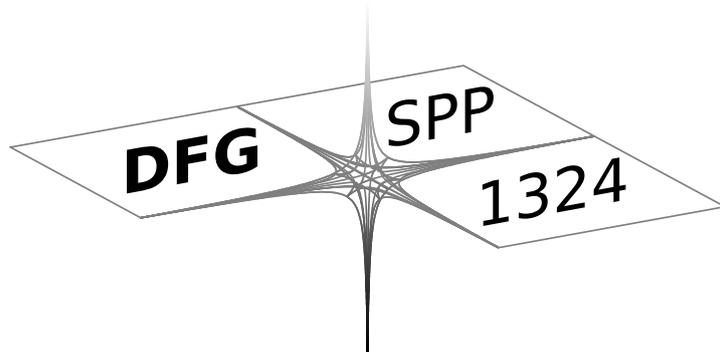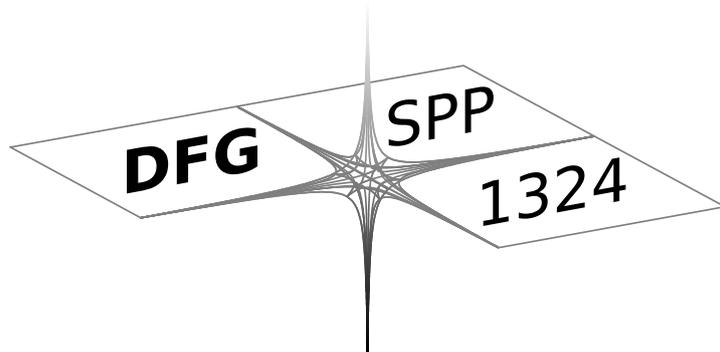# DFG-Schwerpunktprogramm 1324

„Extraktion quantifizierbarer Information aus komplexen Systemen"

Dynamical Low-rank Approximation of Tensors

O. Koch, C. Lubich

Preprint 28

Edited by

DFG-Schwerpunktprogramm 1324

Dynamical Low-rank Approximation of Tensors

O. Koch, C. Lubich

Preprint 28

DYNAMICAL LOW-RANK APPROXIMATION OF TENSORS

OTHMAR KOCH* AND CHRISTIAN LUBICH‡

Abstract. For the low-rank approximation of time-dependent data tensors and of solutions to tensor differential equations, we study a computational approach that can be viewed as a continuous-time updating procedure, working only with the increments rather than the full tensor. In this method, the derivative is projected onto the tangent space of the manifold of rank-$(r_1, \ldots, r_N)$ tensors at the current approximation. This yields nonlinear differential equations for the factors in a Tucker decomposition, suitable for numerical integration. Approximation properties of this approach are analyzed.

Key words. Low-rank approximation, time-varying tensors, continuous updating, Tucker decomposition, tensor differential equations.

AMS subject classifications. 15A18,15A69,65F99,65L05

1. Introduction. Low-rank approximation of tensors is an active research area with both interesting theoretical questions and numerous applications in the compression and retrieval of large structured data; see, e.g., [2, 3, 4, 9, 15] and references therein. While these papers deal with the approximation of a fixed tensor, we here consider the low-rank approximation of time-varying (or parameter-dependent) tensors. There arises the question as how to update a given low-rank approximation, a problem that has been addressed in a different, discrete-time setting in [13, 14]. The approach considered in the present paper can be viewed as a continuous-time updating technique, working only with the increments in the tensors rather than the tensors themselves. It extends the dynamical low-rank approximation of matrices [6, 12] to the higher-order tensor case.

Consider a time-varying family of tensors $\mathcal{A}(t) \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, for $0 \le t \le \bar{t}$. Let $\mathcal{M}_\mathbf{r}$ denote the manifold of all order-N tensors of rank $\mathbf{r} = (r_1, \ldots, r_N)$ (see §2 below for the notion of rank employed in this paper), where $r_n \le I_n$ (and typically $r_n \ll I_n$) for $n = 1, \ldots, N$. The best approximation to $\mathcal{A}(t)$ in $\mathcal{M}_\mathbf{r}$ (with respect to the Frobenius norm $\| \cdot \|$) is

$$\mathcal{X}(t) \in \mathcal{M}_\mathbf{r} \quad \text{such that} \quad \| \mathcal{X}(t) - \mathcal{A}(t) \| = \min. \tag{1.1}$$

Here, we consider instead the *dynamical low-rank approximation* $\mathcal{Y}(t) \in \mathcal{M}_\mathbf{r}$ determined from the condition that for every t the derivative $\dot{\mathcal{Y}}(t)$, which is in the tangent space $\mathcal{T}_{\mathcal{Y}(t)} \mathcal{M}_\mathbf{r}$, be chosen as

$$\dot{\mathcal{Y}}(t) \in \mathcal{T}_{\mathcal{Y}(t)} \mathcal{M}_\mathbf{r} \quad \text{such that} \quad \| \dot{\mathcal{Y}}(t) - \dot{\mathcal{A}}(t) \| = \min. \tag{1.2}$$

This is complemented with an initial condition, ideally $\mathcal{Y}(0) = \mathcal{X}(0)$. Note that for given $\mathcal{Y}(t)$, the derivative $\dot{\mathcal{Y}}(t)$ is obtained in (1.2) by a *linear* projection, though onto a state-dependent vector space. Problem (1.2) yields an initial value problem of nonlinear ordinary differential equations on $\mathcal{M}_\mathbf{r}$, which becomes numerically efficiently accessible with a Tucker decomposition of rank-$(r_1, \ldots, r_N)$ tensors. We will study theoretical properties of this approach in this paper: in §2 we derive the differential

---
*Vienna University of Technology, Institute for Analysis and Scientific Computing, Wiedner Hauptstrasse 8–10, A-1040 Wien, Austria. Email: `othmar@othmar-koch.org`

‡Mathematisches Institut, Universität Tübingen, Auf der Morgenstelle 10, D–72076 Tübingen, Germany. Email: `lubich@na.uni-tuebingen.de`

equations that are to be solved numerically, and after an auxiliary section on the tangent space projection (§3) we study approximation properties in §4. Some numerical experiments are given in §5 (see also [12] for further numerical experiments with time-dependent 3-tensors that arise from a discretized PDE in 3 space dimensions).

In a different context, a closely related approach was developed in the *multiconfiguration time-dependent Hartree method* of multi-particle quantum dynamics [1, 11], where the multidimensional time-dependent wave function is approximated by a linear combination of products of functions depending only on a single spatial variable.

The present paper extends the dynamical low-rank approximation of matrices, studied in our paper [6], to tensors. Though there are many similarities with the matrix case, the treatment of the tensor case is not simply a straightforward extension and therefore requires a careful discussion.

2. Differential equations for low-rank approximation.

2.1. Prerequisites. We use the tensor notation of the review article [7], to which we refer for further details, results and numerous references.

Norm and inner product of tensors. The norm of a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is the Euclidean norm of the vector $\mathbf{y}$ that carries the entries $y_{i_1,\ldots,i_N}$ of $\boldsymbol{\mathcal{Y}}$. The inner product $\langle \boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \rangle$ of two tensors $\boldsymbol{\mathcal{X}}, \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is the Euclidean inner product of the two corresponding vectors $\mathbf{x}$ and $\mathbf{y}$.

Unfolding and reconstruction. The nth unfolding of a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is the matrix $\mathbf{Y}_{(n)} \in \mathbb{R}^{I_n \times I_{n+1} \ldots I_N I_1 \ldots I_{n-1}}$ that aligns all entries $y_{i_1,\ldots,i_N}$ with fixed i_n in the i_n-th row of the matrix, ordered lexicographically. We denote

$$\mathbf{Y}_{(n)} = [\boldsymbol{\mathcal{Y}}]_{(n)},$$

and clearly the tensor $\boldsymbol{\mathcal{Y}}$ can be reshaped from its unfolding $\mathbf{Y}_{(n)}$: we write

$$\boldsymbol{\mathcal{Y}} = \left[\mathbf{Y}_{(n)} \right]^{(n)}.$$

The n-mode product. For a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ and a matrix $\mathbf{V} \in \mathbb{R}^{J_n \times I_n}$, the n-mode product

$$\boldsymbol{\mathcal{Y}} \times_n \mathbf{V} \in \mathbb{R}^{I_1 \times \ldots \times I_{n-1} \times J_n \times I_{n+1} \times \ldots \times I_N}$$

is defined by the relation

$$[\boldsymbol{\mathcal{Y}} \times_n \mathbf{V}]_{(n)} = \mathbf{V} \mathbf{Y}_{(n)}. \tag{2.1}$$

In particular, we then have, for another matrix $\mathbf{W}$ of appropriate dimension,

$$(\boldsymbol{\mathcal{Y}} \times_n \mathbf{V}) \times_n \mathbf{W} = \boldsymbol{\mathcal{Y}} \times_n (\mathbf{W}\mathbf{V}). \tag{2.2}$$

2.2. Manifold of rank-$(r_1, \ldots, r_N)$ tensors and Tucker decomposition. The n-rank of a tensor $\boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$ is

$$r_n = \operatorname{rank}\left(\mathbf{Y}_{(n)} \right).$$

For given $\mathbf{r} = (r_1, \ldots, r_N)$ with $r_n \leq I_n$ (and typically $r_n \ll I_n$), the set

$$\mathcal{M}_{\mathbf{r}} = \left\{ \boldsymbol{\mathcal{Y}} \in \mathbb{R}^{I_1 \times \cdots \times I_N} : \boldsymbol{\mathcal{Y}} \text{ has } n\text{-rank } r_n \text{ for } n = 1, \ldots, N \right\}$$

is a manifold that will serve as an approximation manifold for general tensors $\mathcal{A} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$. As is known (see [8]), every tensor $\mathcal{Y}$ of rank $\mathbf{r} = (r_1, \ldots, r_N)$ can be written as a *Tucker decomposition*

$$\mathcal{Y} = \mathbf{S} \times_1 \mathbf{U}_1 \ldots \times_N \mathbf{U}_N =: \mathbf{S} \underset{n=1}{\overset{N}{\mathsf{X}}} \mathbf{U}_n , \tag{2.3}$$

where the *core tensor* $\mathbf{S} \in \mathbb{R}^{r_1 \times \cdots \times r_N}$ is of full rank $\mathbf{r} = (r_1, \ldots, r_N)$, and the matrices $\mathbf{U}_n \in \mathbb{R}^{I_n \times r_n}$ have orthonormal columns $\mathbf{u}_{j_n}^{(n)}$ $(j_n = 1, \ldots, r_n)$. In terms of the entries of $\mathbf{S} = (s_{j_1 \ldots j_N})$, the above expression can be rewritten as a linear combination of rank-1 tensors that are formed as the outer products of the column vectors:

$$\mathcal{Y} = \sum_{j_1, \ldots, j_N} s_{j_1 \ldots j_N} \, \mathbf{u}_{j_1}^{(1)} \circ \ldots \circ \mathbf{u}_{j_N}^{(N)} .$$

In terms of the nth unfolding, we have the useful matrix formula

$$\mathbf{Y}_{(n)} = \mathbf{U}_n \mathbf{S}_{(n)} (\mathbf{U}_N \otimes \ldots \otimes \mathbf{U}_{n+1} \otimes \mathbf{U}_{n-1} \otimes \ldots \otimes \mathbf{U}_1)^T =: \mathbf{U}_n \mathbf{S}_{(n)} \bigotimes_{k \neq n} \mathbf{U}_k^T , \quad (2.4)$$

where $\otimes$ denotes the Kronecker matrix product.

The representation (2.3) is not unique: replacing $\mathbf{U}_n$ by $\tilde{\mathbf{U}}_n = \mathbf{U}_n \mathbf{Q}_n$ with orthogonal matrices $\mathbf{Q}_n$ and $\mathbf{S}$ by $\tilde{\mathbf{S}} = \mathbf{S} \mathsf{X}_{n=1}^N \mathbf{Q}_n^T$, yields the same tensor $\mathcal{Y} = \mathbf{S} \mathsf{X}_{n=1}^N \mathbf{U}_n = \tilde{\mathbf{S}} \mathsf{X}_{n=1}^N \tilde{\mathbf{U}}_n$.

2.3. Tangent tensors. As a substitute for the non-uniqueness in the decomposition (2.3), we will use a *unique decomposition in the tangent space*. Let $\mathcal{V}_{I,r}$ denote the Stiefel manifold of real $I \times r$ matrices with orthonormal columns. The tangent space at $\mathbf{U} \in \mathcal{V}_{I,r}$ is

$$\mathcal{T}_{\mathbf{U}} \mathcal{V}_{I,r} = \{ \dot{\mathbf{U}} \in \mathbb{R}^{I \times r} : \ \dot{\mathbf{U}}^T \mathbf{U} + \mathbf{U}^T \dot{\mathbf{U}} = \mathbf{0} \} = \{ \dot{\mathbf{U}} \in \mathbb{R}^{I \times r} : \ \mathbf{U}^T \dot{\mathbf{U}} \in \mathrm{so}(r) \},$$

where $\mathrm{so}(r)$ denotes the space of skew-symmetric real $r \times r$ matrices. Consider the extended tangent map of $(\mathbf{S}, \mathbf{U}_1, \ldots, \mathbf{U}_N) \mapsto \mathcal{Y} = \mathbf{S} \mathsf{X}_{n=1}^N \mathbf{U}_n$,

$$\mathbb{R}^{r_1 \times \cdots \times r_N} \times \prod_{n=1}^N \mathcal{T}_{\mathbf{U}_n} \mathcal{V}_{I_n, r_n} \ \to \ \mathcal{T}_{\mathcal{Y}} \mathcal{M}_{\mathbf{r}} \times \prod_{n=1}^N \mathrm{so}(r_n)$$

$$(\dot{\mathbf{S}}, \dot{\mathbf{U}}_1, \ldots, \dot{\mathbf{U}}_N) \ \mapsto \ \left(\dot{\mathbf{S}} \underset{n=1}{\overset{N}{\mathsf{X}}} \mathbf{U}_n + \sum_{n=1}^N \mathbf{S} \times_n \dot{\mathbf{U}}_n \underset{k \neq n}{\mathsf{X}} \mathbf{U}_k, \ \mathbf{U}_1^T \dot{\mathbf{U}}_1, \ldots, \mathbf{U}_N^T \dot{\mathbf{U}}_N \right).$$

This linear map turns out to be an isomorphism, since its inverse can be constructed explicitly by an argument similar to the proof of (2.7) below. Hence, every tangent tensor $\dot{\mathcal{Y}} \in \mathcal{T}_{\mathcal{Y}} \mathcal{M}_{\mathbf{r}}$ is of the form

$$\dot{\mathcal{Y}} = \dot{\mathbf{S}} \underset{n=1}{\overset{N}{\mathsf{X}}} \mathbf{U}_n + \sum_{n=1}^N \mathbf{S} \times_n \dot{\mathbf{U}}_n \underset{k \neq n}{\mathsf{X}} \mathbf{U}_k , \tag{2.5}$$

where $\dot{\mathbf{S}} \in \mathbb{R}^{r_1 \times \cdots \times r_N}$ and $\dot{\mathbf{U}}_n \in \mathcal{T}_{\mathbf{U}_n} \mathcal{V}_{I_n, r_n}$. Moreover, it follows that $\dot{\mathbf{S}}$ and $\dot{\mathbf{U}}_n$ are uniquely determined by $\dot{\mathcal{Y}}$ and the chosen $\mathbf{S}$ and $\mathbf{U}_n$ in (2.3) if we impose the orthogonality constraints

$$\mathbf{U}_n^T \dot{\mathbf{U}}_n = \mathbf{0} . \tag{2.6}$$

We now show that then $\dot{\mathbf{S}}$ and $\dot{\mathbf{U}}_n$ are actually given by the following formulas:

$$
\begin{aligned}
\dot{\mathbf{S}} &= \dot{\mathbfcal{Y}} \underset{n=1}{\overset{N}{\times}} \mathbf{U}_n^T \\
\dot{\mathbf{U}}_n &= \mathbf{P}_n^\perp \big[\dot{\mathbfcal{Y}} \underset{k\neq n}{\times} \mathbf{U}_k^T\big]_{(n)} \mathbf{S}_{(n)}^\dagger
\end{aligned}
\tag{2.7}
$$

with the projection $\mathbf{P}_n^\perp = \mathbf{I} - \mathbf{U}_n \mathbf{U}_n^T$ and the pseudo-inverse $\mathbf{S}_{(n)}^\dagger = \mathbf{S}_{(n)}^T \big(\mathbf{S}_{(n)} \mathbf{S}_{(n)}^T\big)^{-1}$. The $r_n \times r_n$ matrix $\mathbf{S}_{(n)} \mathbf{S}_{(n)}^T$ is invertible since $\mathbf{S}_{(n)}$ is of rank r_n by assumption.

Proof of (2.7): When we multiply (2.5) by $\mathsf{X}_{n=1}^N \mathbf{U}_n^T$, then we obtain by (2.2) and the orthogonality relations $\mathbf{U}_n^T \mathbf{U}_n = \mathbf{I}$ and (2.6) the first equation of (2.7). For the second equation, we multiply (2.5) by $\mathsf{X}_{k\neq n} \mathbf{U}_k^T$, and again using (2.2) and the orthogonality relations we obtain

$$
\dot{\mathbfcal{Y}} \underset{k\neq n}{\times} \mathbf{U}_k^T = \dot{\mathbf{S}} \times_n \mathbf{U}_n + \mathbf{S} \times_n \dot{\mathbf{U}}_n.
\tag{2.8}
$$

By the equation for $\dot{\mathbf{S}}$ and once again (2.2), the first expression on the right-hand side becomes

$$
\dot{\mathbf{S}} \times_n \mathbf{U}_n = \dot{\mathbfcal{Y}} \times_n (\mathbf{U}_n \mathbf{U}_n^T) \underset{k\neq n}{\times} \mathbf{U}_k^T,
$$

or in its mode-n unfolding,

$$
\mathbf{U}_n \dot{\mathbf{S}}_{(n)} = \mathbf{U}_n \mathbf{U}_n^T \big[\dot{\mathbfcal{Y}} \underset{k\neq n}{\times} \mathbf{U}_k^T\big]_{(n)}.
$$

Taking the mode-n unfolding on both sides of (2.8) and rearranging terms then gives

$$
\dot{\mathbf{U}}_n \mathbf{S}_{(n)} = \big(\mathbf{I} - \mathbf{U}_n \mathbf{U}_n^T\big) \big[\dot{\mathbfcal{Y}} \underset{k\neq n}{\times} \mathbf{U}_k^T\big]_{(n)}.
$$

Multiplying this relation from the right with $\mathbf{S}_{(n)}^\dagger$ yields the second equation of (2.7).

2.4. Dynamical low-rank approximation. We now turn to the low-rank approximation (1.2) of time-dependent tensors $\mathbfcal{A}(t)$. The minimization condition (1.2) on the tangent space is equivalent to an orthogonal projection: find $\dot{\mathbfcal{Y}} \in \mathcal{T}_\mathbfcal{Y} \mathcal{M}_\mathbf{r}$ (we omit the argument t) satisfying

$$
\langle \dot{\mathbfcal{Y}} - \dot{\mathbfcal{A}}, \mathbfcal{V} \rangle = 0 \quad \text{for all} \quad \mathbfcal{V} \in \mathcal{T}_\mathbfcal{Y} \mathcal{M}_\mathbf{r}.
\tag{2.9}
$$

This is a Galerkin condition on the tangent space $\mathcal{T}_\mathbfcal{Y} \mathcal{M}_\mathbf{r}$. With this formulation we derive differential equations for the factors in the Tucker decomposition (2.3).

THEOREM 2.1. *For a rank-$(r_1, \ldots, r_N)$ tensor $\mathbfcal{Y} = \mathbf{S} \mathsf{X}_{n=1}^N \mathbf{U}_N \in \mathcal{M}_\mathbf{r}$ with core tensor $\mathbf{S} \in \mathbb{R}^{r_1 \times \cdots \times r_N}$ and n-mode factors $\mathbf{U}_n \in \mathbb{R}^{I_n \times r_n}$ having orthonormal columns, condition (1.2) or (2.9) is equivalent to*

$$
\dot{\mathbfcal{Y}} = \dot{\mathbf{S}} \underset{n=1}{\overset{N}{\times}} \mathbf{U}_n + \sum_{n=1}^N \mathbf{S} \times_n \dot{\mathbf{U}}_n \underset{k\neq n}{\times} \mathbf{U}_k,
\tag{2.10}
$$

where the factors in the decomposition satisfy the system of differential equations

$$
\begin{aligned}
\dot{\mathbf{S}} &= \dot{\mathbfcal{A}} \underset{n=1}{\overset{N}{\times}} \mathbf{U}_n^T \\
\dot{\mathbf{U}}_n &= \mathbf{P}_n^\perp \big[\dot{\mathbfcal{A}} \underset{k\neq n}{\times} \mathbf{U}_k^T\big]_{(n)} \mathbf{S}_{(n)}^\dagger
\end{aligned}
\tag{2.11}
$$

with the projection $\mathbf{P}_n^\perp = \mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T$ *onto the orthogonal complement of the range of* $\mathbf{U}_n$ *and with the pseudo-inverse* $\mathbf{S}_{(n)}^\dagger = \mathbf{S}_{(n)}^T\big(\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T\big)^{-1}$ *of the n-mode unfolding* $\mathbf{S}_{(n)}$ *of* $\mathcal{S}$.

Note that Equations (2.11) are formally like (2.7), with $\dot{\mathcal{Y}}$ replaced by $\dot{\mathcal{A}}$.

Proof. We know from Section 2.3 that $\dot{\mathcal{Y}}$ can be written in the form (2.10) with $\dot{\mathbf{U}}_n$ satisfying the orthogonality condition (2.6). Choosing $\mathcal{V} = \mathcal{T}\mathsf{X}_{n=1}^N\mathbf{U}_n \in \mathcal{T}_{\mathcal{Y}}\mathcal{M}_{\mathbf{r}}$ with an arbitrary $\mathcal{T} \in \mathbb{R}^{r_1\times\cdots\times r_N}$, we have

$$\langle\dot{\mathcal{A}}, \mathcal{T} \underset{n=1}{\overset{N}{\mathsf{X}}} \mathbf{U}_n\rangle = \langle\dot{\mathcal{A}} \underset{n=1}{\overset{N}{\mathsf{X}}} \mathbf{U}_n^T, \mathcal{T}\rangle$$

$$\langle\dot{\mathcal{Y}}, \mathcal{T} \underset{n=1}{\overset{N}{\mathsf{X}}} \mathbf{U}_n\rangle = \langle\dot{\mathcal{S}}, \mathcal{T}\rangle\,.$$

Since this holds for every $\mathcal{T} \in \mathbb{R}^{r_1\times\cdots\times r_N}$, (2.9) yields the first equation of (2.11).

We now choose $\mathcal{V} = \mathcal{S} \times_n \mathbf{V}_n \mathsf{X}_{k\neq n} \mathbf{U}_k$ which is in $\mathcal{T}_{\mathcal{Y}}\mathcal{M}_{\mathbf{r}}$ if $\mathbf{V}_n$ satisfies the orthogonality relation $\mathbf{U}_n^T\mathbf{V}_n = \mathbf{0}$. We then have

$$\langle\dot{\mathcal{A}}, \mathcal{S} \times_n \mathbf{V}_n \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k\rangle = \langle\dot{\mathcal{A}} \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k^T, \mathcal{S} \times_n \mathbf{V}_n\rangle$$

$$= \langle\big[\dot{\mathcal{A}} \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k^T\big]_{(n)}, \mathbf{V}_n\mathbf{S}_{(n)}\rangle = \langle\big[\dot{\mathcal{A}} \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k^T\big]_{(n)}\mathbf{S}_{(n)}^T, \mathbf{V}_n\rangle\,,$$

where the matrix inner product in the second line is the Frobenius inner product. On the other hand we have, using (2.10) and the orthogonality relations,

$$\langle\dot{\mathcal{Y}}, \mathcal{S} \times_n \mathbf{V}_n \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k\rangle = \langle\dot{\mathcal{S}} \times_n \mathbf{U}_n + \mathcal{S} \times_n \dot{\mathbf{U}}_n, \mathcal{S} \times_n \mathbf{V}_n\rangle$$

$$= \langle\mathbf{U}_n\dot{\mathbf{S}}_{(n)} + \dot{\mathbf{U}}_n\mathbf{S}_{(n)}, \mathbf{V}_n\mathbf{S}_{(n)}\rangle$$

$$= \langle\dot{\mathbf{S}}_{(n)}, \mathbf{U}_n^T\mathbf{V}_n\mathbf{S}_{(n)}\rangle + \langle\dot{\mathbf{U}}_n\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T, \mathbf{V}_n\rangle$$

$$= \langle\dot{\mathbf{U}}_n\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T, \mathbf{V}_n\rangle\,.$$

By (2.9), we thus have

$$\langle\dot{\mathbf{U}}_n\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T - \big[\dot{\mathcal{A}} \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k^T\big]_{(n)}\mathbf{S}_{(n)}^T, \mathbf{V}_n\rangle = 0$$

for all $\mathbf{V}_n$ with $\mathbf{U}_n^T\mathbf{V}_n = \mathbf{0}$, that is, for all matrices $\mathbf{V}_n = (\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T)\mathbf{W}_n$ with an arbitrary matrix $\mathbf{W}_n \in \mathbb{R}^{I_n\times r_n}$. We therefore obtain

$$(\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T)\dot{\mathbf{U}}_n\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T = (\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^T)\big[\dot{\mathcal{A}} \underset{k\neq n}{\mathsf{X}} \mathbf{U}_k^T\big]_{(n)}\mathbf{S}_{(n)}^T\,.$$

Since $\mathbf{U}_n^T\dot{\mathbf{U}}_n = \mathbf{0}$ by condition (2.6), this yields the second equation of (2.11). $\square$

REMARK 2.1. Related differential equations were derived earlier in the chemical physics literature in [1] for the multiconfiguration time-dependent Hartree method of quantum mechanics, where an approximation to the multivariate wave function is sought for in the form of a linear combination of products of univariate functions.

3. Tangent space projection and curvature bounds. In operator notation, (2.9) can be written as

$$\dot{\mathcal{Y}} = \mathcal{P}(\mathcal{Y})\dot{\mathcal{A}}\,, \tag{3.1}$$

where $\mathcal{P}(\mathcal{Y})$ is the orthogonal projection onto the tangent space $\mathcal{T}_{\mathcal{Y}}\mathcal{M}_{\mathbf{r}}$. For a theoretical understanding of the dynamical low-rank approximation, an analysis of the properties of this projection operator is therefore essential. We begin by giving an explicit formula.

LEMMA 3.1. *Let* $\mathcal{Y} = \mathcal{S} \mathsf{X}_{n=1}^{N} \mathbf{U}_n \in \mathcal{M}_{\mathbf{r}}$ *with n-mode factors* $\mathbf{U}_n$ *having orthonormal columns. With the orthogonal projections*

$$\mathbf{P}_n = \mathbf{U}_n \mathbf{U}_n^T, \quad \mathbf{P}_n^{\perp} = \mathbf{I} - \mathbf{P}_n,$$
$$\mathbf{Q}_n = \bigotimes_{k \neq n} \mathbf{U}_k \, \mathbf{S}_{(n)}^{\dagger} \mathbf{S}_{(n)} \bigotimes_{k \neq n} \mathbf{U}_k^T,$$

the orthogonal projection onto the tangent space $\mathcal{T}_{\mathcal{Y}}\mathcal{M}_{\mathbf{r}}$ *is given by*

$$\mathcal{P}(\mathcal{Y})\mathcal{B} = \mathcal{B} \mathop{\mathsf{X}}_{n=1}^{N} \mathbf{P}_n + \sum_{n=1}^{N} \left[\mathbf{P}_n^{\perp} \mathbf{B}_{(n)} \mathbf{Q}_n \right]^{(n)} \tag{3.2}$$

for $\mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$, *with* $\mathbf{B}_{(n)} = [\mathcal{B}]_{(n)}$ *the nth unfolding of* $\mathcal{B}$.

Proof. Theorem 2.1 gives an expression for $\dot{\mathcal{Y}}$ determined by (2.9), that is, for $\dot{\mathcal{Y}} = \mathcal{P}(\mathcal{Y})\dot{\mathcal{A}}$. With $\mathcal{B}$ in place of $\dot{\mathcal{A}}$, this reads

$$\mathcal{P}(\mathcal{Y})\mathcal{B} = \mathcal{B} \mathop{\mathsf{X}}_{n=1}^{N} \mathbf{U}_n^T \mathop{\mathsf{X}}_{n=1}^{N} \mathbf{U}_n + \sum_{n=1}^{N} \mathcal{S} \times_n \left(\mathbf{P}_n^{\perp} \big[\mathcal{B} \mathop{\mathsf{X}}_{k \neq n} \mathbf{U}_k^T \big]_{(n)} \mathbf{S}_{(n)}^{\dagger} \right) \mathop{\mathsf{X}}_{k \neq n} \mathbf{U}_k$$

$$= \mathcal{B} \mathop{\mathsf{X}}_{n=1}^{N} \mathbf{P}_n + \sum_{n=1}^{N} \left[\mathbf{P}_n^{\perp} \mathbf{B}_{(n)} \bigotimes_{k \neq n} \mathbf{U}_k \, \mathbf{S}_{(n)}^{\dagger} \mathbf{S}_{(n)} \bigotimes_{k \neq n} \mathbf{U}_k^T \right]^{(n)},$$

where we have used (2.2) and (2.4). This is (3.2). □

The next lemma is the key tool for the approximation results of the following section.

LEMMA 3.2. *There are constants c and C (depending only on the order N and satisfying* $Cc \leq \frac{1}{2}$*) such that the following holds true. Let the rank-$(r_1, \ldots, r_N)$ tensor* $\mathcal{X} \in \mathcal{M}_{\mathbf{r}}$ *be such that the smallest nonzero singular value of the nth unfolding satisfies* $\sigma_{r_n}(\mathbf{X}_{(n)}) \geq \rho > 0$ *for* $n = 1, \ldots, N$, *and let* $\tilde{\mathcal{X}} \in \mathcal{M}_{\mathbf{r}}$ *with* $\|\tilde{\mathcal{X}} - \mathcal{X}\| \leq c\rho$. *Then, for all* $\mathcal{B} \in \mathbb{R}^{I_1 \times \cdots \times I_N}$,

$$\left\| \big(\mathcal{P}(\tilde{\mathcal{X}}) - \mathcal{P}(\mathcal{X}) \big) \mathcal{B} \right\| \leq C\rho^{-1} \|\tilde{\mathcal{X}} - \mathcal{X}\| \cdot \|\mathcal{B}\|, \tag{3.3}$$

$$\left\| \mathcal{P}^{\perp}(\mathcal{X})(\tilde{\mathcal{X}} - \mathcal{X}) \right\| \leq C\rho^{-1} \|\tilde{\mathcal{X}} - \mathcal{X}\|^2, \tag{3.4}$$

where $\mathcal{P}^{\perp}(\mathcal{X}) = I - \mathcal{P}(\mathcal{X})$ *is the projection onto the orthogonal complement of the tangent space* $\mathcal{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}}$.

Proof. (a) Writing $\mathcal{X} \in \mathcal{M}_{\mathbf{r}}$ in Tucker form as $\mathcal{X} = \mathcal{S} \mathsf{X}_{n=1}^{N} \mathbf{U}_n$, we note that the non-zero singular values of $\mathbf{X}_{(n)}$ are those of $\mathbf{S}_{(n)}$, and hence

$$\|\mathbf{S}_{(n)}^{\dagger}\| = \frac{1}{\sigma_{r_n}(\mathbf{S}_{(n)})} = \frac{1}{\sigma_{r_n}(\mathbf{X}_{(n)})} \leq \frac{1}{\rho}.$$

Since we have, by [5, p. 448],

$$|\sigma_{r_n}(\tilde{\mathbf{X}}_{(n)}) - \sigma_{r_n}(\mathbf{X}_{(n)})| \leq \|\tilde{\mathbf{X}}_{(n)} - \mathbf{X}_{(n)}\|_2 \leq \|\tilde{\mathbf{X}}_{(n)} - \mathbf{X}_{(n)}\|_F = \|\tilde{\mathcal{X}} - \mathcal{X}\|,$$

we obtain for $\|\tilde{\mathcal{X}} - \mathcal{X}\| \leq \frac{1}{2}\rho$ that

$$\sigma_{r_n}(\tilde{\mathbf{X}}_{(n)}) \geq \sigma_{r_n}(\mathbf{X}_{(n)}) - |\sigma_{r_n}(\tilde{\mathbf{X}}_{(n)}) - \sigma_{r_n}(\mathbf{X}_{(n)})| \geq \tfrac{1}{2}\rho\,,$$

and hence $\tilde{\mathcal{X}} = \tilde{\mathcal{S}} \, \mathsf{X}_{n=1}^{N} \, \tilde{\mathbf{U}}_n$ with $\|\tilde{\mathbf{S}}_{(n)}^{\dagger}\| \leq 2\rho^{-1}$.

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

$$\mathcal{X} + \tau(\tilde{\mathcal{X}} - \mathcal{X}) = \mathcal{Y}(\tau) + \mathcal{Z}(\tau) \quad \text{with} \quad \mathcal{Y}(\tau) \in \mathcal{M}_{\mathbf{r}}, \; \mathcal{Z}(\tau) \perp \mathcal{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}}.$$

A smooth such decomposition exists at least for small τ, but the arguments below show that it exists in fact for $0 \leq \tau \leq 1$. We denote

$$\Delta = \mathcal{P}(\mathcal{X})(\tilde{\mathcal{X}} - \mathcal{X}) \in \mathcal{T}_{\mathcal{X}}\mathcal{M}_{\mathbf{r}}, \quad \text{with} \;\; \|\Delta\| \leq \delta := \|\tilde{\mathcal{X}} - \mathcal{X}\|.$$

We then have $\mathcal{P}(\mathcal{X})(\mathcal{Y}(\tau) - \mathcal{X}) = \tau\Delta$, which yields

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

Since $\mathcal{P}(\mathcal{Y})\dot{\mathcal{Y}} = \dot{\mathcal{Y}}$, we have

$$\dot{\mathcal{Y}}(\tau) = \Delta + \big(\mathcal{P}(\mathcal{Y}(\tau)) - \mathcal{P}(\mathcal{X})\big)\dot{\mathcal{Y}}(\tau)\,.$$

(c) As long as the operator norm of $\big(\mathcal{P}(\mathcal{Y}(\tau)) - \mathcal{P}(\mathcal{X})\big)$ is bounded by $\frac{1}{2}$, we thus have

$$\|\dot{\mathcal{Y}}\| \leq 2\delta\,.$$

This yields $\|\mathcal{Y}(\tau) - \mathcal{X}\| \leq 2\delta$ for $0 \leq \tau \leq 1$, and for the Tucker factors of $\mathcal{Y}(\tau) = \mathcal{S}(\tau) \, \mathsf{X}_{n=1}^{N} \, \mathbf{U}_n(\tau)$ with $\mathbf{U}_n^T \dot{\mathbf{U}}_n = \mathbf{0}$ we therefore have by (2.7)

$$\|\dot{\mathcal{S}}\| \leq 2\delta\,, \quad \|\dot{\mathbf{U}}_n \mathbf{S}_{(n)}\| \leq 2\delta\,.$$

If $2\delta \leq \frac{1}{2}\rho$, then the argument of (a) applied to $\mathcal{Y}(\tau)$ instead of $\tilde{\mathcal{X}}$ shows that $\|\mathbf{S}_{(n)}^{\dagger}(\tau)\| \leq 2\rho^{-1}$, and hence

$$\|\dot{\mathbf{U}}_n\| \leq \|\dot{\mathbf{U}}_n \mathbf{S}_{(n)}\| \cdot \|\mathbf{S}_{(n)}^{\dagger}\| \leq 4\delta\rho^{-1} =: \gamma\,.$$

With this estimate we further obtain for $\mathbf{P}_n = \mathbf{U}_n \mathbf{U}_n^T$ that

$$\|\dot{\mathbf{P}}_n\| \leq 2\gamma\,.$$

Using the product rule for $\frac{d}{d\tau}\big(\mathbf{S}_{(n)}^T (\mathbf{S}_{(n)}\mathbf{S}_{(n)}^T)^{-1}\mathbf{S}_{(n)}\big)$ and the estimates for the norms of $\dot{\mathbf{S}}_{(n)}$ and $\mathbf{S}_{(n)}^{\dagger}$ by 2δ and $2\rho^{-1}$, respectively, we find that the norm of this derivative is bounded by 4γ, and hence we have for the projection $\mathbf{Q}_n$ of Lemma 3.1

$$\|\dot{\mathbf{Q}}_n\| \leq 2(N-1)\gamma + 4\gamma = 2(N+1)\gamma\,.$$

Expressing $\mathbf{U}_n(\tau) - \mathbf{U}_n(0) = \int_0^{\tau} \dot{\mathbf{U}}_n(s)\,ds$ and similarly for the increments of $\mathcal{S}$, $\mathbf{P}_n$, and $\mathbf{Q}_n$, we obtain from the formula of Lemma 3.1 and the above estimates

$$\|\big(\mathcal{P}(\mathcal{Y}(\tau)) - \mathcal{P}(\mathcal{X})\big)\mathcal{B}\| \leq \big(N \cdot 2\gamma + N \cdot (2\gamma + 2(N+1)\gamma)\big)\tau\|\mathcal{B}\|$$
$$= 2N(N+3)\gamma \cdot \tau\|\mathcal{B}\| = 8N(N+3)\,\delta\rho^{-1}\,\tau\,\|\mathcal{B}\|\,.$$

The operator norm of $\mathcal{P}(\boldsymbol{\mathcal{Y}}(\tau)) - \mathcal{P}(\boldsymbol{\mathcal{X}})$ thus does not exceed $\frac{1}{2}$ for $0 \leq \tau \leq 1$ if

$$\delta \leq c\rho \quad \text{with} \quad c = \frac{1}{16\,N(N+3)}\,,$$

and at $\tau = 1$ we then obtain the bound (3.3) with $C = 8N(N+3)$.

(d) We note

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

By the above estimates, this is bounded by

$$\|\mathcal{P}^{\perp}(\boldsymbol{\mathcal{X}})(\tilde{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}})\| \leq 8N(N+3)\,\rho^{-1}\delta^2\,,$$

which yields (3.4). $\square$

4. Approximation properties. We give approximation results that are extensions to tensors of results for the matrix case in [6].

4.1. Local quasi-optimality. If the low-rank approximation problem (1.1) has a continuously differentiable best approximation $\boldsymbol{\mathcal{X}}(t) \in \mathcal{M}_{\mathbf{r}}$, then the error of (1.2) can be bounded in terms of the best-approximation error $\|\boldsymbol{\mathcal{X}}(t) - \boldsymbol{\mathcal{A}}(t)\|$. The result requires a bound on $\dot{\boldsymbol{\mathcal{A}}}(t)$:

$$\|\dot{\boldsymbol{\mathcal{A}}}(t)\| \leq \mu \qquad \text{for} \ \ 0 \leq t \leq \bar{t}. \tag{4.1}$$

THEOREM 4.1. *Suppose that a continuously differentiable best approximation* $\boldsymbol{\mathcal{X}}(t) \in \mathcal{M}_{\mathbf{r}}$ *to* $\boldsymbol{\mathcal{A}}(t)$ *exists for* $0 \leq t \leq \bar{t}$. *Let* $\rho > 0$ *be such that the smallest nonzero singular value of the nth unfolding of* $\boldsymbol{\mathcal{X}}(t)$ *satisfies* $\sigma_{r_n}(\mathbf{X}_{(n)}(t)) \geq \rho$ *for* $n = 1, \ldots, N$, *and assume that the best-approximation error is bounded by* $\|\boldsymbol{\mathcal{X}}(t) - \boldsymbol{\mathcal{A}}(t)\| \leq c\rho$ *for* $0 \leq t \leq \bar{t}$, *with* c *of* Lemma 3.2. *Then, the approximation error of the dynamical low-rank approximation* (1.2) *with initial value* $\boldsymbol{\mathcal{Y}}(0) = \boldsymbol{\mathcal{X}}(0)$ *is bounded by*

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

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 3.2.

Proof. With Lemma 3.2 at hand, the proof is essentially the same as that of Theorem 5.1 in [6]. The error of the best approximation $\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}}$ is orthogonal to the tangent space $\mathcal{T}_{\boldsymbol{\mathcal{X}}}\mathcal{M}_{\mathbf{r}}$, or equivalently, $\mathcal{P}(\boldsymbol{\mathcal{X}})(\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}}) = \mathbf{0}$. We differentiate this relation with respect to t and denote $(\mathcal{P}'(\boldsymbol{\mathcal{X}}) \cdot \boldsymbol{\mathcal{B}})\dot{\boldsymbol{\mathcal{X}}} = \frac{d}{dt}\mathcal{P}(\boldsymbol{\mathcal{X}}(t))\boldsymbol{\mathcal{B}}$ to obtain

$$\mathcal{P}(\boldsymbol{\mathcal{X}})(\dot{\boldsymbol{\mathcal{X}}} - \dot{\boldsymbol{\mathcal{A}}}) + \big(\mathcal{P}'(\boldsymbol{\mathcal{X}}) \cdot (\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}})\big)\dot{\boldsymbol{\mathcal{X}}} = 0.$$

Since $\dot{\boldsymbol{\mathcal{X}}} \in \mathcal{T}_{\boldsymbol{\mathcal{X}}}\mathcal{M}_{\mathbf{r}}$, we have $\mathcal{P}(\boldsymbol{\mathcal{X}})\dot{\boldsymbol{\mathcal{X}}} = \dot{\boldsymbol{\mathcal{X}}}$, and the equation becomes

$$\big(\boldsymbol{\mathcal{I}} - \mathcal{P}'(\boldsymbol{\mathcal{X}}) \cdot (\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}})\big)\dot{\boldsymbol{\mathcal{X}}} = \mathcal{P}(\boldsymbol{\mathcal{X}})\dot{\boldsymbol{\mathcal{A}}}. \tag{4.2}$$

Lemma 3.2 and the condition $d := \|\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}}\| \leq c\rho$ yield

$$\|\mathcal{P}'(\boldsymbol{\mathcal{X}}) \cdot (\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}})\| \leq C\rho^{-1}d \leq Cc \leq \tfrac{1}{2},$$

and hence (4.2) can be solved for $\dot{\boldsymbol{\mathcal{X}}}$ to yield

$$\dot{\boldsymbol{\mathcal{X}}} = \mathcal{P}(\boldsymbol{\mathcal{X}})\dot{\boldsymbol{\mathcal{A}}} + \boldsymbol{\mathcal{D}} \qquad \text{with} \quad \|\boldsymbol{\mathcal{D}}\| \le 2C\rho^{-1}d\mu = 2\beta d.$$

We subtract this equation from (3.1), viz., $\dot{\boldsymbol{\mathcal{Y}}} = \mathcal{P}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{A}}}$, and integrate from 0 to t. As long as $e(t) := \|\boldsymbol{\mathcal{Y}}(t) - \boldsymbol{\mathcal{X}}(t)\| \le c\rho$, Lemma 3.2 yields

$$\|\big(\mathcal{P}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}})\big)\dot{\boldsymbol{\mathcal{A}}}\| \le C\rho^{-1}e\mu = \beta e,$$

and hence we obtain

$$e(t) \le \beta \int_0^t e(s)\,ds + 2\beta \int_0^t d(s)\,ds.$$

The result now follows with the Gronwall inequality. $\square$

4.2. A farther-reaching error bound. Smaller errors over longer time intervals are obtained if not only $\boldsymbol{\mathcal{X}} - \boldsymbol{\mathcal{A}}$, but also its derivative is small. We assume that $\boldsymbol{\mathcal{A}}(t)$ is of the form

$$\boldsymbol{\mathcal{A}}(t) = \boldsymbol{\mathcal{X}}(t) + \boldsymbol{\mathcal{E}}(t), \qquad 0 \le t \le \bar{t}, \tag{4.3}$$

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

$$\|\dot{\boldsymbol{\mathcal{X}}}(t)\| \le \mu \tag{4.4}$$

and the derivative of the remainder term is bounded by

$$\|\dot{\boldsymbol{\mathcal{E}}}(t)\| \le \varepsilon \tag{4.5}$$

with a small $\varepsilon > 0$.

THEOREM 4.2. *In addition to the above assumptions, suppose that the smallest singular values of the unfoldings of $\boldsymbol{\mathcal{X}}(t)$ are bounded from below by $\rho > 0$. Then, the approximation error of* (1.2) *with initial value $Y(0) = X(0)$ is bounded by*

$$\|\boldsymbol{\mathcal{Y}}(t) - \boldsymbol{\mathcal{X}}(t)\| \le 2t\varepsilon \qquad for \quad t \le \frac{\rho}{C\sqrt{\mu\varepsilon}}\,,$$

provided that $t \le \frac{c\rho}{2\varepsilon}$ and $t \le \bar{t}$. The constants C and c are those of Lemma 3.2.

Proof. The proof follows that of Theorem 5.2 in [6]. We note $\dot{\boldsymbol{\mathcal{X}}} = \mathcal{P}(\boldsymbol{\mathcal{X}})\dot{\boldsymbol{\mathcal{X}}}$, rewrite (3.1) as $\dot{\boldsymbol{\mathcal{Y}}} = \mathcal{P}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{X}}} + \mathcal{P}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{E}}}$, and subtract the two equations. We observe

$$\big(\mathcal{P}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}})\big)\dot{\boldsymbol{\mathcal{X}}} = -\big(\mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}^{\perp}(\boldsymbol{\mathcal{X}})\big)\dot{\boldsymbol{\mathcal{X}}} = -\mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{X}}} = -\mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})^2\dot{\boldsymbol{\mathcal{X}}}.$$

We take the inner product with $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}$ to obtain

$$\langle \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}, \big(\mathcal{P}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}})\big)\dot{\boldsymbol{\mathcal{X}}}\rangle = -\langle \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}, \mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{X}}}\rangle = -\langle \mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}), \mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{X}}}\rangle$$
$$= \langle \mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}), \big(\mathcal{P}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}})\big)\dot{\boldsymbol{\mathcal{X}}}\rangle.$$

With Lemma 3.2 and (4.4), (4.5) this yields

$$\langle \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}, \dot{\boldsymbol{\mathcal{Y}}} - \dot{\boldsymbol{\mathcal{X}}}\rangle = \langle \mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})(\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}), \big(\mathcal{P}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}})\big)\dot{\boldsymbol{\mathcal{X}}}\rangle + \langle \boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}, \mathcal{P}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{E}}}\rangle$$
$$\le C^2\rho^{-2}\mu\|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}\|^3 + \|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}\| \cdot \varepsilon,$$

and on the other hand we have

$$\langle \mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}, \dot{\mathbf{\mathcal{Y}}} - \dot{\mathbf{\mathcal{X}}} \rangle = \frac{1}{2}\frac{d}{dt}\|\mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}\|^2 = \|\mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}\|\frac{d}{dt}\|\mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}\|.$$

Taken together, we obtain for $e(t) = \|\mathbf{\mathcal{Y}}(t) - \mathbf{\mathcal{X}}(t)\|$ the differential inequality

$$\dot{e} \leq \gamma e^2 + \varepsilon, \qquad e(0) = 0,$$

with $\gamma = C^2\rho^{-2}\mu$. Hence, $e(t)$ is majorized by the solution of

$$\dot{y} = \gamma y^2 + \varepsilon, \qquad y(0) = 0,$$

which equals $y(t) = \sqrt{\varepsilon/\gamma}\,\tan(t\sqrt{\gamma\varepsilon})$ and is bounded by $2t\varepsilon$ for $t\sqrt{\gamma\varepsilon} \leq 1$. Lemma 3.2 remains applicable as long as $2t\varepsilon \leq c\rho$. □

4.3. Systems without gaps between the singular values. The results of the preceding subsections only give satisfactory error bounds when there is a gap in the distribution of the singular values of the unfoldings so that essential and inessential singular values are widely separated. We now consider a situation where such a gap need not exist. We make the assumptions of Theorem 4.2 and further that $\mathbf{\mathcal{X}}(t) \in \mathcal{M}_{\mathbf{r}}$ with all singular values of all unfoldings greater than $\rho > 0$ has a Tucker decomposition

$$\mathbf{\mathcal{X}}(t) = \overline{\mathbf{\mathcal{S}}}(t)\mathop{\vphantom{\sum}\mathchoice{\raise-0.5ex\hbox{\huge$\times$}}{\times}{\times}{\times}}_{n=1}^{N}\overline{\mathbf{U}}_n(t) \tag{4.6}$$

such that the following bounds hold for $n = 1, \ldots, N$ and $0 \leq t \leq \bar{t}$:

$$\left\|\overline{\mathbf{S}}_{(n)}^{\dagger}\dot{\overline{\mathbf{S}}}_{(n)}\right\| \leq \kappa, \qquad \|\dot{\overline{\mathbf{U}}}_n(t)\| \leq \nu. \tag{4.7}$$

Under these conditions we can show an $O(\varepsilon)$ error over times $O(1)$ even with $\rho \sim \varepsilon$.

THEOREM 4.3. *Under the conditions of* Theorem 4.2 *and with* (4.6)–(4.7), *the approximation error of* (1.2) *with initial value* $\mathbf{\mathcal{Y}}(0) = \mathbf{\mathcal{X}}(0)$ *is bounded by*

$$\|\mathbf{\mathcal{Y}}(t) - \mathbf{\mathcal{X}}(t)\| \leq 2t\varepsilon \qquad for \quad t \leq \frac{1}{\sqrt{c_1\kappa + c_2\nu}}\sqrt{\frac{\rho}{\varepsilon}}$$

with constants c_1 and c_2 that depend only on the order N. This holds as long as the right-hand side remains bounded by $c\rho$ with c of Lemma 3.2.

Proof. From the proof of Theorem 4.2 we have the equation

$$\langle \mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}, \dot{\mathbf{\mathcal{Y}}} - \dot{\mathbf{\mathcal{X}}} \rangle = -\langle \mathcal{P}^{\perp}(\mathbf{\mathcal{Y}})(\mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}), \mathcal{P}^{\perp}(\mathbf{\mathcal{Y}})\dot{\mathbf{\mathcal{X}}}\rangle + \langle \mathbf{\mathcal{Y}} - \mathbf{\mathcal{X}}, \mathcal{P}(\mathbf{\mathcal{Y}})\dot{\mathbf{\mathcal{E}}}\rangle. \tag{4.8}$$

For $e(t) = \|\mathbf{\mathcal{Y}}(t) - \mathbf{\mathcal{X}}(t)\| \leq c\rho$, the proof of Lemma 3.2 shows that there is a homotopy

$$\mathbf{\mathcal{Y}}(t, \tau) = \mathbf{\mathcal{S}}(t, \tau)\mathop{\vphantom{\sum}\mathchoice{\raise-0.5ex\hbox{\huge$\times$}}{\times}{\times}{\times}}_{n=1}^{N}\mathbf{U}_n(t, \tau)$$

with

$$\mathbf{\mathcal{Y}}(t, 1) = \mathbf{\mathcal{Y}}(t), \qquad \mathbf{\mathcal{Y}}(t, 0) = \mathbf{\mathcal{X}}(t),$$

for which the Tucker factors are bounded by

$$\left\|\frac{\partial\mathbf{\mathcal{S}}}{\partial\tau}(t, \tau)\right\| \leq 2e(t)$$

$$\left\|\frac{\partial\mathbf{U}_n}{\partial\tau}(t, \tau)\,\mathbf{S}_{(n)}(t, \tau)\right\| \leq 2e(t) \tag{4.9}$$

$$\left\|\frac{\partial\mathbf{U}_n}{\partial\tau}(t, \tau)\right\| \leq 4e(t)\rho^{-1} \leq 4c.$$

We write

$$\mathcal{P}^\perp(\boldsymbol{\mathcal{Y}}(t))\dot{\boldsymbol{\mathcal{X}}}(t) = \mathcal{P}^\perp(\boldsymbol{\mathcal{Y}}(t))\Big(\dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,0) + \sum_{n=1}^{N}\boldsymbol{\mathcal{S}}(t,0)\times_n\dot{\mathbf{U}}_n(t,0)\mathop{\mathsf{X}}_{k\neq n}\mathbf{U}_k(t,0)$$

$$- \dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,1) + \sum_{n=1}^{N}\boldsymbol{\mathcal{S}}(t,1)\times_n\dot{\mathbf{U}}_n(t,0)\mathop{\mathsf{X}}_{k\neq n}\mathbf{U}_k(t,1)\Big),$$

noting that $\mathcal{P}^\perp(\boldsymbol{\mathcal{Y}}(t))$ annihilates the terms in the second line, since they are in $\mathcal{T}_{\boldsymbol{\mathcal{Y}}}\mathcal{M}_{\mathbf{r}}$. We have

$$\Big\|\dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,1) - \dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,0)\Big\|$$

$$= \Big\|\int_0^1\frac{\partial}{\partial\tau}\Big(\dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,\tau)\Big)\partial\tau\Big\|$$

$$= \Big\|\int_0^1\sum_{n=1}^{N}\dot{\boldsymbol{\mathcal{S}}}(t,0)\times_n\frac{\partial\mathbf{U}_n}{\partial\tau}(t,\tau)\mathop{\mathsf{X}}_{k\neq n}\mathbf{U}_k(t,\tau)\Big)d\tau\Big\|$$

$$\leq \int_0^1\Big\|\frac{\partial\mathbf{U}_n}{\partial\tau}(t,\tau)\dot{\mathbf{S}}_{(n)}(t,0)\bigotimes_{k\neq n}\mathbf{U}_k(t,\tau)^T\Big\|d\tau$$

$$\leq \int_0^1\Big\|\frac{\partial\mathbf{U}_n}{\partial\tau}(t,\tau)\mathbf{S}_{(n)}(t,0)\Big\|\cdot\Big\|\mathbf{S}_{(n)}(t,0)^\dagger\dot{\mathbf{S}}_{(n)}(t,0)\Big\|d\tau\,.$$

Writing

$$\frac{\partial\mathbf{U}_n}{\partial\tau}(t,\tau)\mathbf{S}_{(n)}(t,0) = \frac{\partial\mathbf{U}_n}{\partial\tau}(t,\tau)\mathbf{S}_{(n)}(t,\tau) - \frac{\partial\mathbf{U}_n}{\partial\tau}(t,\tau)\int_0^\tau\frac{\partial\mathbf{S}_{(n)}}{\partial\sigma}(t,\sigma)\,d\sigma$$

and using the bounds (4.7) and (4.9) thus yields

$$\Big\|\dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,1) - \dot{\boldsymbol{\mathcal{S}}}(t,0)\mathop{\mathsf{X}}_{n=1}^{N}\mathbf{U}_n(t,0)\Big\| \leq \big(2e(t) + 8ce(t)\big)\kappa = \tilde{c}_1\kappa e(t)$$

with $\tilde{c}_1 = 2 + 8c$. In the same way, the remaining terms in the above expression for $\mathcal{P}^\perp(\boldsymbol{\mathcal{Y}}(t))\dot{\boldsymbol{\mathcal{X}}}(t)$ are bounded by $\tilde{c}_2\nu e(t)$, and hence we have

$$\|\mathcal{P}^\perp(\boldsymbol{\mathcal{Y}}(t))\dot{\boldsymbol{\mathcal{X}}}(t)\| \leq (\tilde{c}_1\kappa + \tilde{c}_2\nu)e(t)\,.$$

Using this inequality in (4.8) and also the second estimate of Lemma 3.2, we obtain the differential inequality, as long as $e(t) \leq c\rho$,

$$\dot{e} \leq \gamma e^2 + \varepsilon \tag{4.10}$$

with $\gamma = \rho^{-1}C(\tilde{c}_1\kappa + \tilde{c}_2\nu)$. With $c_1 = C\tilde{c}_1$ and $c_2 = C\tilde{c}_2$, the result now follows as at the end of the proof of Theorem 4.2. $\square$

4.4. Low-rank approximation of tensor differential equations. For the low-rank approximation to a solution of the tensor differential equation

$$\dot{\boldsymbol{\mathcal{A}}} = F(\boldsymbol{\mathcal{A}}), \tag{4.11}$$

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

$$\dot{\boldsymbol{\mathcal{Y}}} \in \mathcal{T}_{\boldsymbol{\mathcal{Y}}}\mathcal{M}_{\mathbf{r}} \quad\text{such that}\quad \|\dot{\boldsymbol{\mathcal{Y}}} - F(\boldsymbol{\mathcal{Y}})\| = \min! \tag{4.12}$$

Equivalently, condition (2.9) is replaced by the Galerkin condition

$$\langle \dot{\mathcal{Y}} - F(\mathcal{Y}), \mathcal{V} \rangle = 0 \quad \text{for all} \ \ \mathcal{V} \in \mathcal{T}_{\mathcal{Y}} \mathcal{M}_{\mathbf{r}}, \tag{4.13}$$

and correspondingly, the expression $\dot{\mathcal{A}}$ is replaced by $F(\mathcal{Y})$ for $\mathcal{Y} = \mathcal{S} \, \mathsf{X}_{n=1}^{N} \, \mathbf{U}_N \in \mathcal{M}_{\mathbf{r}}$ in the differential equations (2.11).

Theorems 4.1–4.3 extend to the low-rank approximation of matrix differential equations (4.11). We assume that F has a moderate bound along the approximations,

$$\|F(\mathcal{X}(t))\| \leq \mu, \qquad \|F(\mathcal{Y}(t))\| \leq \mu \qquad \text{for} \ \ 0 \leq t \leq \bar{t}, \tag{4.14}$$

and satisfies a one-sided Lipschitz condition: there is a real λ (positive or negative or zero) such that

$$\langle F(\mathcal{Y}) - F(\mathcal{X}), \mathcal{Y} - \mathcal{X} \rangle \leq \lambda \, \|\mathcal{Y} - \mathcal{X}\|^2 \tag{4.15}$$

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

$$\|F(\mathcal{X}(t)) - F(\mathcal{A}(t))\| \leq L \, \|\mathcal{X}(t) - \mathcal{A}(t)\| \qquad \text{for} \ \ 0 \leq t \leq \bar{t}, \tag{4.16}$$

which is in particular satisfied if F is Lipschitz continuous with Lipschitz constant L. Furthermore, assume that the best-approximation error is bounded by $\|\mathcal{X}(t) - \mathcal{A}(t)\| \leq c\rho$ for $0 \leq t \leq \bar{t}$, with c of Lemma 3.2.

We then have the following extension of the quasi-optimality result of Theorem 4.1.

THEOREM 4.4. *Suppose that a continuously differentiable best approximation* $\mathcal{X}(t) \in \mathcal{M}_{\mathbf{r}}$ *to a solution* $\mathcal{A}(t)$ *of (4.11) exists for* $0 \leq t \leq \bar{t}$, *and assume the bounds (4.14)–(4.16). Let* $\rho > 0$ *be such that the smallest nonzero singular value of the nth unfolding of* $\mathcal{X}(t)$ *satisfies* $\sigma_{r_n}(\mathbf{X}_{(n)}(t)) \geq \rho$ *for* $n = 1, \ldots, N$, *and assume that the best-approximation error is bounded by* $\|\mathcal{X}(t) - \mathcal{A}(t)\| \leq c\rho$ *with* c *of* Lemma 3.2, *for* $0 \leq t \leq \bar{t}$. *Then, the approximation error of (4.13) with initial value* $\mathcal{Y}(0) = \mathcal{X}(0)$ *is bounded in the Frobenius norm by*

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

for $t \leq \bar{t}$ *and as long as the right-hand side is bounded by* $c\rho$.

Proof. Equation (4.13) rewritten as in (3.1) reads

$$\dot{\mathcal{Y}} = \mathcal{P}(\mathcal{Y})F(\mathcal{Y}). \tag{4.17}$$

As in the proof of Theorem 4.1, we have the equation

$$\dot{\mathcal{X}} = \mathcal{P}(\mathcal{X})F(\mathcal{A}) + \mathcal{D} \qquad \text{with} \quad \|\mathcal{D}\| \leq 2\beta d$$

for $d = \|\mathcal{X} - \mathcal{A}\|$. We subtract the two equations, write

$$\mathcal{P}(\mathcal{Y})F(\mathcal{Y}) - \mathcal{P}(\mathcal{X})F(\mathcal{A}) - \mathcal{D} = (\mathcal{P}(\mathcal{Y}) - \mathcal{P}(\mathcal{X}))F(\mathcal{X}) + \mathcal{P}(\mathcal{X})(F(\mathcal{X}) - F(\mathcal{A}))$$
$$+ (F(\mathcal{Y}) - F(\mathcal{X})) - \mathcal{P}^{\perp}(\mathcal{Y})(F(\mathcal{Y}) - F(\mathcal{X})) - \mathcal{D},$$

and take the inner product with $\mathcal{Y} - \mathcal{X}$. With Lemma 3.2 we obtain

$$\langle \dot{\mathcal{Y}} - \dot{\mathcal{X}}, \mathcal{Y} - \mathcal{X} \rangle \leq \beta \, \|\mathcal{Y} - \mathcal{X}\|^2 + Ld \, \|\mathcal{Y} - \mathcal{X}\|$$
$$+ \lambda \, \|\mathcal{Y} - \mathcal{X}\|^2 + \beta \, \|\mathcal{Y} - \mathcal{X}\|^2 + 2\beta d \, \|\mathcal{Y} - \mathcal{X}\|.$$

For $e = \|\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}\|$ this gives the differential inequality

$$\dot{e} \leq (2\beta + \lambda)e + (2\beta + L)d, \qquad e(0) = 0, \tag{4.18}$$

which yields the result. $\quad\square$

We refer to [10, Theorem 4.1] for a related quasi-optimality result in a situation of a linear differential equation with an unbounded operator.

In the differential equation analogue of Theorem 4.2 with the splitting (4.3), we start from the equations $\dot{\boldsymbol{\mathcal{Y}}} - \dot{\boldsymbol{\mathcal{X}}} = \mathcal{P}(\boldsymbol{\mathcal{Y}})F(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}})\dot{\boldsymbol{\mathcal{X}}}$ and $\dot{\boldsymbol{\mathcal{X}}} = F(\boldsymbol{\mathcal{A}}) - \dot{\boldsymbol{\mathcal{E}}}$, yielding

$$\begin{aligned}
\dot{\boldsymbol{\mathcal{Y}}} - \dot{\boldsymbol{\mathcal{X}}} &= (\mathcal{P}(\boldsymbol{\mathcal{Y}}) - \mathcal{P}(\boldsymbol{\mathcal{X}}))\dot{\boldsymbol{\mathcal{X}}} - \mathcal{P}^{\perp}(\boldsymbol{\mathcal{Y}})(F(\boldsymbol{\mathcal{Y}}) - F(\boldsymbol{\mathcal{X}})) \\
&\quad + (F(\boldsymbol{\mathcal{Y}}) - F(\boldsymbol{\mathcal{X}})) + \mathcal{P}(\boldsymbol{\mathcal{Y}})(F(\boldsymbol{\mathcal{X}}) - F(\boldsymbol{\mathcal{A}})) + \mathcal{P}(\boldsymbol{\mathcal{Y}})\dot{\boldsymbol{\mathcal{E}}},
\end{aligned}$$

where we now take the inner product with $\boldsymbol{\mathcal{Y}} - \boldsymbol{\mathcal{X}}$. If it is additionally assumed that F has Lipschitz constant L, then this leads to the differential inequality

$$\dot{e} \leq C\rho^{-1}(\beta + L)e^2 + \lambda e + Ld + \varepsilon, \qquad e(0) = 0. \tag{4.19}$$

With $\widehat{\gamma} = C\rho^{-1}(\beta + L)$ and $\widehat{\varepsilon} = \varepsilon + L\max_{0 \leq t \leq \bar{t}} d(t)$, and with $\varphi(x) = (e^x - 1)/x$, this yields the error bound

$$\|\boldsymbol{\mathcal{Y}}(t) - \boldsymbol{\mathcal{X}}(t)\| \leq 2t\,\varphi(\lambda t)\,\widehat{\varepsilon} \qquad \text{for} \quad t\varphi(\lambda t) \leq \tfrac{1}{2}(\widehat{\gamma}\widehat{\varepsilon})^{-1/2} \tag{4.20}$$

and as long as $t \leq \bar{t}$ and $2t\varphi(\lambda t)\widehat{\varepsilon} \leq c\rho$.

Theorem 4.3 is extended similarly.

5. Numerical Experiments. To illustrate the theoretical error analysis, we compute the dynamical low rank approximation of a time-dependent tensor $\boldsymbol{\mathcal{A}}(t) \in \mathbb{R}^{15 \times 15 \times 15 \times 15}$, which is elaborately constructed to give control over the order of magnitude of the truncation error of the approximation: With $U = (U_1, \ldots, U_4)$ a set of orthogonal matrices in $\mathbb{R}^{15 \times 10}$, $\boldsymbol{\mathcal{B}}$ a random core tensor in $\mathbb{R}^{10 \times 10 \times 10 \times 10}$, and $\boldsymbol{\mathcal{C}} \in \mathbb{R}^{15 \times 15 \times 15 \times 15}$ a random perturbation, the time-dependent data tensor is defined as

$$\boldsymbol{\mathcal{A}}(t) = \exp(t)\boldsymbol{\mathcal{B}} \times_1 U_1 \times_2 \cdots \times_4 U_4 + \varepsilon(t + 1 + \sin(3t))\boldsymbol{\mathcal{C}}, \quad t \in [0, 1]. \tag{5.1}$$

We compute the dynamical tensor approximation (2.9), where $\boldsymbol{\mathcal{Y}}$ is defined with a core tensor of dimension $10 \times 10 \times 10 \times 10$. As time integrator (starting at a best approximation computed by alternating least squares) of the equations (2.11), we use the implicit midpoint rule with step size $h = 10^{-4}$, where the nonlinear equations in each step are solved by fixed point iteration. Figure 5 shows the approximation errors as functions of t for (5.1) with $\varepsilon \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. We observe that indeed the error of the approximation is proportional to ε and grows only moderately as a function of t (note the logarithmic scaling of the vertical axis).
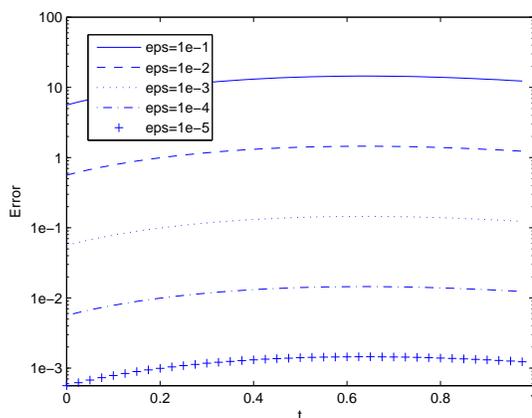
FIG. 5.1. *Error of the approximation of rank* $(10, 10, 10, 10)$ *for* $\mathcal{A}$ *from* (5.1) *with* $\varepsilon \in \{10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}\}$. *Interval* $t \in [0, 1]$, *step-size* $h = 10^{-4}$.

REFERENCES

[1] 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. Rep., 324 (2000), pp. 1–105.

[2] V. DE SILVA AND L.-H. LIM, *Tensor rank and the ill-posedness of the best low-rank approximation problem*, SIAM J. Matrix Anal. Appl., 30 (2008), pp. 1084–1172.

[3] L. GRASEDYCK, *Existence and computation of a low Kronecker-rank approximant to the solution of a tensor system with tensor right-hand side*, Computing, 72 (2004), pp. 247–265.

[4] W. HACKBUSCH AND B. KHOROMSKIJ, *Tensor-product approximation of operators and functions in high dimensions*, J. Complexity, 23 (2007), pp. 137–168.

[5] R.A. HORN AND C.R. JOHNSON, *Matrix Analysis*, Cambridge Univ. Press, Cambridge, 1985.

[6] O. KOCH AND C. LUBICH, *Dynamical low rank approximation*, SIAM J. Matrix Anal. Appl., 29 (2007), pp. 434–454.

[7] T. G. KOLDA AND B. W. BADER, *Tensor decompositions and applications*, Sandia Report SAND2007-6702, Sandia National Laboratories, Albuquerque, New Mexico 87185 and Livermore, California 94550, 2007.

[8] L. DE LATHAUWER, B. DE MOOR, AND J. VANDEWALLE, *A multilinear singular value decomposition*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1253–1278.

[9] ———, *On the best rank-1 and rank-$(R_1, \ldots, R_n)$ approximation of higher-order tensors*, SIAM J. Matrix Anal. Appl., 21 (2000), pp. 1324–1342.

[10] C. LUBICH, *On variational approximations in quantum molecular dynamics*, Math. Comp., 74 (2005), pp. 765–779.

[11] H.-D. MEYER, F. GATTI, AND G. A. WORTH, eds., *Multidimensional Quantum Dynamics: MCTDH Theory and Applications*, Wiley-VCH, Weinheim, Berlin, 2009.

[12] A. NONNENMACHER AND C. LUBICH, *Dynamical low rank approximation: Applications and numerical experiments*, Math. Comput. Simulation, 79 (2009), pp. 1346–1357.

[13] J. SUN, S. PAPADIMITRIOU, AND P. S. YU, *Window-based tensor analysis on high-dimensional and multi-aspect streams*, in ICDM '06: Proceedings of the Sixth International Conference on Data Mining, Washington, DC, USA, 2006, IEEE Computer Society, pp. 1076–1080.

[14] J. SUN, D. TAO, AND C. FALOUTSOS, *Beyond streams and graphs: dynamic tensor analysis*, in KDD '06: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2006, ACM, pp. 374–383.

[15] T. ZHANG AND G. GOLUB, *Rank-one approximation to high order tensors*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 534–550.

Preprint Series DFG-SPP 1324

http://www.dfg-spp1324.de

Reports

[1] R. Ramlau, G. Teschke, and M. Zhariy. A Compressive Landweber Iteration for Solving Ill-Posed Inverse Problems. Preprint 1, DFG-SPP 1324, September 2008.

[2] G. Plonka. The Easy Path Wavelet Transform: A New Adaptive Wavelet Transform for Sparse Representation of Two-dimensional Data. Preprint 2, DFG-SPP 1324, September 2008.

[3] E. Novak and H. Woźniakowski. Optimal Order of Convergence and (In-) Tractability of Multivariate Approximation of Smooth Functions. Preprint 3, DFG-SPP 1324, October 2008.

[4] M. Espig, L. Grasedyck, and W. Hackbusch. Black Box Low Tensor Rank Approximation Using Fibre-Crosses. Preprint 4, DFG-SPP 1324, October 2008.

[5] T. Bonesky, S. Dahlke, P. Maass, and T. Raasch. Adaptive Wavelet Methods and Sparsity Reconstruction for Inverse Heat Conduction Problems. Preprint 5, DFG-SPP 1324, January 2009.

[6] E. Novak and H. Woźniakowski. Approximation of Infinitely Differentiable Multivariate Functions Is Intractable. Preprint 6, DFG-SPP 1324, January 2009.

[7] J. Ma and G. Plonka. A Review of Curvelets and Recent Applications. Preprint 7, DFG-SPP 1324, February 2009.

[8] L. Denis, D. A. Lorenz, and D. Trede. Greedy Solution of Ill-Posed Problems: Error Bounds and Exact Inversion. Preprint 8, DFG-SPP 1324, April 2009.

[9] U. Friedrich. A Two Parameter Generalization of Lions' Nonoverlapping Domain Decomposition Method for Linear Elliptic PDEs. Preprint 9, DFG-SPP 1324, April 2009.

[10] K. Bredies and D. A. Lorenz. Minimization of Non-smooth, Non-convex Functionals by Iterative Thresholding. Preprint 10, DFG-SPP 1324, April 2009.

[11] K. Bredies and D. A. Lorenz. Regularization with Non-convex Separable Constraints. Preprint 11, DFG-SPP 1324, April 2009.

[12] M. Döhler, S. Kunis, and D. Potts. Nonequispaced Hyperbolic Cross Fast Fourier Transform. Preprint 12, DFG-SPP 1324, April 2009.

[13] C. Bender. Dual Pricing of Multi-Exercise Options under Volume Constraints. Preprint 13, DFG-SPP 1324, April 2009.

[14] T. Müller-Gronbach and K. Ritter. Variable Subspace Sampling and Multi-level Algorithms. Preprint 14, DFG-SPP 1324, May 2009.

[15] G. Plonka, S. Tenorth, and A. Iske. Optimally Sparse Image Representation by the Easy Path Wavelet Transform. Preprint 15, DFG-SPP 1324, May 2009.

[16] S. Dahlke, E. Novak, and W. Sickel. Optimal Approximation of Elliptic Problems by Linear and Nonlinear Mappings IV: Errors in L_2 and Other Norms. Preprint 16, DFG-SPP 1324, June 2009.

[17] B. Jin, T. Khan, P. Maass, and M. Pidcock. Function Spaces and Optimal Currents in Impedance Tomography. Preprint 17, DFG-SPP 1324, June 2009.

[18] G. Plonka and J. Ma. Curvelet-Wavelet Regularized Split Bregman Iteration for Compressed Sensing. Preprint 18, DFG-SPP 1324, June 2009.

[19] G. Teschke and C. Borries. Accelerated Projected Steepest Descent Method for Nonlinear Inverse Problems with Sparsity Constraints. Preprint 19, DFG-SPP 1324, July 2009.

[20] L. Grasedyck. Hierarchical Singular Value Decomposition of Tensors. Preprint 20, DFG-SPP 1324, July 2009.

[21] D. Rudolf. Error Bounds for Computing the Expectation by Markov Chain Monte Carlo. Preprint 21, DFG-SPP 1324, July 2009.

[22] M. Hansen and W. Sickel. Best m-term Approximation and Lizorkin-Triebel Spaces. Preprint 22, DFG-SPP 1324, August 2009.

[23] F.J. Hickernell, T. Müller-Gronbach, B. Niu, and K. Ritter. Multi-level Monte Carlo Algorithms for Infinite-dimensional Integration on $\mathbb{R}^{\mathbb{N}}$. Preprint 23, DFG-SPP 1324, August 2009.

[24] S. Dereich and F. Heidenreich. A Multilevel Monte Carlo Algorithm for Lévy Driven Stochastic Differential Equations. Preprint 24, DFG-SPP 1324, August 2009.

[25] S. Dahlke, M. Fornasier, and T. Raasch. Multilevel Preconditioning for Adaptive Sparse Optimization. Preprint 25, DFG-SPP 1324, August 2009.

[26] S. Dereich. Multilevel Monte Carlo Algorithms for Lévy-driven SDEs with Gaussian Correction. Preprint 26, DFG-SPP 1324, August 2009.

[27] G. Plonka, S. Tenorth, and D. Roşca. A New Hybrid Method for Image Approximation using the Easy Path Wavelet Transform. Preprint 27, DFG-SPP 1324, October 2009.

[28] O. Koch and C. Lubich. Dynamical Low-rank Approximation of Tensors. Preprint 28, DFG-SPP 1324, November 2009.