau}$ . We refrain from giving a complete derivation and instead only explain the implementation of (3.13) that computes  $Z_{\tau_1}$ .

By construction, we impose that each  $Z_{\tau}$  has a dimension tree which is in a specific way compatible to that of  $X$ ; see Figure 7.1. In the first step, the subtree at  $\tau_2$  of  $Z_{\tau}$  is contracted with  $\mathbf{U}_{\tau_2}$  so as to replace the transfer tensor  $\mathbf{B}_{\tau_2}$  with a  $(k_Z)_{\tau_2} \times (k_X)_{\tau_2}$  dimensional matrix  $\mathbf{N}_{\tau_2}$ . (It turns out that all these matrices  $\mathbf{N}_{\tau}$  are computed when evaluating the inner product  $\langle X, Z \rangle$  which is a standard operation for tensor networks.) In addition, the last mode of  $Z_{\tau}$  is contracted with the matrix  $\mathbf{M}_{\tau}$ , which amounts to a simple matrix multiplication. Finally, the resulting tensor is contracted again with the small tensor  $C_{\tau}$  that represents an unfolding of  $(\mathbf{B}_{\tau}^{(1)})^T \mathbf{M}_{\tau_1}^{-1}$ .

The remaining computations follow a similar pattern; counting operations, we obtain that the work for

$$P_X(Z) \text{ is } O(dn(k_X^2 + k_X k_Z) + d(k_X^4 + k_X k_Z^3)).$$

For comparison, the work for

$$\text{hierarchical SVD of } X + Z \text{ is } O(dn(k_X^2 + k_X k_Z + k_Z^2) + d(k_X^4 + k_Z^4)).$$

Clearly, the projection is faster when  $k_Z \gg k_X$ .

In Figure 7.2, we have supplied timing results of the application of  $P_X(Z)$  for random tensors  $X$  and  $Z$  which verify the computational cost estimates from above. In all cases the ranks were taken equal along the tree. In addition, we have also supplied the results for the hierarchical SVD of the sum  $X + Z$  as implemented by

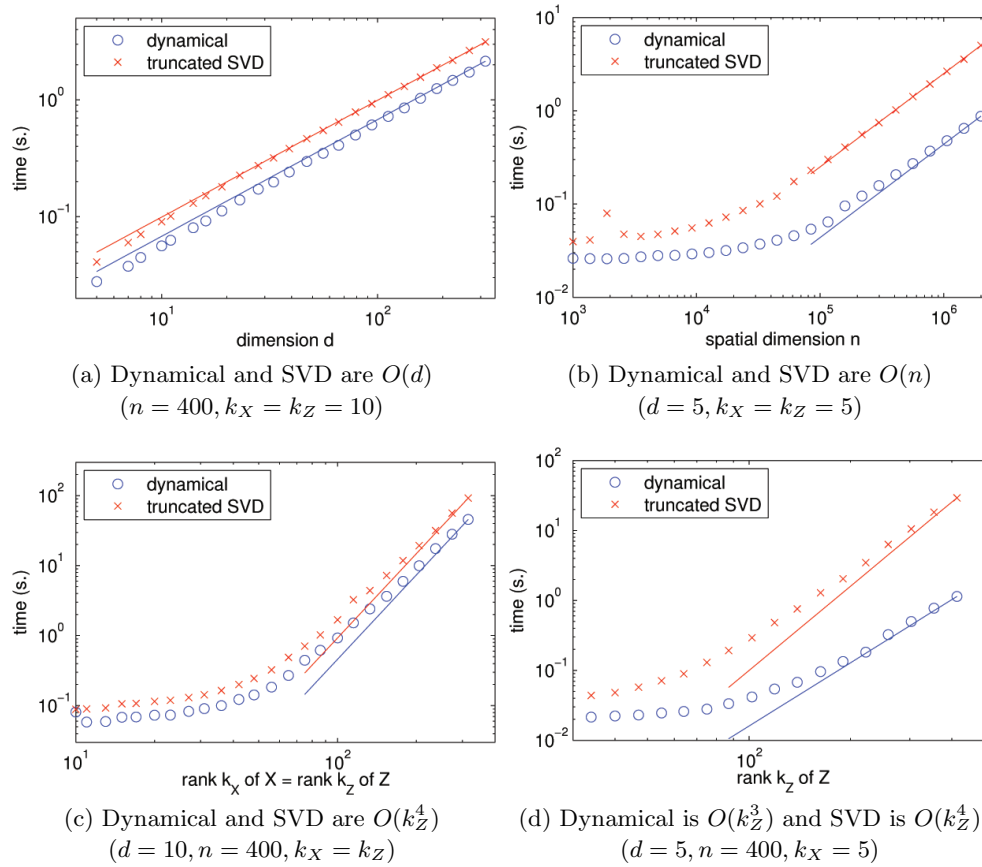


FIG. 7.2. Computational cost in seconds of  $P_X(Z)$  and truncation by SVD for random tensors with various values of  $d, n, k_X$ , and  $k_Z$ . The full lines are the indicated polynomial scalings.

`htensor.truncate_sum` in the `htucker` toolbox. As is evident from the figures, the projection is always faster than the SVD-based truncation, but usually only by a constant factor of about 2 to 5. However, in the case of fixed  $k_X$  and varying  $k_Z$ , the projection  $P_X(Z)$  indeed improves the order of complexity.

**7.3. Example 1: Approximating a periodic tensor.** The theory from section 5 predicts that the dynamical approach gives exact reconstruction of tensors which have fixed hierarchical rank; in other words, when the constant  $c$  in the theorems of section 5 is zero. We investigate this by integrating the HT tensor  $A(t) = A(\mathbf{U}_L(t), \mathbf{B}_I(t))$  such that

$$\begin{aligned} \mathbf{B}_{\tau_r}(t) &= e^{\sin(2\pi t)} \mathbf{B}_{\tau_r}(0) && \text{for } \tau = \tau_r, \\ \mathbf{B}_{\tau}(t) &= \mathbf{B}_{\tau}(0) && \text{for all } \tau \in I \setminus \{\tau_r\}, \\ \mathbf{U}_{\tau}(t) &= \mathbf{U}_{\tau}(0)\mathbf{Q}(t) && \text{for all } \tau \in L, \end{aligned}$$

with  $A(0)$  a random HT tensor, and  $\mathbf{Q}(t)$  a Givens rotation in the first two coordinates with  $\mathbf{Q}_{1,1}(t) = \cos(t)$ . The case for  $n = 25, d = 3$ , and  $k_{\tau} = 3$  is displayed in Figure 7.3 and it is evident that  $X(t)$  indeed approximates  $A(t)$  to the order of the integration tolerance  $10^{-9}$ .

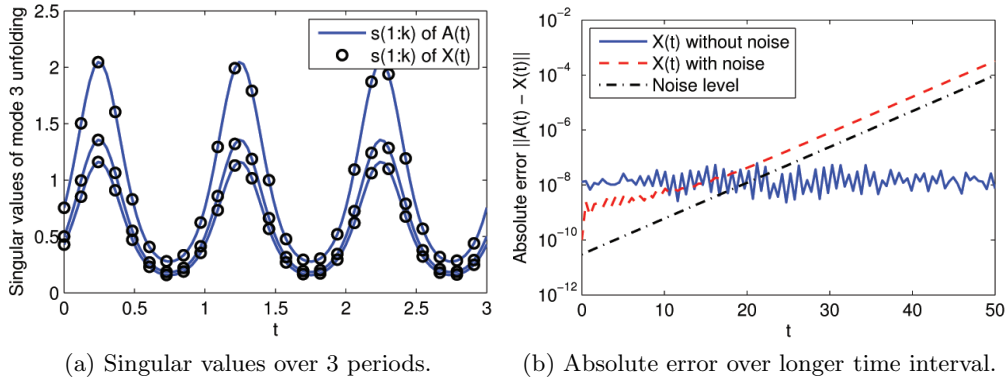


FIG. 7.3. Approximation  $X(t)$  of an HT tensor  $A(t)$  which is periodic in  $t$ .

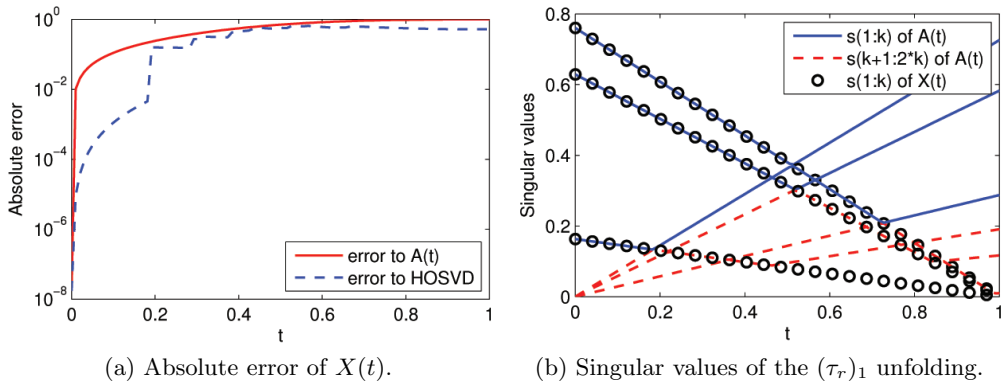


FIG. 7.4. Dynamical approximation  $Y(t)$  to the linear interpolant  $A(t)$  of two HT tensors.

When  $A(t)$  leaves  $\mathcal{M}$ , the approximation will deteriorate as  $t$  increases. In the right panel of Figure 7.3, we have repeated the same integration but now for

$$\hat{A}(t) = A(t) + 10^{-10} e^{0.3t} C,$$

with  $C$  a random full tensor such that there is growing noise. One can clearly observe that as expected, the approximation error is completely dominated by this noise.

**7.4. Example 2: Interpolating two tensors.** As the next illustration, we interpolate two HT tensors  $A_0$  and  $A_1$  with the same dimension tree but different ranks:  $A_0$  has rank  $k_0$ , whereas the rank of  $A_1$  is  $2k_0$ . The dynamical approach will then integrate

$$A(t) = t A_0 + (1 - t) A_1$$

from  $t = 0$  with  $A(0) = A_0 \in \mathcal{M}$  up to  $t = 1$ . Since generically  $A(t)$  does not lie in  $\mathcal{M}$  for  $t \neq 0$ , we cannot expect that  $X(t)$  will approximate  $A(t)$  well on the whole interval  $[0, 1]$ . However, as long as the singular values have a sufficiently big gap,  $X(t)$  should be a reasonable approximation to  $A(t)$ .

In Figure 7.4 we have verified this for  $n = 1000$ ,  $d = 5$ , and  $k_0 = 3$ . In the left panel, one clearly observes that the error to  $A(t)$  quickly deteriorates but the error to the quasi-best approximation by the SVD-based truncation behaves much better

up to  $t \simeq 0.2$ . In the right panel, the singular values of the  $(\tau_r)_1$  unfolding explain this behavior: Around  $t \simeq 0.2$ , the singular values cross and the dynamical approach tracks the smooth, but wrong, branch of the singular values. The result is a large increase of the error which cannot be undone except by restarting.

## REFERENCES

- [1] P.-A. ABSIL, R. MAHONY, AND R. SEPULCHRE, *Optimization Algorithms on Matrix Manifolds*, Princeton University Press, Princeton, NJ, 2008.
- [2] B. W. BADER, T. G. KOLDA, ET AL., *Matlab Tensor Toolbox Version 2.5*, available online at <http://www.sandia.gov/~tgkolda/TensorToolbox/>, 2012.
- [3] J. BALLANI AND L. GRASEDYCK, *A projection method to solve linear systems in tensor format*, Numer. Linear Algebra Appl., 20 (2013), pp. 27–43.
- [4] M. H. BECK, A. JÄCKLE, G. A. WORTH, AND H.-D. MEYER, *The multiconfiguration time-dependent Hartree (MCTDH) method: A highly efficient algorithm for propagating wavepackets*, Phys. Reports, 324 (2000), pp. 1–105.
- [5] D. CONTE AND C. LUBICH, *An error analysis of the multi-configuration time-dependent Hartree method of quantum dynamics*, M2AN Math. Model. Numer. Anal., 44 (2010), pp. 759–780.
- [6] S. DOLGOV, B. KHOROMSKIJ, AND I. OSELEDETS, *Fast Solution of Multi-dimensional Parabolic Problems in the TT/QTT-format with Initial Application to the Fokker–Planck Equation*, 80/2011, Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig, Germany, 2012.
- [7] L. GRASEDYCK, *Hierarchical singular value decomposition of tensors*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2029–2054.
- [8] W. HACKBUSCH, *Tensor Spaces and Numerical Tensor Calculus*, Springer-Verlag, Berlin, Heidelberg, 2012.
- [9] W. HACKBUSCH AND S. KÜHN, *A new scheme for the tensor representation*, J. Fourier Anal. Appl., 15 (2009), pp. 706–722.
- [10] J. HAEGEMAN, I. CIRAC, T. OSBORNE, I. PIŻORN, H. VERSHELDE, AND F. VERSTRAETE, *Time-dependent variational principle for quantum lattices*, Phys. Rev. Lett., 107 (2011) 070601.
- [11] S. HOLTZ, T. ROHWEDDER, AND R. SCHNEIDER, *The alternating linear scheme for tensor optimisation in the tensor train format*, SIAM J. Sci. Comput., 34 (2012), pp. A683–A713.
- [12] S. HOLTZ, T. ROHWEDDER, AND R. SCHNEIDER, *On manifolds of tensors of fixed TT-rank*, Numer. Math., 120 (2012), pp. 701–731.
- [13] B. KHOROMSKIJ AND I. OSELEDETS, *DMRG + QTT Approach to High-dimensional Quantum Molecular Dynamics*, Technical report 69/2010, MPI Leipzig, Leipzig, Germany, 2010.
- [14] B. KHOROMSKIJ, I. OSELEDETS, AND R. SCHNEIDER, *Efficient time-stepping scheme for dynamics on TT-manifolds*, 24/2012, Max-Planck-Institut für Mathematik in den Naturwissenschaften, Leipzig, Germany, 2012.
- [15] O. KOCH AND C. LUBICH, *Dynamical low-rank approximation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 434–454.
- [16] O. KOCH AND C. LUBICH, *Dynamical tensor approximation*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2360–2375.
- [17] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, SIAM Rev., 51 (2009), pp. 455–500.
- [18] D. KRESSNER AND C. TOBLER, *Preconditioned low-rank methods for high-dimensional elliptic PDE eigenvalue problems*, Comput. Methods. Appl. Math., 11 (2011), pp. 363–381.
- [19] D. KRESSNER AND C. TOBLER, *htucker—A MATLAB Toolbox for Tensors in Hierarchical Tucker Format*, Technical report 2012-02, Seminar for Applied Mathematics, ETH Zurich, Zürich, Switzerland, 2012.
- [20] C. LUBICH, *From Quantum to Classical Molecular Dynamics: Reduced Models and Numerical Analysis*, European Mathematical Society, Zürich, 2008.
- [21] K. H. MARTI, B. BAUER, M. REIHER, M. TROYER, AND F. VERSTRAETE, *Complete-graph tensor network states: a new fermionic wave function ansatz for molecules*, New Journal of Physics, 12 (2010), 103008.
- [22] H. MEYER, U. MANTHEA, AND L. S. CEDERBAUMA, *The multi-configurational time-dependent Hartree approach*, Chem. Phys. Lett., 165 (1990), pp. 73–78.
- [23] V. MURG, F. VERSTRAETE, O. LEGEZA, AND R. M. NOACK, *Simulating strongly correlated quantum systems with tree tensor networks*, Phys. Rev. B., 82 (2010) 205105.
- [24] A. NONNENMACHER AND C. LUBICH, *Dynamical low-rank approximation: applications and nu-*



- merical experiments*, Math. Comput. Simulation, 79 (2008), pp. 1346–1357.
- [25] I. V. OSELEDETS, *Tensor-train decomposition*, SIAM J. Sci. Comput., 33 (2011), pp. 2295–2317.
  - [26] I. V. OSELEDETS AND E. E. TYRTYSHNIKOV, *Breaking the curse of dimensionality, or how to use SVD in many dimensions*, SIAM J. Sci. Comput., 31 (2009), pp. 3744–3759.
  - [27] U. SCHOLLWÖCK, *The density-matrix renormalization group in the age of matrix product states*, Ann. Physics, 326 (2011), pp. 96–192.
  - [28] G. W. STEWART, *Matrix Algorithms: Vol. 1, Basic Decompositions*, SIAM, Philadelphia, 1998.
  - [29] G. W. STEWART AND J. SUN, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
  - [30] A. USCHMAJEV AND B. VANDEREYCKEN, *The geometry of algorithms using hierarchical tensors*, Linear Algebra Appl, in press.
  - [31] B. VANDEREYCKEN AND S. VANDEWALLE, *A Riemannian optimization approach for computing low-rank solutions of Lyapunov equations*, SIAM J. Matrix Anal. Appl., 31 (2010), pp. 2553–2579.
  - [32] H. WANG AND M. THOSS, *Multilayer formulation of the multiconfiguration time-dependent Hartree theory*, J. Chem. Phys., 119 (2003), pp. 1289–1299.