Dynamical low-rank tensor approximations to high-dimensional parabolic problems: existence and convergence of spatial discretizations

Markus Bachmayr, Henrik Eisenmann, André Uschmajew

Angaben zur Veröffentlichung / Publication details:

Bachmayr, Markus, Henrik Eisenmann, and André Uschmajew. 2025. "Dynamical low-rank tensor approximations to high-dimensional parabolic problems: existence and convergence of spatial discretizations." *Numerische Mathematik* 157 (3): 781–822. https://doi.org/10.1007/s00211-025-01465-8.

Numerische Mathematik (2025) 157:781–822
https://doi.org/10.1007/s00211-025-01465-8

Numerische
Mathematik

Dynamical low-rank tensor approximations to high-dimensional parabolic problems: existence and convergence of spatial discretizations

Markus Bachmayr[1] · Henrik Eisenmann[1] · André Uschmajew[2]

Received: 14 September 2023 / Revised: 3 February 2025 / Accepted: 19 March 2025 /
Published online: 23 April 2025

Abstract
We consider dynamical low-rank approximations to parabolic problems on higher-order tensor manifolds in Hilbert spaces. In addition to existence of solutions and their stability with respect to perturbations to the problem data, we show convergence of spatial discretizations. Our framework accommodates various standard low-rank tensor formats for multivariate functions, including tensor train and hierarchical tensors.

Mathematics Subject Classification Primary 35K15 · 35R01; Secondary 15A69 · 65M12

1 Introduction

Dynamical low-rank approximation (DLRA) is a nonlinear model for time evolution of high-dimensional functions on low-dimensional submanifolds. In its simplest form for matrices, as presented in the seminal work [20], one seeks approximate solutions of matrix valued ODEs

$$\dot{X}(t) = F(X(t), t), \quad X(t) \in \mathbb{R}^{M \times N}$$

✉ Henrik Eisenmann
eisenmann@igpm.rwth-aachen.de

Markus Bachmayr
bachmayr@igpm.rwth-aachen.de

André Uschmajew
andre.uschmajew@uni-a.de

1 Institut für Geometrie und Praktische Mathematik, RWTH Aachen University, Templergraben 55, 52062 Aachen, Germany

2 Institute of Mathematics & Centre for Advanced Analytics and Predictive Sciences, University of Augsburg, 86159 Augsburg, Germany

constrained to a low-rank model

$$X(t) = U(t)V(t)^{\mathsf{T}},$$

where $U(t) \in \mathbb{R}^{M\times r}$ and $V(t) \in \mathbb{R}^{N\times r}$. Hence $X(t)$ has rank at most r for every time t, which if $r \ll M, N$ allows for an efficient numerical treatment. From a geometric perspective, by enforcing $X(t)$ to have rank exactly r, the problem can be formulated as an ODE on the manifold $\mathcal{M}_r$ of matrices of fixed rank r by projecting the vector field F onto the tangent space $T_{X(t)}\mathcal{M}_r$ at every time t,

$$\dot{X}(t) = P_{X(t)}F(X(t), t). \tag{1.1}$$

In this way, assuming a starting value X_0 on $\mathcal{M}_r$, the time evolution is automatically constrained to the manifold. This constrained problem admits the time-dependent variational formulation

$$\langle \dot{X}(t), Y\rangle = \langle F(X(t), t), Y\rangle \quad \text{for all } Y \in T_{X(t)}\mathcal{M}_r, \tag{1.2}$$

which in physics is also called the Dirac–Frenkel principle [9, 15]. The state-of-the-art tool for the numerical solution of these equations is based on a splitting of the tangent space projector $P_{X(t)}$, see [24].

The DLRA approach can also be applied to time-dependent problems where the solution $X(t)$ is a tensor of size $N_1 \times \cdots \times N_d$. In this case, various low-rank models are possible, notably those based on tree tensor networks [1, 21, 26]. The most basic example is the low-rank Tucker format

$$X(t; i_1, \ldots, i_d) = \sum_{j_1=1}^{r_1} \cdots \sum_{j_d=1}^{r_d} C(t; j_1, \ldots, j_d)U_1(t; i_1, j_1) \cdots U_d(t; j_d, i_d)$$

with multilinear rank $\mathbf{r} = (r_1, \ldots, r_d)$. Another example is the widely used tensor train model [29], also known as matrix product states in physics, in which the tensor $X(t)$ element-wise reads as

$$X(t; i_1, \ldots, i_d) = G_1(t; i_1)G_2(t; i_2) \cdots G_d(t; i_d)\,. \tag{1.3}$$

Here $G_\mu(t, i_\mu)$ for $\mu = 1, \ldots, d$ is a matrix of fixed size $k_{\mu-1} \times k_\mu$, with the convention that $k_0 = k_d = 1$. Hence, the parameters in this model are the so-called core tensors $G_\mu(t) \in \mathbb{R}^{k_{\mu-1}\times N_\mu \times k_\mu}$. The componentwise minimal possible tuple $\mathbf{k} = (k_1, \ldots, k_{d-1})$ is called the TT rank of X, and the set of tensors with a fixed TT rank $\mathbf{k} = (k_1, \ldots, k_{d-1})$ can be shown to form an embedded manifold $\mathcal{M}_{\mathbf{k}}$ in $\mathbb{R}^{N_1\times\cdots\times N_d}$, see [18, 33]. Note that for $d = 2$, one recovers matrices of rank $k = k_1$. The DLRA formulation then essentially takes the same form (1.1) or (1.2) with the according tangent space projection, and the resulting problem again can be solved numerically using a projector splitting approach [7, 25]. We refer to [34] for a survey on DLRA for matrix and tensor problems.

So far we have outlined DLRA for problems in finite-dimensional matrix and tensor spaces. In applications of DLRA related to PDE problems, such as in [1, 5, 6, 8, 10, 11, 16, 19, 23, 27], a finite-dimensional problem appears as the second step after a spatial discretization of multivariate functions. Then the natural question arises whether the solutions of discrete problems converge to a meaningful solution of a continuous DLRA formulation when the spatial discretization is refined. To the best of our knowledge, this convergence issue has not been addressed in the existing literature. Since DLRA is based on a nonlinear variational formulation, the convergence of spatial discretizations is a nontrivial matter and cannot be treated by classical results from the numerical analysis of PDEs.

1.1 Dynamical low-rank approximation for parabolic problems

In this work, we establish such results for certain types of parabolic PDEs. As a model, we consider the diffusion equation

$$
\begin{aligned}
\frac{\partial}{\partial t}u(x,t) - \nabla_x \cdot (B(t)\nabla_x u(x,t)) &= f(x,t) \qquad &&\text{for } (x,t) \in \Omega \times (0,T),\\
u(x,t) &= 0 &&\text{for } (x,t) \in \partial\Omega \times (0,T),\\
u(x,0) &= u_0(x) &&\text{for } x \in \Omega
\end{aligned}
\tag{1.4}
$$

on the product domain $\Omega = (0,1)^d$. Here $B(t)$ is a $d \times d$-matrix, which allows for anisotropic diffusion in (1.4), and we assume it to be uniformly bounded and positive definite, as well as Lipschitz continuous with respect to t. The problem (1.4) is typically formulated in weak form on function spaces as follows: given $u_0 \in L_2(\Omega)$ and $f \in L_2(0,T;H^{-1}(\Omega))$, find

$$
u \in W(0,T;H_0^1(\Omega),H^{-1}(\Omega)) = \{u \in L_2(0,T;H_0^1(\Omega)) \colon u' \in L_2(0,T;H^{-1}(\Omega))\}
$$

such that for almost all $t \in (0,T)$,

$$
\begin{aligned}
\langle u'(t), v\rangle + a(u(t), v; t) = \langle f(t), v\rangle \quad \text{for all } v \in H_0^1(\Omega),\\
u(0) = u_0.
\end{aligned}
\tag{1.5}
$$

Here, by $\langle \cdot, \cdot \rangle$ we denote the dual pairing on $H^{-1}(\Omega) \times H_0^1(\Omega)$, and $a \colon H_0^1(\Omega) \times H_0^1(\Omega) \times [0,T] \to \mathbb{R}$ is the bounded, symmetric and coercive bilinear form

$$
a(u,v;t) = \int_\Omega (B(t)\nabla u(x)) \cdot \nabla v(x)\,dx.
$$

Classical theory provides a unique solution to (1.5); see for example [37, Theorem 23.A] and [38, Theorem 30.A].

A DLRA formulation for (1.5) is obtained by selecting a low-rank model class

$$
\mathcal{M} \subset L_2(\Omega)
$$

based on the tensor product structure of $L_2(\Omega) = L_2(0,1) \otimes \cdots \otimes L_2(0,1)$, such as the tensor train format in a Hilbert space setting to be made precise in Sect. 4, and restricting the test function in (1.5) to the tangent space of $\mathcal{M}$ at $u(t)$. In other words, the problem reads: given $u_0 \in \mathcal{M}$ and $f \in L_2(0,T; H^{-1}(\Omega))$, find

$$u \in W(0,T; H_0^1(\Omega), H^{-1}(\Omega)) := \{u \in L_2(0,T; H_0^1(\Omega)) : u' \in L_2(0,T; H^{-1}(\Omega))\}$$

such that $u(t) \in \mathcal{M}$ for all $t \in (0,T)$ and

$$\begin{aligned} \langle u'(t), v\rangle + a(u(t), v; t) &= \langle f(t), v\rangle \quad \text{for all } v \in T_{u(t)}\mathcal{M} \cap H_0^1(\Omega), \\ u(0) &= u_0. \end{aligned} \tag{1.6}$$

For the matrix case $d = 2$, existence and uniqueness of DLRA solutions on a maximal time interval $(0, T^*)$ has been shown in [3] under the additional regularity assumptions $u_0 \in H_0^1(\Omega)$ and $f \in L_2(0,T; L_2(\Omega))$. The proof given there is carried out in a more general framework of parabolic problems on manifolds in Gelfand triplets and is based on a variational time stepping scheme in Hilbert space. As we will verify in Sect. 4, this framework also applies to the tensor train model of d-variate functions, and can analogously be used to other tree based low-rank tensor models. To this end, we establish a general concept of tensor manifolds in Hilbert space which allows to deduce the required manifold properties, such as curvature estimates, from their finite-dimensional counterparts. As a result, we obtain a meaningful and rigorous notion of a continuous DLRA solution for parabolic problems for higher-order tensors based on the results from [3].

However, the existence proof in [3] does not provide the convergence of solutions of spatial semidiscretizations to the continuous solution, and we address this open problem in the present work as well. Spatial discretizations of low-rank problems can be obtained in a very natural way via tensor products of discretization spaces. By considering finite-dimensional subspaces $V_h^\mu \subset H_0^1(0,1)$ of dimension N_μ one obtains a DLRA problem analogous to (1.6) by restricting to the tensor product space

$$\mathcal{V}_h = V_h^1 \otimes \cdots \otimes V_h^d.$$

We hence seek a solution $u_h(t) \in \mathcal{M} \cap \mathcal{V}_h$ satisfying

$$\begin{aligned} \langle u_h'(t), v_h\rangle + a(u_h(t), v_h; t) &= \langle f(t), v_h\rangle \quad \text{for all } v_h \in T_{u_h(t)}\mathcal{M} \cap \mathcal{V}_h, \\ u(0) &= u_0. \end{aligned} \tag{1.7}$$

By writing $u_h(t)$ in a tensor product basis of $\mathcal{V}_h$,

$$u_h(t) = \sum_{i_1=1}^{N_1} \cdots \sum_{i_d=1}^{N_d} X(t; i_1, \ldots, i_d) \varphi_{i_1}^1 \otimes \cdots \otimes \varphi_{i_d}^d,$$

we obtain a DLRA problem for the coefficient tensor $X(t) \in \mathbb{R}^{N_1\times\cdots\times N_d}$. The question is then whether for $h \to 0$, the functions u_h converge to the (unique) solution u of (1.7).

1.2 Contributions and outline

In this work, we extend the existence and uniqueness result for the DLRA evolution obtained in [3] from the bivariate to the multivariate case. The abstract framework from [3] is recalled in Sect. 2. This section also contains the new stability estimate Theorem 2.5, which complements the uniqueness result of [3]. The main result of Sect. 3 is the convergence result for spatial semidiscretizations (Theorem 3.2) in the abstract framework of Gelfand triplets as developed in [3]. In Sect. 4, we present a general notion of low-rank manifolds in Hilbert spaces and obtain new curvature estimates. We then in Sect. 5 apply these results to dynamical low-rank approximations of the model problem (1.6) using the tensor train format. Based on the general setting of Sect. 4, analogous results can be obtained for other low-rank manifolds; in particular, our results apply directly also to the Tucker format. Note that in this work, we focus on a theoretical framework for DLRA of parabolic problems and do not consider specific numerical realizations. Numerical results for DLRA in the parabolic case can be found, e.g., in [19, 28].

2 Abstract formulation

For developing the essential aspects of the theory we investigate the DLRA problem in an abstract context as in [3]. We consider a Gelfand triplet

$$\mathcal{V} \hookrightarrow \mathcal{H} \cong \mathcal{H}^* \hookrightarrow \mathcal{V}^*$$

of Hilbert spaces, where $\mathcal{V}$ is compactly embedded in $\mathcal{H}$. This implies that the embedding is also continuous, that is,

$$\|u\|_{\mathcal{H}} \lesssim \|u\|_{\mathcal{V}}. \tag{2.1}$$

By $\langle\cdot,\cdot\rangle : \mathcal{V}^* \times \mathcal{V} \to \mathbb{R}$ we denote the dual pairing of $\mathcal{V}^*$ and $\mathcal{V}$. Note that for $u \in \mathcal{H} \subset \mathcal{V}^*$ and $v \in \mathcal{V} \subset \mathcal{H}$, the dual pairing and the inner product on $\mathcal{H}$ coincide, that is, $\langle u, v\rangle_{\mathcal{H}} = \langle u, v\rangle$. We will frequently identify $u \in \mathcal{V}$ as an element of $\mathcal{H}$ and in turn also as an element in $\mathcal{V}^*$.

For every $t \in [0, T]$, let $a(\cdot,\cdot;t) : \mathcal{V}\times\mathcal{V} \to \mathbb{R}$ be a bilinear form which is assumed to be symmetric,

$$a(u, v; t) = a(v, u; t) \quad \text{for all } u, v \in \mathcal{V} \text{ and } t \in [0, T],$$

uniformly bounded,

$$|a(u, v; t)| \le \beta\|u\|_{\mathcal{V}}\|v\|_{\mathcal{V}} \quad \text{for all } u, v \in \mathcal{V} \text{ and } t \in [0, T]$$

for some $\beta > 0$, and uniformly coercive,

$$a(u, u; t) \geq \mu \|u\|_{\mathcal{V}}^2 \quad \text{for all } u \in \mathcal{V} \text{ and } t \in [0, T]$$

for some $\mu > 0$. Under these assumptions, $a(\cdot, \cdot; t)$ is an inner product on $\mathcal{V}$ defining an equivalent norm. Furthermore, it defines a bounded operator

$$A(t) : \mathcal{V} \to \mathcal{V}^* \tag{2.2}$$

such that

$$a(u, v; t) = \langle A(t)u, v \rangle \quad \text{for all } u, v \in \mathcal{V}.$$

We also assume that $a(u, v; t)$ is Lipschitz continuous with respect to t. In other words, there exists an $L \geq 0$ such that

$$|a(u, v; t) - a(u, v; s)| \leq L\beta \|u\|_{\mathcal{V}} \|v\|_{\mathcal{V}} |t - s| \tag{2.3}$$

for all $u, v \in \mathcal{V}$ and $s, t \in [0, T]$, which in the model problem corresponds to the Lipschitz continuity of the function $t \mapsto B(t)$.

We deal with evolution equations on subsets $\mathcal{M} \subset \mathcal{H}$ that are submanifolds in the following sense: for every point $u \in \mathcal{M}$ there exists a closed subspace $T_u\mathcal{M} \subset \mathcal{H}$ such that $T_u\mathcal{M}$ contains all tangent vectors to $\mathcal{M}$ at u. Here a tangent vector is any $v \in \mathcal{H}$ for which there exists a (strongly) differentiable curve $\varphi \colon (-\varepsilon, \varepsilon) \to \mathcal{H}$ (for some $\varepsilon > 0$) such that $\varphi(t) \in \mathcal{M}$ for all t and

$$\varphi(0) = u, \quad \varphi'(0) = v.$$

By $P_u : \mathcal{H} \to T_u\mathcal{M}$ we denote the $\mathcal{H}$-orthogonal projection onto $T_u\mathcal{M}$. We will also assume $\mathcal{M} \cap \mathcal{V}$ to be nonempty as well as $T_u\mathcal{M} \cap \mathcal{V}$ to be nonempty for $u \in \mathcal{M} \cap \mathcal{V}$. By $\overline{\mathcal{M}}^w$ we denote the weak closure of $\mathcal{M}$ in $\mathcal{H}$.

The abstract problem takes the following form.

Problem 2.1 In the above setting, given $f \in L_2(0, T; \mathcal{V}^*)$ and $u_0 \in \mathcal{M} \cap \mathcal{H}$, find

$$u \in W(0, T; \mathcal{V}, \mathcal{V}^*) := \{u \in L_2(0, T; \mathcal{V}) \colon u' \in L_2(0, T; \mathcal{V}^*)\}$$

such that for almost all $t \in [0, T]$,

$$\begin{aligned} u(t) &\in \mathcal{M}, \\ \langle u'(t), v \rangle + a(u(t), v; t) &= \langle f(t), v \rangle \quad \text{for all } v \in T_{u(t)}\mathcal{M} \cap \mathcal{V}, \\ u(0) &= u_0. \end{aligned} \tag{2.4}$$

2.1 Basic assumptions and existence of solutions

The main challenge of the weak formulation in Problem 2.1 is that according to the Dirac-Frenkel principle, the test functions are from the tangent space only. The existence result in [3] requires several assumptions, including additional regularity of the data as in assumption **A0** below. The other assumption **A1**–**A4** are abstractions of corresponding properties of the model problem and will be discussed for the tensor train format in Sect. 4, and hence the main results of this paper apply to this setting. The assumptions are the following.

A0 (Regularity of data) We have $f \in L_2(0, T; \mathcal{H})$ and $u_0 \in \mathcal{M} \cap \mathcal{V}$.

A1 (Cone property) $\mathcal{M}$ is a cone, that is, $u \in \mathcal{M}$ implies $su \in \mathcal{M}$ for all $s > 0$.

A2 (Curvature bound) For every subset $\mathcal{M}'$ of $\mathcal{M}$ that is weakly compact in $\mathcal{H}$, there exists a constant $\kappa = \kappa(\mathcal{M}')$ such that

$$\|P_u - P_v\|_{\mathcal{H}\to\mathcal{H}} \leq \kappa \|u - v\|_{\mathcal{H}}$$

and

$$\|(I - P_u)(u - v)\|_{\mathcal{H}} \leq \kappa \|u - v\|_{\mathcal{H}}^2$$

for all $u, v \in \mathcal{M}'$.

A3 (Compatibility of tangent spaces)

(a) For $u \in \mathcal{M} \cap \mathcal{V}$ and $v \in T_u\mathcal{M} \cap \mathcal{V}$ an admissible curve with $\varphi(0) = u$, $\varphi'(0) = v$ can be chosen such that

$$\varphi(t) \in \mathcal{M} \cap \mathcal{V}$$

for all $|t|$ small enough.

(b) If $u \in \mathcal{M} \cap \mathcal{V}$ and $v \in \mathcal{V}$ then $P_u v \in T_u\mathcal{M} \cap \mathcal{V}$.

A4 (Operator splitting) The associated operator $A(t)$ in (2.2) admits a splitting

$$A(t) = A_1(t) + A_2(t)$$

into two uniformly bounded operators $\mathcal{V} \to \mathcal{V}^*$ such that for all $t \in [0, T]$, all $u \in \mathcal{M} \cap \mathcal{V}$ and all $v \in \mathcal{V}$, the following holds:

(a) "$A_1(t)$ maps to the tangent space":

$$\langle A_1(t)u, v\rangle = \langle A_1(t)u, P_u v\rangle.$$

(b) "$A_2(t)$ is locally bounded from $\mathcal{M} \cap \mathcal{V}$ to $\mathcal{H}$": For every subset $\mathcal{M}'$ of $\mathcal{M}$ that is weakly compact in $\mathcal{H}$, there exists $\gamma = \gamma(\mathcal{M}') > 0$ such that

$$A_2(t)u \in \mathcal{H} \quad \text{and} \quad \|A_2(t)u\|_{\mathcal{H}} \leq \gamma \|u\|_{\mathcal{V}}^{\eta} \quad \text{for all } u \in \mathcal{M}'$$

with an $\eta > 0$ independent of $\mathcal{M}'$.

The following existence and uniqueness result has been obtained in [3, Theorem 4.3]. Here $W(0,T;\mathcal{V},\mathcal{H})$ denotes the subspace of $W(0,T;\mathcal{V},\mathcal{H})$ with $u' \in L_2(0,T;\mathcal{H})$.

Theorem 2.2 *Let the Assumptions* **A0**–**A4** *hold and let u_0 have positive $\mathcal{H}$-distance from $\overline{\mathcal{M}}^w \setminus \mathcal{M}$. There exist $T^* \in (0,T]$ and $u \in W(0,T^*;\mathcal{V},\mathcal{H}) \cap L_\infty(0,T^*;\mathcal{V})$ such that u solves Problem 2.1 on the time interval $[0,T^*]$, and its continuous representative $u \in C(0,T^*;\mathcal{H})$ satisfies $u(t) \in \mathcal{M}$ for all $t \in [0,T^*)$. Here T^* is maximal for the evolution on $\mathcal{M}$ in the sense that if $T^* < T$, then*

$$\liminf_{t\to T^*} \inf_{v\in\overline{\mathcal{M}}^w\setminus\mathcal{M}} \|u(t)-v\|_{\mathcal{H}} = 0.$$

In either case, u is the unique solution of Problem 2.1 in $W(0,T^*;\mathcal{V},\mathcal{H}) \cap L_\eta(0,T^*;\mathcal{V})$.

In particular, let $\sigma = \operatorname{dist}_{\mathcal{H}}(u_0, \overline{\mathcal{M}}^w \setminus \mathcal{M})$, then there exists a constant $c > 0$ such that $T^* \geq \min(\sigma^2/c, T)$.

The solution satisfies the following estimates:

$$\|u\|^2_{L_2(0,T^*;\mathcal{V})} \leq \|u_0\|^2_{\mathcal{H}} + C_1\|f\|^2_{L_2(0,T^*;\mathcal{H})}, \tag{2.5}$$

$$\|u'\|^2_{L_2(0,T^*;\mathcal{H})} \leq C_2\left(\|u_0\|^2_{\mathcal{V}} + \|f\|^2_{L_2(0,T^*;\mathcal{H})}\right), \tag{2.6}$$

$$\|u\|^2_{L_\infty(0,T^*;\mathcal{V})} \leq C_3\left(\|u_0\|^2_{\mathcal{V}} + \|f\|^2_{L_2(0,T^*;\mathcal{H})}\right), \tag{2.7}$$

where C_1, C_2, and C_3 are the constants from [3, Lemma 4.4].

Note that the energy estimates (2.5)–(2.7) are not explicitly stated in [3, Theorem 4.3], but immediately follow from its proof in combination with [3, Lemma 4.4].

Remark 2.3 We can take c as the right-hand side of (2.6).

Remark 2.4 Our Assumption **A2** is stronger than the one used in [3], which requires the curvature estimates to be valid for $\|u - v\|_{\mathcal{H}} \leq \varepsilon$ for some $\varepsilon > 0$ (the constant κ then may depend on ε). The stronger assumption made here is used in the proof of Theorem 2.5. As our new curvature estimates in Sect. 4.1.4 show, this stronger assumption is in fact satisfied for the low-rank manifolds under consideration.

2.2 A stability estimate

As a first extension to the above existence and uniqueness theorem, we now provide a stability estimate that in particular ensures continuity of the solution with respect to the data. This result was obtained in [12]. The proof follows a similar idea as the uniqueness result in [3, Theorem 4.1].

Theorem 2.5 *Let $u, v \in W(0,T^*;\mathcal{V},\mathcal{H})$ be two solutions of Problem 2.1 on a time interval $[0,T^*]$ corresponding to right-hand sides $f, g \in L_2(0,T;\mathcal{H})$, and initial values $u_0, v_0 \in \mathcal{M}$, respectively. Assume that the continuous representatives $u, v \in$*

$C(0, T^*; \mathcal{H})$ have values in a weakly compact subset $\mathcal{M}' \subset \mathcal{M}$ (in particular their $\mathcal{H}$-distance to $\overline{\mathcal{M}}^w \setminus \mathcal{M}$ remains bounded from below). Moreover, assume that $u, v \in L_\eta(0, T^*; \mathcal{V})$ where η is from Assumption **A4**(b). *Then for any $\varepsilon > 0$,*

$$\|u(t) - v(t)\|_{\mathcal{H}}^2 \leq \left(\|u_0 - v_0\|_{\mathcal{H}}^2 + \frac{1}{\varepsilon} \int_0^t \|f(s) - g(s)\|_{\mathcal{H}}^2 \, ds \right) \exp(\Lambda(t) + \varepsilon t),$$

where

$$\begin{aligned}\Lambda(t) := 2\kappa \int_0^t & \|u'(s)\|_{\mathcal{H}} + \|v'(s)\|_{\mathcal{H}} + \gamma \left(\|u(s)\|_{\mathcal{V}}^\eta + \|v(s)\|_{\mathcal{V}}^\eta \right) + \|f(s)\|_{\mathcal{H}} \\ & + \|g(s)\|_{\mathcal{H}} \, ds < \infty\end{aligned}$$

with $\kappa = \kappa(\mathcal{M}')$ from Assumption **A2**.

Proof We use integration by parts in the sense of [37, Proposition 23.23(iv)]. This results in

$$\begin{aligned}\frac{1}{2} \frac{d}{dt} \|u(t) - v(t)\|_{\mathcal{H}}^2 \leq \langle & u'(t) - v'(t) + A(t)(u(t) - v(t)) - f(t) + g(t) \\ & + f(t) - g(t), u(t) - v(t) \rangle\end{aligned}$$

for almost all $t \in [0, T^*]$ by coercivity of $A(t)$ and adding and subtracting $\langle f(t) - g(t), u(t) - v(t) \rangle$. We add and subtract (2.4) for the solutions u and v with $w = P_{v(t)}(u(t) - v(t))$ and $w = P_{u(t)}(u(t) - v(t))$, respectively. This results in

$$\begin{aligned}&\frac{1}{2} \frac{d}{dt} \|u(t) - v(t)\|_{\mathcal{H}}^2 \\ &\quad \leq \langle f(t) - g(t), u(t) - v(t) \rangle + \langle u'(t) + A(t)u(t) - f(t), (\mathrm{id} - P_{u(t)})(u(t) - v(t)) \rangle \\ &\quad\quad - \langle v'(t) + A(t)v(t) - g(t), (\mathrm{id} - P_{v(t)})(u(t) - v(t)) \rangle.\end{aligned}$$

We use Young's inequality to estimate

$$\langle f(t) - g(t), u(t) - v(t) \rangle \leq \frac{1}{2\varepsilon} \|f(t) - g(t)\|_{\mathcal{H}}^2 + \frac{\varepsilon}{2} \|u(t) - v(t)\|_{\mathcal{H}}^2$$

and Assumption **A4** to get

$$\begin{aligned}&\frac{1}{2} \frac{d}{dt} \|u(t) - v(t)\|_{\mathcal{H}}^2 \\ &\quad \leq \left(\|u'(t)\|_{\mathcal{H}} + \gamma \|u(t)\|_{\mathcal{V}}^\eta + \|f(t)\|_{\mathcal{H}} \right) \|(\mathrm{id} - P_{u(t)})(u(t) - v(t))\|_{\mathcal{H}} \\ &\quad\quad + \left(\|v'(t)\|_{\mathcal{H}} + \gamma \|v(t)\|_{\mathcal{V}}^\eta + \|g(t)\|_{\mathcal{H}} \right) \|(\mathrm{id} - P_{v(t)})(u(t) - v(t))\|_{\mathcal{H}} \\ &\quad\quad + \frac{1}{2\varepsilon} \|f(t) - g(t)\|_{\mathcal{H}}^2 + \frac{\varepsilon}{2} \|u(t) - v(t)\|_{\mathcal{H}}^2.\end{aligned}$$

Finally, Assumption **A2** implies

$$\frac{d}{dt}\|u(t)-v(t)\|_{\mathcal{H}}^2 \le \Big(2\kappa\Big(\|u'(t)\|_{\mathcal{H}}+\|v'(t)\|_{\mathcal{H}}+\gamma\big(\|u(t)\|_{\mathcal{V}}^{\eta}+\|u(t)\|_{\mathcal{V}}^{\eta}\big) \\ +\|f(t)\|_{\mathcal{H}}+\|g(t)\|_{\mathcal{H}}\Big)+\varepsilon\Big)\|u(t)-v(t)\|_{\mathcal{H}}^2+\frac{1}{\varepsilon}\|f(t)-g(t)\|_{\mathcal{H}}^2$$

and the result follows from Grönwall's lemma; see for example [31, Lemma 2.7]. Here we take into account that $L_2(0,T^*;\mathcal{H})\subset L_1(0,T^*;\mathcal{H})$. □

3 Convergence of spatial discretizations

From the perspective of numerical analysis, an important question is whether the unique solution of Problem 2.1 can be obtained as the limit of solutions of spatially discretized problems. We now provide such a result under assumptions on the compatibility of the discrete spaces $\mathcal{V}_h\subset\mathcal{V}$ with $\mathcal{M}$. The spatially discretized problems are of the following form.

Problem 3.1 Given $f\in L_2(0,T;\mathcal{H})$ and $u_{0,h}\in\mathcal{M}\cap\mathcal{V}_h$, find $u_h\in W(0,T;\mathcal{V},\mathcal{H})$ such that for almost all $t\in[0,T]$,

$$\begin{aligned} u_h(t)&\in\mathcal{M}\cap\mathcal{V}_h,\\ \langle u_h'(t),v_h\rangle+a(u_h(t),v_h;t)&=\langle f(t),v_h\rangle \quad\text{for all } v_h\in T_{u_h(t)}\mathcal{M}\cap\mathcal{V}_h,\\ u_h(0)&=u_{h,0}. \end{aligned} \tag{3.1}$$

We require that the discrete subspaces $\mathcal{V}_h\subset\mathcal{V}$ have the following properties.

B1 (Approximation property)

(a) For every $v\in\mathcal{V}$, the $\mathcal{V}$-orthogonal projections $v_h\in\mathcal{V}_h$ satisfy $\|v-v_h\|_{\mathcal{V}}\to 0$ as $h\searrow 0$.
(b) For every $u\in\mathcal{M}\cap\mathcal{V}$ there is a sequence (u_h) with $u_h\in\mathcal{M}\cap\mathcal{V}_h$ such that u_h converges to u in $\mathcal{V}$ as $h\searrow 0$ and $\|u_h\|_{\mathcal{V}}\le\|u\|_{\mathcal{V}}$.

B2 (Compatibility of tangent spaces)

(a) For $u_h\in\mathcal{M}\cap\mathcal{V}_h$ and $v_h\in T_u\mathcal{M}\cap\mathcal{V}_h$ a continuously differentiable curve with $\varphi(0)=u_h$, $\varphi'(0)=v_h$ can be chosen such that

$$\varphi(t)\in\mathcal{M}\cap\mathcal{V}_h$$

for all $|t|$ small enough.
(b) If $u_h\in\mathcal{M}\cap\mathcal{V}_h$ and $v_h\in\mathcal{V}_h$ then $P_{u_h}v_h\in T_{u_h}\mathcal{M}\cap\mathcal{V}_h$.

The model problem (1.5) allows for such a space discretization, as verified in Sect. 5.1.3 for tensor train manifolds. The following result was obtained in [12].

Theorem 3.2 *Let the Assumptions* **A0–A4** *and* **B1–B2** *hold. Let $u_{0,h} \in \mathcal{M} \cap \mathcal{V}_h$ define a sequence that converges to u_0 in $\mathcal{V}$ as $h \searrow 0$ and let u_0 have positive $\mathcal{H}$-distance σ to the relative boundary $\overline{\mathcal{M}}^{\mathsf{w}} \setminus \mathcal{M}$. Then there exists a constant $c > 0$ independent of σ and a constant $h_0 > 0$ such that there is a unique u_h in $W(0, T^*; \mathcal{V}, \mathcal{H}) \cap L_\eta(0, T^*; \mathcal{V})$ that solves Problem 3.1 on the time interval $[0, T^*]$ when $T^* < \sigma^2/c$ for all $h \le h_0$. Furthermore, u_h converges to the unique solution u of Problem 2.1 in $W(0, T^*; \mathcal{V}, \mathcal{H}) \cap L_\eta(0, T^*; \mathcal{V})$ weakly in $L_2(0, T^*; \mathcal{V})$ and strongly in $C(0, T^*; \mathcal{H})$, while the weak derivatives u_h' converge weakly to u' in $L_2(0, T^*, \mathcal{H})$.*

Proof Since $u_{0,h}$ converges to u_0 in $\mathcal{V}$, there is an $h_0 > 0$ such that $\|u_{0,h} - u_0\|_{\mathcal{V}} \le \sigma/2$ and $\|u_{0,h} - u_0\|_{\mathcal{H}} \le \sigma/2$ for all $h \le h_0$ due to (2.1). Furthermore, we can choose h_0 small enough, such that $\|u_{0,h}\|_{\mathcal{V}}^2 \le 2\|u_0\|_{\mathcal{V}}^2$ and $\|u_{0,h}\|_{\mathcal{H}}^2 \le 2\|u_0\|_{\mathcal{H}}^2$. Therefore, the $\mathcal{H}$-distance of $u_{0,h}$ from $\overline{\mathcal{M}}^{\mathsf{w}} \setminus \mathcal{M}$ is at least $\sigma/2$. Hence, applying Theorem 2.2 with $\mathcal{V}_h$ in place of $\mathcal{V}$ provides us with solutions u_h to Problem 3.1 on a time interval $[0, T^*]$ with $T^* < \sigma^2/(4c)$ for every $h \le h_0$, where c can be chosen as the right-hand side of the following estimate (3.3). Theorem 2.2 provides us with the estimates

$$\|u_h\|_{L_2(0,T^*;\mathcal{V})}^2 \le 2\|u_0\|_{\mathcal{H}}^2 + C_1\|f\|_{L_2(0,T^*;\mathcal{H})}^2, \tag{3.2}$$

$$\|u_h'\|_{L_2(0,T^*;\mathcal{H})}^2 \le C_2\left(2\|u_0\|_{\mathcal{V}}^2 + \|f\|_{L_2(0,T^*;\mathcal{H})}^2\right), \tag{3.3}$$

$$\|u_h\|_{L_\infty(0,T^*;\mathcal{V})}^2 \le C_3\left(2\|u_0\|_{\mathcal{V}}^2 + \|f\|_{L_2(0,T^*;\mathcal{H})}^2\right). \tag{3.4}$$

Note that by (3.3), we can assume that for h sufficiently small, $\|u_h(t) - u_0\|_{\mathcal{H}} \le \sigma - \delta$ for a $\delta > 0$. As a consequence, there is a subsequence (u_h) converging weakly to $\tilde{u}$ in $L_2(0, T^*; \mathcal{V})$ and weakly* in $L_\infty(0, T^*; \mathcal{V})$ and the derivatives (u_h') converging weakly to $\tilde{w}$ in $L_2(0, T^*; \mathcal{H})$.

We next show that $\tilde{w}$ is the weak derivative of $\tilde{u}$. For this, we need to verify that

$$\int_0^{T^*} \langle \tilde{w}(t), v\rangle\, \phi(t) + \langle \tilde{u}(t), v\rangle\, \phi'(t)\, dt = 0$$

for arbitrary $v \in \mathcal{V}$ and $\phi \in C_0^\infty(0, T^*)$. For any $v_h \in \mathcal{V}_h$ we may add and subtract the weak derivative of u_h to obtain

$$\begin{aligned}&\int_{T^*} \langle \tilde{w}(t), v_h\rangle\, \phi(t)\, dt + \langle \tilde{u}(t), v_h\rangle\, \phi'(t)\, dt\\ &\quad = \int_{T^*} \langle \tilde{w}(t) - u_h'(t), v_h\rangle\, \phi(t)\; + \langle \tilde{u}(t) - u_h(t), v_h\rangle\, \phi'(t)\, dt.\end{aligned}$$

Now let (v_h) be a sequence converging to v in $\mathcal{V}$. Then

$$\begin{aligned}&\int_{T^*} \langle \tilde{w}(t), v\rangle\, \phi(t)\, dt + \langle \tilde{u}(t), v\rangle\, \phi'(t)\, dt = \lim_{h\searrow 0} \int_{T^*} \langle \tilde{w}(t), v_h\rangle\, \phi(t)\, dt + \langle \tilde{u}(t), v_h\rangle\, \phi'(t)\, dt\\ &= \lim_{h\searrow 0} \int_{T^*} \langle \tilde{w}(t) - u_h'(t), v_h\rangle\, \phi(t)\; + \langle \tilde{u}(t) - u_h(t), v_h\rangle\, \phi'(t)\, dt = 0\end{aligned}$$

since $v_h\phi$ converges strongly to $v\phi$ in $L_2(0, T^*; \mathcal{V})$. Therefore, the sequence (u_h) converges weakly in $W(0, T^*; \mathcal{V}, \mathcal{H})$ to $\tilde{u}$. Due to the Aubin-Lions theorem, and by boundedness in $L_\infty(0, T^*; \mathcal{V})$, it also converges strongly in $C(0, T^*; \mathcal{H})$ to $\tilde{u}$, and $\tilde{u}(0) = \lim_{h\searrow 0} u_h(0) = \lim_{h\searrow 0} u_{0,h} = u_0$.

It remains to show that $\tilde{u}$ satisfies (2.4) and therefore agrees with the unique solution u of Problem 2.1 in $W(0, T^*; \mathcal{V}, \mathcal{H}) \cap L_\eta(0, T^*; \mathcal{V})$ provided by Theorem 2.2. By a subsequence-of-subsequence argument, it then follows that the entire sequence converges to $\tilde{u}$. Let Q_h be the $\mathcal{V}$-orthogonal projection onto $\mathcal{V}_h$. For $v \in T_{\tilde{u}(t)}\mathcal{M} \cap \mathcal{V}$, let $v_h = Q_h v$, which converges strongly to v in $\mathcal{V}$ by Property **B1**(a). This also implies that the sequence is uniformly bounded in $\mathcal{H}$. By (3.1), we have

$$\langle u_h'(t), P_{u_h(t)} v_h \rangle + a(u_h(t), P_{u_h(t)} v_h; t) = \langle f(t), P_{u_h(t)} v_h \rangle$$

for almost every t, since $P_{u_h(t)} v_h \in T_{u_h(t)}\mathcal{M} \cap \mathcal{V}_h$ by Property **B2**(b). We have chosen the time interval such that $u_h(t) \in \mathcal{M}' \subset \mathcal{M}$ lie in a weakly compact subset for all $t \in [0, T^*]$. Hence, using Assumption **A2**,

$$\begin{aligned} \|v - P_{u_h(t)} v_h\|_{\mathcal{H}} &\le \|v - P_{u_h(t)} v\|_{\mathcal{H}} + \|P_{u_h(t)}(v - v_h)\|_{\mathcal{H}} \\ &\le \kappa(\mathcal{M}')\|\tilde{u}(t) - u_h(t)\|_{\mathcal{H}}\|v\|_{\mathcal{H}} + \|(v - v_h)\|_{\mathcal{H}}, \end{aligned} \tag{3.5}$$

and thus $P_{u_h(t)} v_h$ converges strongly to v in $\mathcal{H}$. Using a similar argument as in the proof of Theorem 2.2 in [3], it suffices to show

$$\int_0^{T^*} \langle \tilde{u}'(t), v(t) \rangle + a(\tilde{u}(t), v(t); t) - \langle f(t), v(t) \rangle \, dt = 0$$

for all $v \in L_\infty(0, T^*; \mathcal{V})$ with $v(t) \in T_{\tilde{u}(t)}\mathcal{M} \cap \mathcal{V}$ for almost every t.

Since $P_{u_h(t)} Q_h v(t)$ converges to $v(t)$ in $\mathcal{H}$ for almost all $t \in [0, T^*]$, and we have the square integrable bound (3.5), the sequence $P_{u_h(t)} Q_h v(t)$ converges strongly to v in $L_2(0, T^*; \mathcal{H})$. This together with weak convergence of (u_h') in $L_2(0, T^*; \mathcal{H})$ implies

$$\begin{aligned} \lim_{h\searrow 0} \int_0^{T^*} &\langle u_h'(t), P_{u_h(t)} Q_h v(t) \rangle - \langle f(t), P_{u_h(t)} Q_h v(t) \rangle \, dt \\ &= \int_0^{T^*} \langle \tilde{u}'(t), v(t) \rangle - \langle f(t), v(t) \rangle \, dt. \end{aligned}$$

Finally, we use Assumption **A4**. We have

$$\begin{aligned} &a(u_h(t), P_{u_h(t)} Q_h v(t); t) - a(\tilde{u}(t), v(t); t) \\ &\quad = \langle A_1(t) u_h(t), P_{u_h(t)} Q_h v(t) \rangle - \langle A_1(t) \tilde{u}(t), v(t) \rangle \\ &\qquad + \langle A_2(t) u_h(t), P_{u_h(t)} Q_h v(t) \rangle - \langle A_2(t) \tilde{u}(t), v(t) \rangle \end{aligned} \tag{3.6}$$

and due to Assumption **A4**(a)

$$\langle A_1(t)u_h(t), P_{u_h(t)}Q_h v(t)\rangle = \langle A_1(t)u_h(t), Q_h v(t)\rangle.$$

This implies

$$\lim_{h\searrow 0}\int_0^{T^*} \left|\langle A_1(t)u_h(t), P_{u_h(t)}Q_h v(t)\rangle - \langle A_1(t)u(t), v(t)\rangle\right| \, dt = 0$$

as u_h converges weakly to $\tilde{u}$ and $Q_h v$ converges strongly to v in $L_2(0, T^*; \mathcal{V})$. For the second summand in (3.6) we have

$$\begin{aligned}&\langle A_2(t)u_h(t), P_{u_h(t)}Q_h v(t)\rangle - \langle A_2(t)\tilde{u}(t), v(t)\rangle \\ &\quad = \langle A_2(t)u_h(t), P_{u_h(t)}Q_h v(t) - v(t)\rangle + \langle A_2(t)(\tilde{u}(t) - u_h(t)), v(t)\rangle\end{aligned}$$

where

$$\left|\langle A_2(t)u_h(t), P_{u_h(t)}Q_h v(t) - v(t)\rangle\right| \le \gamma \|u_h(t)\|_{\mathcal{V}}^{\eta}\|P_{u_h(t)}Q_h v(t) - v(t)\|_{\mathcal{H}}$$

and $\int_0^{T^*} \|u_h(t)\|_{\mathcal{V}}^{\eta}\|P_{u_h(t)}Q_h v(t) - v(t)\|_{\mathcal{H}}\, dt \to 0$. Moreover,

$$\int_0^{T^*} \langle A_2(t)(\tilde{u}(t) - u_h(t)), v(t)\rangle \, dt \to 0 \quad \text{as}\, h \searrow 0,$$

since u_h converges weakly to $\tilde{u}$ in $L_2(0, T^*; \mathcal{V})$. Taken together with (3.5) and the uniform bound of u_h in $L_\infty(0, T^*; \mathcal{V})$, we have

$$\int_0^{T^*} a(u_h(t), P_{u_h(t)}Q_h v(t); t) - a(\tilde{u}(t), v(t); t)\, dt \to 0 \quad \text{as}\, h \searrow 0$$

and hence

$$\begin{aligned}&\int_0^{T^*} \langle \tilde{u}'(t), v(t)\rangle + a(\tilde{u}(t), v(t); t) - \langle f(t), v(t)\rangle \, dt \\ &\quad = \lim_{h\searrow 0}\int_0^{T^*} \langle u_h'(t), P_{u_h(t)}v_h\rangle + a(u_h(t), P_{u_h(t)}v_h; t) - \langle f(t), P_{u_h(t)}v_h\rangle \, dt = 0\end{aligned}$$

for all $v \in L_\infty(0, T^*; \mathcal{V})$ with $v(t) \in T_{\tilde{u}(t)}\mathcal{M} \cap \mathcal{V}$ for almost every t. □

4 Properties of low-rank tensor manifolds in Hilbert space

In this section we return to our model problem (1.4) in its weak formulation (1.5) and apply the theory developed above for low-rank models of multivariate functions. In the model problem $\mathcal{H} = L_2(\Omega)$ and $\mathcal{V} = H_0^1(\Omega)$, the compact embedding $\mathcal{V} \hookrightarrow \mathcal{H}$ is due to the Rellich-Kondrachov theorem and (2.1) is the Poincaré inequality.

4.1 Low-rank tensor manifolds in function space

Let $\Omega = \Omega_1 \times \cdots \times \Omega_d$, where Ω_μ is a bounded domain in a Euclidean space for $\mu = 1, \dots, d$. We write $H^\mu = L_2(\Omega_\mu)$ for abbreviation. The space $\mathcal{H} = L_2(\Omega_1 \times \cdots \times \Omega_d) = H^1 \otimes \cdots \otimes H^d$ is a tensor product Hilbert space. In DLRA, one considers low-rank manifolds $\mathcal{M}$ in such spaces. We consider manifolds of the general form

$$\mathcal{M} = \Big\{ u = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} C(k_1, \dots, k_d)\, u^1_{k_1} \otimes \cdots \otimes u^d_{k_d} : \\ C \in \mathcal{M}_\mathrm{c} \subset \mathbb{R}^{r_1 \times \cdots \times r_d}_*,\ G(u^\mu) \text{ is invertible} \Big\}. \tag{4.1}$$

Here $\mathbb{R}^{r_1 \times \cdots \times r_d}_*$ denotes the dense and open subset of "regular" $r_1 \times \cdots \times r_d$ tensors with full multilinear rank $(r_1, \dots, r_d)$, and $G(u^\mu) = [\langle u^\mu_i, u^\mu_j \rangle]_{ij} \in \mathbb{R}^{r_\mu \times r_\mu}$ is the Gramian of the system $\{u^\mu_1, \dots, u^\mu_{r_\mu}\}$. We assume that $\mathcal{M}_\mathrm{c}$ is a smooth submanifold in $\mathbb{R}^{r_1 \times \cdots \times r_d}_*$ that we additionally assume to be invariant under changes of basis in the sense that for all $C \in \mathcal{M}_\mathrm{c}$,

$$C \times_1 T^1 \times_2 \cdots \times_d T^d \in \mathcal{M}_\mathrm{c} \quad \text{for all invertible matrices } T^1, \dots, T^d. \tag{4.2}$$

Here we use the notation $\times_\mu$ for left multiplication of a matrix onto the μ-th mode a tensor [22].

The definition of $\mathcal{M}$ results in a constrained version of the Tucker format (for which $\mathcal{M}_\mathrm{c} = \mathbb{R}^{r_1 \times \cdots \times r_d}_*$), and covers continuous versions of general tree based tensor formats, for example by letting $\mathcal{M}_\mathrm{c}$ be a corresponding finite-dimensional low-rank tensor manifold in $\mathbb{R}^{r_1 \times \cdots \times r_d}$, such as manifolds of tensor trains [18, 34] or of hierarchical Tucker tensors [4, 33] with fixed ranks.

Note that it follows from the assumed properties of $\mathcal{M}_\mathrm{c}$, that this set is not closed in $\mathbb{R}^{r_1 \times \cdots \times r_d}$, and hence $\overline{\mathcal{M}_\mathrm{c}} \setminus \mathcal{M}_\mathrm{c} \neq \emptyset$. To see this, let $C \in \mathcal{M}_\mathrm{c}$ and $T^1_n, \dots, T^d_n$ be invertible matrices converging to $T^1_*, \dots, T^d_*$ such that at least one of the limits is not invertible. Then $C \times_1 T^1_n \times_2 \cdots \times_d T^d_n \in \mathcal{M}_\mathrm{c}$ for all n by (4.2), but the limit $C \times_1 T^1_* \times_2 \cdots \times_d T^d_*$ does not have full multilinear rank, and hence is not in $\mathcal{M}_\mathrm{c}$. This example also shows that the set $\mathcal{M}$ is not closed in $\mathcal{H}$, and in particular not weakly closed. Moreover, in Lemma 4.4 we will see that the closure and weak closure of $\mathcal{M}$ coincide with the set described in (4.1) with $\mathcal{M}_\mathrm{c}$ replaced by $\overline{\mathcal{M}_\mathrm{c}}$.

For investigating the manifold properties of $\mathcal{M}$, the concepts of matricizations and minimal subspaces play a crucial role. For every $\mu = 1, \dots, d$, we can identify u with an element M^μ_u, called the μ-th matricization of u, in the subspace $H^\mu \otimes H^{\neq\mu}$, where $H^{\neq\mu} = \bigotimes_{\nu\neq\mu} H^\nu$. Assuming $u \in \mathcal{M}$ as above and letting

$$v^\mu_{k_\mu} = \sum_{k_1=1}^{r_1} \cdots \sum_{k_{\mu-1}=1}^{r_{\mu-1}} \sum_{k_{\mu+1}=1}^{r_{\mu+1}} \cdots \sum_{k_d=1}^{r_d} C(k_1, \dots, k_d)\, u^1_{k_1} \otimes \cdots \otimes u^{\mu-1}_{k_{\mu-1}} \otimes u^{\mu+1}_{k_{\mu+1}} \otimes \cdots \otimes u^d_{k_d}, \tag{4.3}$$

one has

$$M_u^\mu = \sum_{k_\mu=1}^{r_\mu} u_{k_\mu}^\mu \otimes v_{k_\mu}^\mu. \tag{4.4}$$

Since the core tensor C has full multilinear rank by (4.1), one can show that the $v_{k_\mu}^\mu$ are also linearly independent. Now define

$$\mathcal{U}^\mu = \operatorname{span}\{u_1^\mu, \ldots, u_{r_\mu}^\mu\}$$

and

$$\mathcal{V}^\mu = \operatorname{span}\{v_1^\mu, \ldots, v_{r_\mu}^\mu\},$$

then (4.4) expresses the fact, that M_u^μ is an element of the "matrix subspace" $\mathcal{U}^\mu \otimes \mathcal{V}^\mu$ and $\operatorname{rank}(M_u^\mu) = r_\mu$. We call $\mathcal{U}^\mu$ the μ-th minimal subspace of $u \in \mathcal{M}$.

Choosing an orthonormal basis for each space $H^\mu = L_2(\Omega_\mu)$, we obtain an isomorphism between H^μ and $\ell_2(\mathbb{N})$ for each μ. This in turn defines a tensor space isomorphism between $\mathcal{H}$ and $\ell_2(\mathbb{N}) \otimes \cdots \otimes \ell_2(\mathbb{N})$. In what follows, in order to use a more common matrix and tensor notation, we can thus assume without loss of generality that

$$\mathcal{H} = \ell_2(\mathbb{N}^d) = \ell_2(\mathbb{N}) \otimes \cdots \otimes \ell_2(\mathbb{N}),$$

and thus consider $\mathcal{M}$ as a set in the tensor product Hilbert space of square summable infinite arrays. The definition of $\mathcal{M}$ remains the same as in (4.1), only that now $u_{k_\mu}^\mu \in \ell_2(\mathbb{N})$.

We will, however, denote the elements of $\ell_2(\mathbb{N}^d)$ as X instead of u, in order to clearly distinguish these sequences from functions. The corresponding matricizations are $M_X^\mu \in \ell_2(\mathbb{N}) \otimes \ell_2(\mathbb{N}^{d-1})$. The Tucker format (4.1) can then be written in the usual abbreviated form

$$X = C \times_1 U^1 \times_2 \cdots \times_d U^d$$

where $U^\mu = [u_1^\mu, \ldots, u_{r_\mu}^\mu] \in (\ell_2(\mathbb{N}))^{r_\mu}$ contains a basis for $\mathcal{U}^\mu$. Here the multiplications $\times_\mu$ are defined as for finite tensors.

4.1.1 Manifold structure

Using the concept of manifolds in Banach space as presented in [36, Ch. 73] we can prove the following result.

Theorem 4.1 *Let $X = C \times_1 U^1 \times_2 \cdots \times_d U^d$ be in $\mathcal{M}$ defined as in (4.1) satisfying (4.2). Then the following statements hold.*

(i) *There exists an open neighborhood* $\mathcal{O}$ *of* X *and a submersion* g *defined on* $\mathcal{O}$ *such that* $\mathcal{M} \cap \mathcal{O} = g^{-1}(0)$. *Consequently,* $\mathcal{M} \cap \mathcal{O}$ *is a smooth submanifold in the Hilbert space* $\mathcal{H}$. *The tangent space* $T_X\mathcal{M}$ *at* $X \in \mathcal{M} \cap \mathcal{O}$ *is the null space of* $g'(X)$.

(ii) *There exists a continuously Fréchet-differentiable homeomorphism* φ *from a neighborhood of zero in* $T_{X_*}\mathcal{M}$ *to* $\mathcal{M} \cap \mathcal{O}$ *satisfying* $\varphi(\xi) = X_* + \xi + o(\|\xi\|_{\mathcal{H}})$ *for all* ξ *in that neighborhood. Moreover,* φ *is also an immersion and hence a local embedding for* $\mathcal{M}$.

(iii) *The tangent space equals the subspace spanned by elements of the form*

$$\xi = \dot{C} \times_1 U^1 \times_2 \cdots \times_d U^d + C \times_1 \dot{U}^1 \times_2 \cdots \times_d U^d + \cdots + C \times_1 U^1 \times_2 \cdots \times_d \dot{U}^d \tag{4.5}$$

with $\dot{C} \in T_C\mathcal{M}_{\mathrm{c}}$ *and* $(U^\mu)^\mathsf{T}\dot{U}^\mu = 0_{r_\mu \times r_\mu}$ *for* $\mu = 1, \ldots, d$ *(that is, the columns of* $\dot{U}^\mu$ *span a subspace orthogonal to* $\mathcal{U}^\mu$*).*

The proof of this theorem is given in the appendix. For an alternative treatment of low-rank manifolds in Banach spaces, see [13, 14].

4.1.2 Tangent space projection

We now consider the orthogonal projection onto the tangent space $T_X\mathcal{M}$ at given $X = C \times_1 U^1 \times_2 \cdots \times_d U^d$. By Theorem 4.1(iii), $T_X\mathcal{M}$ is spanned by elements $\xi = \xi_0 + \xi_1 + \cdots + \xi_d$ of the form (4.5). Here the elements $\xi_0 = \dot{C} \times_1 U^1 \times_2 \cdots \times_d U^d$ with $\dot{C} \in T_C\mathcal{M}_{\mathrm{c}}$ span a subspace of $\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^d$ which we denote by $\mathcal{S}_X$. For $\mu = 1, \ldots, d$, the elements $\xi_\mu = C \times_1 U^1 \times_2 \cdots \times_\mu \dot{U}^\mu \times_{\mu+1} \cdots \times_d U^d$ with $(U^\mu)^\mathsf{T}\dot{U}^\mu = 0$ are equivalently described via their matricization as

$$M^\mu_{\xi_\mu} = \dot{U}^\mu (V^\mu)^\mathsf{T} = \sum_{k_\mu=1}^{r_\mu} \dot{u}^\mu_{k_\mu} \otimes v^\mu_{k_\mu}$$

due to the definition (4.3) of V^μ. Since actually any element in the space $(\mathcal{U}^\mu)^\perp \otimes \mathcal{V}^\mu$ can be written in this way, the $M^\mu_{\xi_\mu}$ span this space. Treating the $(\mathcal{U}^\mu)^\perp \otimes \mathcal{V}^\mu$ as subspaces of $\ell_2(\mathbb{N}^d)$ (in a slight abuse of notation) we conclude that

$$T_X\mathcal{M} = \mathcal{S}_X \oplus [(\mathcal{U}^1)^\perp \otimes \mathcal{V}^1] \oplus \cdots \oplus [(\mathcal{U}^d)^\perp \otimes \mathcal{V}^d], \tag{4.6}$$

which indeed is an orthogonal decomposition as can be seen from the fact that $\mathcal{V}^\mu \subseteq \mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^{\mu-1} \otimes \mathcal{U}^{\mu+1} \otimes \cdots \otimes \mathcal{U}^d$ for every μ.

In the following proposition, we compute the tangent space projection under the assumption that the matrices U^μ have orthonormal columns; see Remark 4.3 for the general formula.

Proposition 4.2 *Let* $X = C \times_1 U_1 \times_2 \cdots \times_d U_d \in \mathcal{M}$, *and assume* $(U^\mu)^\mathsf{T}U^\mu = \mathrm{id}$. *The orthogonal projection onto the tangent space* $T_X\mathcal{M}$ *is given as*

$$P_X = P^0_X + P^1_X + \cdots + P^d_X \tag{4.7}$$

with $P_X^1, \dots, P_X^d$ being implicitly defined via their action on matricizations as

$$M^{\mu}_{P_X^{\mu}(Z)} = (I - P_{\mathcal{U}^{\mu}}) M_Z^{\mu} P_{\mathcal{V}^{\mu}}, \quad \mu = 1, \dots, d. \tag{4.8}$$

The projector P_X^0 is defined as

$$P_X^0(Z) = P_C(C_Z) \times_1 U^1 \times_2 \cdots \times_d U^d, \tag{4.9}$$

where P_C is the orthogonal tangent space projector onto $T_C\mathcal{M}_{\mathrm{c}}$ in $\mathbb{R}^{r_1 \times \cdots \times r_d}$, and $C_Z = Z \times_1 (U^1)^T \times_2 \cdots \times_d (U^d)^T$.

Proof By (4.6), the single terms $\xi_0, \dots, \xi_d$ in the tangent vector representation (4.5) belong to mutually orthogonal subspaces. Therefore, the orthogonal projection $P_X(Z) = \xi_0 + \xi_1 + \cdots + \xi_d$ onto $T_X\mathcal{M}$ can be decomposed accordingly as in (4.7). Here the P_X^{μ} for $\mu = 1, \dots, d$ are the projections on $(\mathcal{U}^{\mu})^{\perp} \otimes \mathcal{V}^{\mu}$ which have the asserted form. We consider the projection $\xi_0 = P_X^0(Z)$ of a given Z onto the space $\mathcal{S}_X$ in (4.6). We write

$$\xi_0 = K \times_1 U^1 \times_2 \cdots \times_d U^d$$

and need to determine $K \in T_C\mathcal{M}_{\mathrm{c}}$. The orthogonality condition for the projection is

$$0 = \langle Z - \xi_0, \dot{C} \times_1 U^1 \times_2 \cdots \times_d U^d \rangle$$

for all $\dot{C} \in T_C\mathcal{M}_{\mathrm{c}}$. Using properties of tensor-matrix multiplication, we rewrite this as

$$\begin{aligned}
0 &= \langle Z \times_1 (U^1)^{\mathsf{T}} \times_2 \cdots \times_d (U^d)^{\mathsf{T}} - \xi_0 \times_1 (U^1)^{\mathsf{T}} \times_2 \cdots \times_d (U^d)^{\mathsf{T}}, \dot{C} \rangle \\
&= \langle Z \times_1 (U^1)^{\mathsf{T}} \times_2 \cdots \times_d (U^d)^{\mathsf{T}} - K, \dot{C} \rangle = \langle C_Z - K, \dot{C} \rangle.
\end{aligned}$$

Since this holds for all $\dot{C} \in T_C\mathcal{M}_{\mathrm{c}}$, it follows that K equals the orthogonal projection of C_Z onto $T_C\mathcal{M}_{\mathrm{c}}$. □

Remark 4.3 If the U^{μ} are not orthonormal, then the formula for $P_X^0(Z)$ needs to be adjusted to

$$P_X^0(Z) = \Pi_C(C_Z),$$

where Π_C is the orthogonal projection in $\mathbb{R}^{r_1 \times \cdots \times r_d}$ onto $T_C\mathcal{M}_{\mathrm{c}}$ with respect to the inner product induced by the operator $\mathbf{A} = (U^1)^T U^1 \otimes \cdots \otimes (U^d)^T U^d$, which is symmetric and positive definite on $\mathbb{R}^{r_1 \times \cdots \times r_d}$. This projection is given by $\Pi_C = (P_C \mathbf{A} P_C)^{-1} P_C \mathbf{A}$, where $(P_C \mathbf{A} P_C)^{-1}$ denotes the inverse of $P_C \mathbf{A} P_C$ on $T_C\mathcal{M}_{\mathrm{c}}$.

4.1.3 Distance to boundary

As we will see in Sect. 4.1.4, curvature estimates as in **A2** for low-rank tensor manifolds can be expressed in terms of inverses of smallest singular values of certain matricizations. In this subsection, we therefore estimate the smallest singular values of matricizations of a tensor $X \in \mathcal{M}$ from below by its distance to the boundary $\overline{\mathcal{M}}^w \setminus \mathcal{M}$. This will have the effect, that on every weakly compact subset $\mathcal{M}' \subseteq \mathcal{M}$ these singular values remain bounded from below.

We first give a characterization of the weak closure of $\mathcal{M}$. In what follows, by $\|\cdot\|$ without further specification we denote the Frobenius norm of tensors.

Lemma 4.4 *Let $\mathcal{M}$ be of the form (4.1) with $\mathcal{M}_{\mathrm{c}}$ satisfying (4.2). Then*

$$\overline{\mathcal{M}}^w = \overline{\mathcal{M}} = \left\{X = C \times_1 U^1 \times_2 \cdots \times_d U^d : C \in \overline{\mathcal{M}_{\mathrm{c}}}, \quad (U^\mu)^\mathsf{T} U^\mu \in \mathrm{GL}_{\mathrm{r}_\mu}\right\},$$

that is, the weak closure and closure of $\mathcal{M}$ coincide and are of the form (4.1) with $\mathcal{M}_{\mathrm{c}}$ replaced by $\overline{\mathcal{M}_{\mathrm{c}}}$.

Proof Let $(X_n) \subset \mathcal{M}$ be a sequence converging weakly to $X \in \overline{\mathcal{M}}^w$. By [17, Thm. 6.29], there are r_μ dimensional subspaces $\mathcal{U}^\mu$ such that $X \in \mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^d$. In particular, let $U^\mu \in (\ell_2(\mathbb{N}))^{r_\mu}$ be orthonormal bases of the spaces $\mathcal{U}^\mu$. Then

$$X = \left(X \times_1 (U^1)^\mathsf{T} \times_2 \cdots \times_d (U^d)^\mathsf{T}\right) \times_1 U^1 \times_2 \cdots \times_d U^d = C \times_1 U^1 \times_2 \cdots \times_d U^d$$

with $C \in \mathbb{R}^{r_1 \times \cdots} \times r_d$. Moreover, since $X_n \in \mathcal{M}$, we have

$$X_n = C_n \times_1 U_n^1 \times_2 \cdots \times_d U_n^d$$

and by weak convergence, we have

$$\lim_{n\to\infty} C_n \times_1 (U^1)^\mathsf{T} U_n^1 \times_2 \cdots \times_d (U^d)^\mathsf{T} U_n^d = C.$$

By the invariance condition (4.2) for $\mathcal{M}_{\mathrm{c}}$, it follows that $C \in \overline{\mathcal{M}_{\mathrm{c}}}$. This shows

$$\overline{\mathcal{M}}^w = \left\{X = C \times_1 U^1 \times_2 \cdots \times_d U^d : C \in \overline{\mathcal{M}_{\mathrm{c}}}, \quad (U^\mu)^\mathsf{T} U^\mu \in \mathrm{GL}_{\mathrm{r}_\mu}\right\}.$$

Now let $(\tilde{C}_n) \subset \mathcal{M}_{\mathrm{c}}$ be a sequence converging to C. Then the sequence defined by

$$\tilde{X}_n = \tilde{C}_n \times_1 U^1 \times_2 \cdots \times_d U^d$$

converges strongly to X as $\|\tilde{X}_n - X\|_{\ell_2(\mathbb{N}^d)} = \|\tilde{C}_n - C\|$. This proves the assertion. ◻

For $X \in \mathcal{M}$ and $\mu = 1, \dots, d$, let $\{u_1^\mu, \dots, u_{r_\mu}^\mu\}$ be the left singular vectors of the matricization M_X^μ. Then

$$M_X^\mu = \sum_{k=1}^{r_\mu} u_k^\mu \otimes v_k^\mu$$

with $\{v_1^\mu, \dots, v_{r_\mu}^\mu\}$ orthogonal in $\ell_2(\mathbb{N}^{d-1})$ such that

$$\sigma_k^\mu = \sigma_k^\mu(X) = \|v_k^\mu\|_{\ell_2(\mathbb{N}^{d-1})}, \quad k = 1, \dots, r_\mu, \tag{4.10}$$

are the singular values of M_X^μ, for which we may assume

$$\sigma_1^\mu \geq \sigma_2^\mu \geq \ldots \geq \sigma_{r_\mu}^\mu.$$

Further, we define

$$\sigma = \operatorname{dist}(X, \overline{\mathcal{M}}^w \setminus \mathcal{M}). \tag{4.11}$$

Proposition 4.5 *Let σ_k^μ for $\mu = 1, \dots, d$ and $k = 1, \dots, r_\mu$ be defined as above. Then*

$$\min_{\mu \in \{1,\dots,d\}} \sigma_{r_\mu}^\mu \geq \sigma.$$

Proof Let $\mu \in \{1, \dots, d\}$ and let $u_1^\mu, \dots, u_{r_\mu}^\mu$ be the left singular vectors of M_X^μ associated to $\sigma_1^\mu, \dots, \sigma_{r_\mu}^\mu$, respectively. Then for the tensor $\tilde{X}$ defined by its matricization

$$M_{\tilde{X}}^\mu = \sum_{k=1}^{r_\mu - 1} u_k^\mu (u_k^\mu)^\top M_X^\mu,$$

we have $\|X - \tilde{X}\|_{\ell_2(\mathbb{N}^d)} = \sigma_{r_\mu}^\mu$, and $\tilde{X} \notin \mathcal{M}$. Furthermore, we have

$$X = C \times_1 U^1 \times_2 \cdots \times_d U^d,$$

with $U^\mu = [u_1^\mu, \dots, u_{r_\mu}^\mu]$ and

$$\begin{aligned} \tilde{X} &= C \times_1 U^1 \times_2 \cdots \times_\mu U^\mu (\tilde{U}^\mu)^\top U^\mu \times_{\mu+1} \cdots \times_d U^d \\ &= (C \times_\mu (\tilde{U}^\mu)^\top U^\mu) \times_1 U^1 \times_2 \cdots \times_d U^d \end{aligned}$$

with $\tilde{U}^\mu = [u_1^\mu, \dots, u_{r_\mu - 1}^\mu, 0]$. It follows from Lemma 4.4 that $\tilde{X} \in \overline{\mathcal{M}}$ and the claim is proven. □

It is important to note that the distance σ defined in (4.11) can be expressed as the distance of the core tensor C to the relative boundary of $\mathcal{M}_c$.

Proposition 4.6 *Let* $X = C \times_1 U_1 \times_2 \cdots \times_d U_d$ *with* $C \in \mathcal{M}_c$ *and orthonormal* $U_1, \ldots, U_d$. *Then* $\operatorname{dist}(C, \overline{\mathcal{M}_c} \backslash \mathcal{M}_c) = \operatorname{dist}(X, \overline{\mathcal{M}}^w \backslash \mathcal{M})$.

Proof First, let $Y \in \overline{\mathcal{M}}^w \backslash \mathcal{M}$ satisfy $\|X - Y\|_{\ell_2(\mathbb{N}^d)} = \operatorname{dist}(X, \overline{\mathcal{M}}^w \backslash \mathcal{M})$. Then the tensor $D = Y \times_1 U_1^\mathsf{T} \times_2 \cdots \times_d U_d^\mathsf{T}$ satisfies $\|C - D\| \le \|X - Y\|_{\ell_2(\mathbb{N}^d)}$ and by Lemma 4.4, we also have $D \in \overline{\mathcal{M}_c} \backslash \mathcal{M}_c$, and hence $\operatorname{dist}(C, \overline{\mathcal{M}_c} \backslash \mathcal{M}_c) \le \operatorname{dist}(X, \overline{\mathcal{M}}^w \backslash \mathcal{M})$. To show equality, we consider a $D \in \overline{\mathcal{M}_c} \backslash \mathcal{M}_c$ with $\|C - D\| = \operatorname{dist}(C, \overline{\mathcal{M}_c} \backslash \mathcal{M}_c)$. Set $Y = D \times_1 U_1 \times_2 \cdots \times_d U_d \in \overline{\mathcal{M}}^w \backslash \mathcal{M}$. Then $\|X - Y\|_{\ell_2(\mathbb{N}^d)} = \|C - D\|$ holds, and thus $\operatorname{dist}(C, \overline{\mathcal{M}_c} \backslash \mathcal{M}_c) \ge \operatorname{dist}(X, \overline{\mathcal{M}}^w \backslash \mathcal{M})$. □

4.1.4 Curvature estimates

We now turn to the curvature bounds in Assumption **A2**. We first derive an estimate for the norm difference $\|P_X - P_Y\|_{\ell_2(\mathbb{N}^d) \to \ell_2(\mathbb{N}^d)}$ of two such projections in operator norm, which can be regarded as a curvature estimate for the manifold $\mathcal{M}$. It will be required for Assumption **A2**.

Proposition 4.7 *Assume a curvature estimate*

$$\max_{\|Z\|=1} \|(P_C - P_{\tilde{C}})Z\| \le \frac{c}{\hat{\sigma}} \|C - \tilde{C}\| \quad \textit{for all } C, \tilde{C} \in \mathcal{M}_c, \tag{4.12}$$

where $\hat{\sigma} = \operatorname{dist}(C, \overline{\mathcal{M}_c} \backslash \mathcal{M}_c)$ *and where* $c > 0$ *is independent of* $C, \tilde{C}$. *Let* $X, Y \in \mathcal{M}$ *with corresponding tangent space projections* P_X *and* P_Y. *Then*

$$\|P_X - P_Y\|_{\ell_2(\mathbb{N}^d) \to \ell_2(\mathbb{N}^d)} \le \left(\frac{\sqrt{2}c}{\sigma} + 2(\sqrt{2} + 1) \sum_{\mu=1}^{d} \frac{1}{\sigma^\mu_{r_\mu}} \right) \|X - Y\|_{\ell_2(\mathbb{N}^d)}$$

where $\sigma = \operatorname{dist}(X, \overline{\mathcal{M}}^w \backslash \mathcal{M})$ *and* $\sigma^\mu_{r_\mu}$ *is the smallest singular value of the* μ*-th matricization.*

Note that $\sigma^\mu_{r_\mu} \ge \sigma$ for each μ as a consequence of Proposition 4.5. Therefore, we have the simpler estimate

$$\|P_X - P_Y\|_{\ell_2(\mathbb{N}^d) \to \ell_2(\mathbb{N}^d)} \le \frac{\sqrt{2}c + 2d\left(\sqrt{2} + 1\right)}{\sigma} \|X - Y\|_{\ell_2(\mathbb{N}^d)}.$$

Since on every weakly compact subset $\mathcal{M}' \subseteq \mathcal{M}$ the distance σ to the boundary is bounded from below (recall that $\mathcal{M}$ itself is not weakly closed), we obtain the first curvature estimate in **A2**. In the proof of Proposition 4.7, we use the following lemma.

Lemma 4.8 *Let* $U, V \in [\ell_2(\mathbb{N})]^r$ *be orthonormal (that is,* $U^\mathsf{T} U = V^\mathsf{T} V = \mathrm{id}$*) such that the* $r \times r$ *matrix* $U^\mathsf{T} V$ *is symmetric and positive semidefinite.*

(i) *The corresponding subspace projections* $P_{\mathcal{U}} = UU^{\mathsf{T}}$ *and* $P_{\mathcal{V}} = VV^{\mathsf{T}}$ *satisfy*

$$\|U_V\|_{\mathbb{R}^r \to \ell_2(\mathbb{N})} \leq \sqrt{2}\|P_{\mathcal{U}} - P_{\mathcal{V}}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})}.$$

(ii) *For all* $x, y \in \mathbb{R}^r$, $\|x - y\| \leq \sqrt{2}\|Ux - Vy\|_{\ell_2(\mathbb{N})}$.

Proof After orthogonal change of basis, we may assume the matrix $U^{\mathsf{T}}V = \Sigma$ to be diagonal with entries $1 \geq \sigma_i \geq 0$, that is $u_j^{\mathsf{T}} v_i = \sigma_i \delta_{ij}$.

Ad (i). We define the spaces $W_i = \operatorname{span}\{u_i, v_i\}$ for $i = 1, \ldots, r$. These are pairwise orthogonal. Furthermore, let $W_{r+1} = \left(\bigoplus_{i=1}^r W_i\right)^{\perp}$. Then the difference of projections is block-diagonal with respect to the spaces W_i, that is, $(P_{\mathcal{U}} - P_{\mathcal{V}})(xu_i + yv_i) = (x + \sigma_i y)u_i - (y + \sigma_i x)v_i$. Therefore, the operator norm is given by

$$\|P_{\mathcal{U}} - P_{\mathcal{V}}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} = \max_i \max_{x \neq 0 \neq y} \frac{1}{\|xu_i + yv_i\|_{\ell_2(\mathbb{N})}} \|(P_{\mathcal{U}} - P_{\mathcal{V}})(xu_i + yv_i)\|_{\ell_2(\mathbb{N})}.$$

We note the norm equality $\|xu_i + yv_i\|_{\ell_2(\mathbb{N})}^2 = x^2 + y^2 + 2\sigma_i xy$. Then on the one hand,

$$\begin{aligned}
\|(P_{\mathcal{U}} - P_{\mathcal{V}})(xu_i + yv_i)\|_{\ell_2(\mathbb{N})}^2 &= (x + \sigma_i y)^2 + (y + \sigma_i x)^2 - 2\sigma_i(x + \sigma_i y)(y + \sigma_i x) \\
&= x^2 + y^2 + 2\sigma_i xy - \sigma_i^2(x^2 + y^2 + 2\sigma_i xy) \\
&= (1 - \sigma_i^2)\|xu_i + yv_i\|_{\ell_2(\mathbb{N})}^2 \\
&\geq (1 - \sigma_i)\|xu_i + yv_i\|_{\ell_2(\mathbb{N})}^2,
\end{aligned}$$

that is, $\|P_{\mathcal{U}} - P_{\mathcal{V}}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})}^2 \geq 1 - \sigma_i$. On the other hand, we have

$$(U - V)^{\mathsf{T}}(U - V) = 2(\operatorname{id}_r - \Sigma),$$

and thus $\max_{\|w\|=1} \|(U - V)w\|_{\ell_2(\mathbb{N})}^2 = 2\max_i(1 - \sigma_i)$, which leads to the desired inequality.

Ad (ii). Using inequality $(a - b)^2 \leq 2a^2 + 2b^2$ componentwise, we get

$$\begin{aligned}
\|x - y\|^2 &= (x - y)^{\mathsf{T}}\Sigma(x - y) + (x - y)^{\mathsf{T}}(\operatorname{id}_r - \Sigma)(x - y) \\
&\leq (x - y)^{\mathsf{T}}\Sigma(x - y) + 2x^{\mathsf{T}}(\operatorname{id}_r - \Sigma)x + 2y^{\mathsf{T}}(\operatorname{id}_r - \Sigma)y \\
&\leq 2\|x\|^2 + 2\|y\|^2 - 4x^{\mathsf{T}}\Sigma y \\
&= 2\|Ux\|_{\ell_2(\mathbb{N})} + 2\|Vy\|_{\ell_2(\mathbb{N})} - 4(Ux)^{\mathsf{T}}(Vy) = 2\|Ux - Vy\|_{\ell_2(\mathbb{N})}^2,
\end{aligned}$$

which is the claim. □

Proof of Proposition 4.7 Assume representations

$$X = C \times_1 U_1 \times_2 \cdots \times_d U_d, \quad Y = \tilde{C} \times_1 \tilde{U}_1 \times_2 \cdots \times_d \tilde{U}_d$$

as in Proposition 4.2. By using polar decompositions $(U^\mu)^\mathsf{T}\tilde{U}^\mu = Q^\mu S^\mu$, where Q^μ is orthogonal and S^μ is positive semidefinite, we can replace the U^μ with $U^\mu Q^\mu$ and the core tensor C accordingly such that $(U^\mu)^\mathsf{T}\tilde{U}^\mu$ is positive semidefinite, which we assume to be the case for all $\mu = 1, \dots, d$. By Proposition 4.2,

$$P_X - P_Y = P_X^0 - P_Y^0 + \sum_{\mu=1}^{d} P_X^\mu - P_Y^\mu .$$

We will estimate the single differences separately. Applying the triangle inequality will then prove the assertion.

We first consider any of the projector differences $P_X^\mu - P_Y^\mu$ for $\mu = 1, \dots, d$. By (4.8), they can be written in the μ-th matricization space as

$$\begin{aligned} P_X^\mu - P_Y^\mu &= (\mathrm{id} - P_{\mathcal{U}^\mu}) \otimes P_{\mathcal{V}^\mu} - (\mathrm{id} - P_{\tilde{\mathcal{U}}^\mu}) \otimes P_{\tilde{\mathcal{V}}^\mu} \\ &= (\mathrm{id} - P_{\mathcal{U}^\mu}) \otimes (P_{\mathcal{V}^\mu} - P_{\tilde{\mathcal{V}}^\mu}) + (P_{\tilde{\mathcal{U}}^\mu} - P_{\mathcal{U}^\mu}) \otimes P_{\tilde{\mathcal{V}}^\mu} . \end{aligned}$$

We have

$$\|P_{\tilde{\mathcal{U}}^\mu} - P_{\mathcal{U}^\mu}\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})} \le \frac{1}{\sigma_{r_\mu}^\mu} \|M_X^\mu - M_Y^\mu\|_{\ell_2(\mathbb{N}^{d-1})\to\ell_2(\mathbb{N})} \le \frac{1}{\sigma_{r_\mu}^\mu} \|X - Y\|_{\ell_2(\mathbb{N}^d)}, \tag{4.13}$$

where again M_X^μ and M_Y^μ denote the matricizations of X and Y and $\sigma_{r_\mu}^\mu = \sigma_{r_\mu}^\mu(X)$ denotes the smallest positive singular value of M_X^μ as in (4.10). For the first inequality see, for example, the proof of [3, Lemma A.2], the second one is trivial. The same upper bound holds for $\|P_{\tilde{\mathcal{V}}^\mu} - P_{\mathcal{V}^\mu}\|_{\ell_2(\mathbb{N}^{d-1})\to\ell_2(\mathbb{N}^{d-1})}$. Thus we conclude

$$\|P_X^\mu - P_Y^\mu\|_{\ell_2(\mathbb{N}^d)\to\ell_2(\mathbb{N}^d)} \le \frac{2}{\sigma} \|X - Y\|_{\ell_2(\mathbb{N}^d)} .$$

We now proceed with estimating the operator norm of the difference $P_X^0 - P_Y^0$. By (4.9),

$$\begin{aligned} (P_X^0 - P_Y^0)(Z) &= P_C(C_Z) \times_1 U^1 \times_2 \cdots \times_d U^d - P_{\tilde{C}}(\tilde{C}_Z) \times_1 \tilde{U}^1 \times_2 \cdots \times_d \tilde{U}^d \\ &= [P_C(C_Z) - P_{\tilde{C}}(\tilde{C}_Z)] \times_1 \tilde{U}^1 \times_2 \cdots \times_d \tilde{U}^d \\ &\quad + P_C(C_Z) \times_1 [U^1 - \tilde{U}^1] \times_2 \tilde{U}^2 \times_3 \cdots \times_d \tilde{U}^d \\ &\quad \vdots \\ &\quad + P_C(C_Z) \times_1 U^1 \times_2 \cdots \times_{d-1} U^{d-1} \times_d [U^d - \tilde{U}^d] \end{aligned}$$

where $C_Z = Z \times_1 (U^1)^T \times_2 \cdots \times_d (U^d)^T$ and similar for $\tilde{C}_Z$. The first term in the right sum is bounded by

$$\|P_C(C_Z) - P_{\tilde{C}}(\tilde{C}_Z)\|_{\ell_2(\mathbb{N}^d)} \le \|(P_C - P_{\tilde{C}})C_Z\| + \|C_Z - \tilde{C}_Z\|,$$

since U^μ and $\tilde{U}^\mu$ have orthonormal columns and hence spectral norm one. For the other terms we use Lemma 4.8(i) and (4.13), which leads to

$$
\begin{aligned}
&\| P_C(C_Z) \times_1 [U^1 - \tilde{U}^1] \times_2 \tilde{U}^2 \times_3 \cdots \times_d \tilde{U}^d \|_{\ell_2(\mathbb{N}^d)} \\
&\le \sqrt{2} \| P_{\mathcal{U}^1} - P_{\tilde{\mathcal{U}}^1} \|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} \| P_C(C_Z) \| \\
&\le \frac{\sqrt{2}}{\sigma_{r_1^1}(X)} \| X - Y \|_{\ell_2(\mathbb{N}^d)} \| C_Z \| \le \frac{\sqrt{2}}{\sigma_{r_1}^1} \| X - Y \|_{\ell_2(\mathbb{N}^d)} \| Z \|_{\ell_2(\mathbb{N}^d)}
\end{aligned}
$$

and we proceed similarly for the further modes. So far we have shown

$$
\begin{aligned}
\|(P_X^0 - P_Y^0)(Z)\|_{\ell_2(\mathbb{N}^d)} \le \|(P_C - P_{\tilde{C}}) C_Z\| + \|C_Z - \tilde{C}_Z\| \\
+ \sqrt{2} \left(\frac{1}{\sigma_{r_1}^1} + \cdots + \frac{1}{\sigma_{r_d}^d} \right) \|X - Y\|_{\ell_2(\mathbb{N}^d)} \|Z\|_{\ell_2(\mathbb{N}^d)}.
\end{aligned}
$$

It remains to estimate $\|C_Z - \tilde{C}_Z\|$ and $\|(P_C - P_{\tilde{C}}) C_Z\|$. Using again a telescopic expansion of

$$
C_Z - \tilde{C}_Z = Z \times_1 (U^1)^\mathsf{T} \times_2 \cdots \times_d (U^d)^\mathsf{T} - Z \times_1 (\tilde{U}^1)^\mathsf{T} \times_2 \cdots \times_d (\tilde{U}^d)^T,
$$

one obtains in a similar way as above that

$$
\|C_Z - \tilde{C}_Z\| \le \sqrt{2} \left(\frac{1}{\sigma_{r_1}^1} + \cdots + \frac{1}{\sigma_{r_d}^d} \right) \|X - Y\|_{\ell_2(\mathbb{N}^d)} \|Z\|_{\ell_2(\mathbb{N}^d)}.
$$

We need to bound $\|(P_C - P_{\tilde{C}}) C_Z\|$ in terms of $\|X - Y\|_{\ell_2(\mathbb{N}^d)}$. Note that

$$
X - Y = C \times_1 U_1 \times_2 \cdots \times_d U_d - \tilde{C} \times_1 \tilde{U}_1 \times_2 \cdots \times_d \tilde{U}_d
$$

where U_μ and $\tilde{U}_\mu$ satisfy the assumptions in Lemma 4.8. It follows that $\|C - \tilde{C}\| \le \sqrt{2} \|X - Y\|_{\ell_2(\mathbb{N}^d)}$. Hence by (4.12) and Proposition 4.6 we have

$$
\|(P_C - P_{\tilde{C}}) C_Z\| \le \frac{c}{\sigma} \|C - \tilde{C}\| \|C_Z\| \le \frac{\sqrt{2} c}{\sigma} \|X - Y\|_{\ell_2(\mathbb{N}^d)} \|Z\|_{\ell_2(\mathbb{N}^d)}.
$$

In total, we have

$$
\|(P_X^0 - P_Y^0)(Z)\|_{\ell_2(\mathbb{N}^d)} \le \sqrt{2} \left(\frac{c}{\sigma} + \frac{2}{\sigma_{r_1}^1} + \cdots + \frac{2}{\sigma_{r_d}^d} \right) \|X - Y\|_{\ell_2(\mathbb{N}^d)} \|Z\|_{\ell_2(\mathbb{N}^d)}.
$$

In summary, this allows to conclude the asserted curvature estimate. □

In Assumption **A2** we also need an estimate for the projection $\mathrm{id} - P_X$.

Proposition 4.9 *Assume a curvature estimate of the form*

$$\|(\mathrm{id} - P_{\tilde{C}})(C - \tilde{C})\| \leq \frac{c}{\hat{\sigma}} \|C - \tilde{C}\|^2 \quad \text{for all } C \in \mathcal{M}_\mathrm{c} \text{ and } \tilde{C} \in \overline{\mathcal{M}_\mathrm{c}},$$

where $\hat{\sigma} = \mathrm{dist}(C, \overline{\mathcal{M}_\mathrm{c}} \backslash \mathcal{M}_\mathrm{c})$ *and where* $c > 0$ *is independent of* $C, \tilde{C}$. *Let* $X, Y \in \mathcal{M}$ *with corresponding tangent space projections* P_X *and* P_Y. *Then*

$$\|(\mathrm{id} - P_X)(X - Y)\|_{\ell_2(\mathbb{N}^d)} \leq \sqrt{\frac{c^2}{\sigma^2} + \sum_{\mu=1}^{d} \frac{1}{(\sigma^\mu_{r_\mu})^2}} \|X - Y\|^2_{\ell_2(\mathbb{N}^d)} \leq \frac{\sqrt{d + c^2}}{\sigma} \|X - Y\|^2_{\ell_2(\mathbb{N}^d)}$$

where $\sigma = \mathrm{dist}(X, \overline{\mathcal{M}}^w \setminus \mathcal{M})$.

Proof We use the same notation as in the proof of Proposition 4.7. We decompose the identity into

$$\mathrm{id} = P_{\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^d} + P_{(\mathcal{U}^1)^\perp \otimes \cdots \otimes \mathcal{U}^d} + P_{\ell_2(\mathbb{N}) \otimes (\mathcal{U}^2)^\perp \cdots \otimes \mathcal{U}^d} + \ldots + P_{\ell_2(\mathbb{N}^{d-1}) \otimes (\mathcal{U}^d)^\perp}. \tag{4.14}$$

Then

$$\begin{aligned}(\mathrm{id} - P_X)(X - Y) &= (P_{\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^d} - P^0_X)(X - Y) + (P_{(\mathcal{U}^1)^\perp \otimes \cdots \otimes \mathcal{U}^d} - P^1_X)(X - Y) \\ &+ (P_{\ell_2(\mathbb{N}) \otimes (\mathcal{U}^2)^\perp \otimes \cdots \otimes \mathcal{U}^d} - P^2_X)(X - Y) + \ldots + (P_{\ell_2(\mathbb{N}^{d-1}) \otimes (\mathcal{U}^d)^\perp} - P^d_X)(X - Y)\end{aligned}$$

holds. For the first summand, we have

$$\begin{aligned}(P_{\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^d} - P^0_X)(X - Y) &= \Big((\mathrm{id} - P_C)(X - Y) \times_1 (U^1)^\mathsf{T} \times_2 \cdots \times_d (U^d)^\mathsf{T}\Big) \\ &\quad \times_1 U^1 \times_2 \cdots \times_d U^d\end{aligned}$$

where $\|(X - Y) \times_1 (U^1)^\mathsf{T} \times_2 \cdots \times_d (U^d)^\mathsf{T}\| \leq \|X - Y\|_{\ell_2(\mathbb{N}^d)}$. Since $\mathrm{dist}(X, \overline{\mathcal{M}}^w \backslash \mathcal{M}) = \mathrm{dist}(C, \overline{\mathcal{M}_\mathrm{c}} \backslash \mathcal{M}_\mathrm{c})$ by Proposition 4.6, we have

$$\|(P_{\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^d} - P^0_X)(X - Y)\|_{\ell_2(\mathbb{N}^d)} \leq \frac{c}{\sigma} \|X - Y\|^2_{\ell_2(\mathbb{N}^d)},$$

where we use that $Y \times_1 (U^1)^\mathsf{T} \times_2 \cdots \times_d (U^d)^\mathsf{T} \in \overline{\mathcal{M}_\mathrm{c}}$. For the next summand, we obtain

$$(P_{(\mathcal{U}^1)^\perp \otimes \cdots \otimes \mathcal{U}^d} - P^1_X)(X - Y) = P_{(\mathcal{U}^1)^\perp} \otimes (\mathrm{id}_{\ell_2(\mathbb{N}^{d-1})} - P_{\mathcal{V}^1})(X - \mathrm{id}_{\ell_2(\mathbb{N})} \otimes P_{\mathcal{U}^2 \otimes \cdots \otimes \mathcal{U}^d} Y).$$

Then, $\|X - \mathrm{id}_{\ell_2(\mathbb{N})} \otimes P_{\mathcal{U}^2 \otimes \cdots \otimes \mathcal{U}^d} Y\|_{\ell_2(\mathbb{N}^d)} \leq \|X - Y\|_{\ell_2(\mathbb{N}^d)}$ and $\mathrm{id}_{\ell_2(\mathbb{N})} \otimes P_{\mathcal{U}^2 \otimes \cdots \otimes \mathcal{U}^d} Y \in \overline{\mathcal{M}}^w$. If $\mathrm{id}_{\ell_2(\mathbb{N})} \otimes P_{\mathcal{U}^2 \otimes \cdots \otimes \mathcal{U}^d} Y \in \mathcal{M}$, then we have corresponding spaces $\tilde{\mathcal{U}}^1$ and $\tilde{\mathcal{V}}^1$ and

$$\|P_{\mathcal{V}^1} - P_{\tilde{\mathcal{V}}^1}\|_{\ell_2(\mathbb{N}^d) \to \ell_2(\mathbb{N}^d)} \leq \frac{1}{\sigma^1_{r_1}} \|X - \mathrm{id}_{\ell_2(\mathbb{N})} \otimes P_{\mathcal{U}^2 \otimes \cdots \otimes \mathcal{U}^d} Y\|_{\ell_2(\mathbb{N}^d)} \leq \frac{1}{\sigma^1_{r_1}} \|X - Y\|_{\ell_2(\mathbb{N}^d)}.$$

As a consequence,

$$\|(P_{(\mathcal{U}^1)^\perp\otimes\cdots\otimes\mathcal{U}^d} - P_X^1)(X-Y)\|_{\ell_2(\mathbb{N}^d)} \le \frac{1}{\sigma_{r_1}^1}\|X-Y\|_{\ell_2(\mathbb{N}^d)}^2.$$

If $\mathrm{id}_{\ell_2(\mathbb{N})}\otimes P_{\mathcal{U}^2\otimes\cdots\otimes\mathcal{U}^d} Y \notin \mathcal{M}$, the same estimate follows by continuity of the linear operator $P_{(\mathcal{U}^1)^\perp\otimes\cdots\otimes\mathcal{U}^d} - P_X^1$. Similar considerations show

$$\|(P_{\ell_2(\mathbb{N}^{\mu-1})\otimes(\mathcal{U}^\mu)^\perp\otimes\cdots\otimes\mathcal{U}^d} - P_X^\mu)(X-Y)\|_{\ell_2(\mathbb{N}^d)} \le \frac{1}{\sigma_{r_\mu}^\mu}\|X-Y\|_{\ell_2(\mathbb{N}^d)}^2$$

for $\mu = 2, \dots, d$. Finally, we obtain

$$\|(\mathrm{id} - P_X)(X-Y)\|_{\ell_2(\mathbb{N}^d)} \le \sqrt{\frac{c^2}{\sigma^2} + \sum_{\mu=1}^{d}\frac{1}{(\sigma_{r_\mu}^\mu)^2}}\,\|X-Y\|_{\ell_2(\mathbb{N}^d)}^2$$

utilizing that the images of the operators appearing in (4.14) are orthogonal. The final inequality follows with Proposition 4.6. □

4.2 Application to tensor train manifolds

In order to give the above estimates a more concrete meaning, we now consider the popular example of the fixed-rank tensor-train (TT) format discussed in the introduction. Here $\mathcal{M}_c = \mathcal{M}_{\mathbf{k}}$ consists of all finite dimensional tensors $C \in \mathbb{R}^{r_1\times\cdots\times r_d}$ with the fixed TT rank $\mathbf{k}$ of the form (1.3). We denote the resulting manifold $\mathcal{M}$ in (4.1) by $\mathcal{M}_{\mathbf{r},\mathbf{k}}$. It thus contains infinite tensors in $\ell_2(\mathbb{N}^d)$ of "outer" multilinear rank $\mathbf{r} = (r_1, \dots, r_d)$ and "inner" TT rank $\mathbf{k} = (k_1, \dots, k_{d-1})$. This can be seen as a special case of the hierarchical tensor format with linear dimension tree, see [2, Rem. 2.27].

Proposition 4.10 *Let $X, Y \in \mathcal{M}_{\mathbf{r},\mathbf{k}}$ with corresponding tangent space projections P_X and P_Y. Then*

$$\|P_X - P_Y\|_{\ell_2(\mathbb{N}^d)\to\ell_2(\mathbb{N}^d)} \le \frac{2d\left(3\sqrt{2}+1\right)}{\sigma}\|X-Y\|_{\ell_2(\mathbb{N}^d)},$$
$$\|(\mathrm{id} - P_X)(X-Y)\|_{\ell_2(\mathbb{N}^d)} \le \frac{\sqrt{2d-1}}{\sigma}\|X-Y\|_{\ell_2(\mathbb{N}^d)}^2,$$

where $\sigma = \mathrm{dist}(X, \overline{\mathcal{M}_{\mathbf{r},\mathbf{k}}}^w \setminus \mathcal{M}_{\mathbf{r},\mathbf{k}})$.

The result follows directly follows from Propositions 4.7 and 4.9 and the following refined curvature estimates for the finite-dimensional TT manifold $\mathcal{M}_{\mathbf{k}}$, which under the given assumptions seem to be new.

Proposition 4.11 *Let $\mathcal{M}_{\mathbf{k}} \subset \mathbb{R}^{N_1\times\cdots\times N_d}$ be a finite-dimensional TT manifold of fixed TT rank* $\mathbf{k}$. *Let $X, Y \in \mathcal{M}_{\mathbf{k}}$ and $\sigma = \operatorname{dist}(X, \overline{\mathcal{M}_{\mathbf{k}}}\setminus\mathcal{M}_{\mathbf{k}})$. Then*

$$\max_{\|Z\|=1} \|(P_X - P_Y)Z\| \le \frac{4d}{\sigma}\|X-Y\|, \quad \|(\mathrm{id} - P_X)(X-Y)\| \le \frac{\sqrt{d-1}}{\sigma}\|X-Y\|^2$$

and

$$\|(\mathrm{id} - P_X)(X-Y)\| \le \frac{\sqrt{d-1}}{\sigma}\|X-Y\|^2.$$

Furthermore, the last inequality holds more generally for $Y \in \overline{\mathcal{M}_{\mathbf{k}}}$.

The proof is given in the appendix.

5 Application to the model problem

We now return to the model problem (1.6) under the regularity assumption **A0**.

Problem 5.1 Given $f \in L_2(0,T; L_2(\Omega))$ and $u_0 \in \mathcal{M} \cap H_0^1(\Omega)$, find

$$u \in W(0,T; H_0^1(\Omega), L_2(\Omega)) = \{u \in L_2(0,T; H_0^1(\Omega)) \colon u' \in L_2(0,T;\mathcal{V}^*)\}$$

such that for almost all $t \in [0,T]$,

$$\begin{aligned} u(t) &\in \mathcal{M}, \\ \langle u'(t), v\rangle + a(u(t), v; t) &= \langle f(t), v\rangle \quad \text{for all } v \in T_{u(t)}\mathcal{M} \cap H_0^1(\Omega), \\ u(0) &= u_0. \end{aligned} \tag{5.1}$$

Here

$$a(u,v;t) = \int_\Omega (B(t)\nabla u(x)) \cdot \nabla v(x)\, dx$$

with a symmetric positive definite matrix $B(t)$ that is entrywise Lipschitz continuous in t and $\mathcal{M}$ is a manifold of functions in $L_2(\Omega) = L_2(\Omega_1 \times \cdots \times \Omega_d)$ as described in (4.1), that is, $u \in \mathcal{M}$ is of the form

$$u = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} C(k_1,\ldots,k_d)\, u_{k_1}^1 \otimes \cdots \otimes u_{k_d}^d, \tag{5.2}$$

with $C \in \mathcal{M}_{\mathrm{c}}$.

5.1 Discussion of main assumptions

Our goal is to apply Theorems 2.2, 2.5 and 3.2 to Problem 5.1. It suffices to verify the Assumptions **A1**–**A4**, **B1**, and **B2**. Assumption **A1** holds since $\mathcal{M}_\mathrm{c}$ is a cone, which follows from the invariance assumption (4.2). We already proved **A2** in Proposition 4.7 and Proposition 4.9 assuming that corresponding curvature bounds for $\mathcal{M}_\mathrm{c}$ are available. Indeed, in the special case where $\mathcal{M}_\mathrm{c}$ is the manifold of tensors with constant TT-rank $\mathcal{M}_\mathbf{k}$, such curvature bounds are stated in Proposition 4.10. We now consider the remaining Assumptions **A3** and **A4** as well as **B1** and **B2**. We make use of the following well-known technique for estimating norms of factors in low-rank representations, see for example [32].

Lemma 5.2 *Let $u \in H_0^1(\Omega)$ admit a singular value decomposition*

$$u(x) = \sum_{k=1}^{r_\mu} \sigma_k u_{1,k}(x_\mu) u_{2,k}(x_{\{1,\dots,d\}\setminus\{\mu\}})$$

with respect to the μ-th variable. Then the singular vectors satisfy $u_{1,k} \in H_0^1(\Omega_\mu)$ and $u_{2,k} \in H_0^1(\times_{\nu\neq\mu} \Omega_\nu)$ with

$$\|u_{1,k}\|_{H_0^1(\Omega_\mu)} \le \frac{1}{\sigma_k}\|u\|_{H_0^1(\Omega)} \quad \textit{and} \quad \|u_{2,k}\|_{H_0^1(\times_{\nu\neq\mu} \Omega_\nu)} \le \frac{1}{\sigma_k}\|u\|_{H_0^1(\Omega)}.$$

Proof We state the proof for $\mu = 1$. Then

$$\sigma_k u_{1,k}(x_1) = \int_{\times_{\nu=2}^d \Omega_\nu} u(x_1, x_2, \dots, x_d) u_{2,k}(x_2, \dots, x_d)\, d(x_2, \dots, x_d)$$

and

$$\sigma_k u_{2,k}(x_2, \dots, x_d) = \int_{\Omega_1} u(x_1, x_2, \dots, x_d) u_{1,k}(x_1)\, dx_1.$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}\sigma_k^2 \|u_{1,k}\|^2_{H_0^1(\Omega_1)} &= \sigma_k^2 \int_{\Omega_1} \left|\nabla_{x_1} u_{1,k}\right|^2 dx_1 = \int_{\Omega_1} \left| \int_{\times_{\nu=2}^d \Omega_\nu} \nabla_{x_1} u\; u_{2,k}\, d(x_2, \dots, x_d)\right|^2 dx_1 \\ &\le \|u_{2,k}\|^2_{L_2((0,1)^{d-1})} \int_\Omega \left|\nabla_{x_1} u\right|^2 d(x_1, \dots, x_d) = \|u\|^2_{H_0^1(\Omega_1)\otimes L^2(\times_{\nu=2}^d \Omega_\nu)} \le \|u\|^2_{H_0^1(\Omega)}.\end{aligned}$$

This proves the first estimate. The other one follows in analogy. □

5.1.1 Assumption A3: Compatibility of tangent spaces

We now verify **A3**(a).

Lemma 5.3 *Let $u \in H^1_0(\Omega) \cap \mathcal{M}$ and $v \in H^1_0(\Omega) \cap T_u\mathcal{M}$. Then there exists an admissible curve $\varphi(t) \in H^1_0(\Omega) \cap \mathcal{M}$ for $|t|$ small enough with $\varphi(0) = u$ and $\varphi'(0) = v$.*

Proof Let u be of the form (5.2). Then by Lemma 5.2 all basis functions $u^\mu_{k_\mu} \in H^1_0(\Omega_\mu)$. We write v is in (4.5), that is, $v = v_0 + v_1 + \ldots + v_d$, where

$$v_0 = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} \dot{C}(k_1, \ldots, k_d)\, u^1_{k_1} \otimes \cdots \otimes u^d_{k_d}$$

with $\dot{C} \in T_C\mathcal{M}_{\rm c}$ and

$$v_\mu = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} C(k_1, \ldots, k_d)\, u^1_{k_1} \otimes \cdots \otimes u^{\mu-1}_{k_{\mu-1}} \otimes \dot{u}^\mu_{k_\mu} \otimes u^{\mu+1}_{k_{\mu+1}} \otimes \cdots \otimes u^d_{k_d}$$

where $\dot{u}^\mu_{k_\mu}$ is orthogonal to all $u^\mu_1, \ldots, u^\mu_{r_\mu}$. By similar reasoning as in Lemma 5.2 one can show that $\dot{u}^\mu_{k_\mu} \in H^1_0(\Omega_\mu)$.

Furthermore, there exists a curve $D(t) \in \mathcal{M}_{\rm c}$ for $|t|$ small enough such that $D(0) = C$ and $D'(0) = \dot{C}$. We now choose

$$\varphi(t) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} D(k_1, \ldots, k_d)(t)\, (u^1_{k_1} + t\dot{u}^1_{k_1}) \otimes \cdots \otimes (u^d_{k_d} + t\dot{u}^d_{k_d}),$$

which is a differentiable curve in $\mathcal{M} \cap H^1_0(\Omega)$ for small enough $|t|$. By the product rule, it satisfies $\varphi(0) = u$ and $\varphi'(0) = v$. $\square$

Assumption **A3**(b), which states that for $u \in H^1_0(\Omega) \cap \mathcal{M}$ the L_2-orthogonal projection onto its tangent space is also a bounded operator with respect to the H^1_0-norm, follows with a similar technique as Lemma 5.2.

Proposition 5.4 *Let $\mathcal{M}$ be of the form (4.1), let $u \in H^1_0(\Omega) \cap \mathcal{M}$ and $v \in H^1_0(\Omega)$. Let $P^0_u, \ldots, P^d_u$ be the projections in (4.7). Then*

$$\|P^\mu_u v\|_{H^1_0(\Omega)} \leq c\|v\|_{H^1_0(\Omega)}$$

Proof First, we consider P^0_u. We note the norm bound

$$\|C_Z\| \leq \|Z\|_{\ell_2(\mathbb{N}^d)}$$

in the definition of P^0_X in Proposition 4.2. In terms of the represented function,

$$\|C_v\| \leq \|v\|_{L_2(\Omega)}.$$

Furthermore, we have

$$\|P_C(C_v)\| \leq \|C_v\|$$

since P_C is an ℓ_2-orthogonal projection. We now consider the summands of

$$\|P_u^0 v\|^2_{H_0^1(\Omega)} = \|P_u^0 v\|^2_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} + \cdots + \|P_u^0 v\|^2_{L_2(\times_{\mu=1}^{d-1} \Omega_\mu)\otimes H_0^1(\Omega_d)}$$

independently. Let

$$u(x) = \sum_{k=1}^{r_1} \sum_{\ell=1}^{r_1} a_{k\ell} u_{1,k}(x_1) u_{2,\ell}(x_2, \ldots, x_d) \tag{5.3}$$

be a decomposition of u separating the first variable, where $\{u_{1,k}\}$ and $\{u_{2,\ell}\}$ are L_2-orthonormal (as, for example, in a singular value decomposition). After an orthogonal change of basis, we may assume the vectors $u_{1,k}$ to be both L_2-orthonormal and H_0^1-orthogonal. Note that the basis vectors are given by

$$u_{1,k} = \int_{\times_{\mu=2}^d \Omega_\mu} u \sum_{\ell=1}^{r_1} b_{k\ell} u_{2,\ell} \, d(x_2, \ldots, x_d) \quad \text{and} \quad u_{2,\ell} = \int_{\Omega_1} u \sum_{k=1}^{r_1} b_{k\ell} u_{1,k} \, dx_1$$

where $\sum_{\ell=1}^{r_1} a_{k_1\ell} b_{k_2\ell} = \delta_{k_1,k_2}$.

The singular values of the corresponding decomposition satisfy $\sigma_k \geq \sigma$ for $k = 1, \ldots, r_\nu$, and hence the vectors in (5.3) satisfy

$$\begin{aligned} \sigma \|u_{1,k}\|_{H_0^1(\Omega_1)} &\leq \|u\|_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)}, \\ \sigma \|u_{2,k}\|_{H_0^1(\times_{\mu=2}^d \Omega_\mu)} &\leq \|u\|_{L_2(\Omega_1)\otimes H_0^1(\times_{\mu=2}^d \Omega_\mu)}, \end{aligned} \tag{5.4}$$

since $\sigma \|\sum_{k=1}^{r_1} b_{k\ell} u_{1,k}\|_{L_2(\times_{\mu=2}^d \Omega_\mu)} \leq 1$, $\sigma \|\sum_{\ell=1}^{r_1} b_{k\ell} u_{2,\ell}\|_{L_2(\Omega_1)} \leq 1$ and by the last argument of the proof of Lemma 5.2. By L_2- and H_0^1-orthogonality and (5.4), we obtain the estimate

$$\begin{aligned} &\|P_u^0 v\|^2_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} \\ &\quad = \left\| \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} P_C(C_v)(k_1, \ldots, k_d)\, \partial_{x_1} u_{k_1}^1 \otimes u_{k_2}^2 \otimes \cdots \otimes u_{k_d}^d \right\|^2_{L_2(\Omega)} \\ &\quad = \sum_{k_1=1}^{r_1} \left\| \sum_{k_2=1}^{r_2} \cdots \sum_{k_d=1}^{r_d} P_C(C_v)(k_1, \ldots, k_d)\, u_{k_2}^2 \otimes \cdots \otimes u_{k_d}^d \right\|^2_{L_2(\times_{\mu=2}^d \Omega_\mu)} \|u_{k_1}^1\|^2_{H_0^1(\Omega_1)} \\ &\quad \leq \frac{1}{\sigma^2} \|P_C(C_v)\|^2 \|u\|^2_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)}. \end{aligned}$$

Using the Poincaré inequality, we have

$$\|P_C(C_v)\| \le \|C_v\| = \|v\|_{L_2(\Omega)} \le c_\Omega \|v\|_{H_0^1(\Omega)}$$

with a $c_\Omega > 0$ depending only on Ω. We thus arrive at

$$\|P_u^0 v\|_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} \le \frac{c_\Omega}{\sigma} \|v\|_{H_0^1(\Omega)} \|u\|_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)}.$$

Similarly,

$$\begin{aligned} &\|P_u^0 v\|_{L_2(\times_{\nu=1}^{\mu-1} \Omega_\nu)\otimes H_0^1(\Omega_\mu)\otimes L_2(\times_{\nu=\mu+1}^d \Omega_\nu)} \\ &\quad \le \frac{c_\Omega}{\sigma} \|v\|_{H_0^1(\Omega)} \|u\|_{L_2(\times_{\nu=1}^{\mu-1} \Omega_\nu)\otimes H_0^1(\Omega_\mu)\otimes L_2(\times_{\nu=\mu+1}^d \Omega_\nu)}. \end{aligned}$$

This yields

$$\|P_u^0 v\|_{H_0^1(\Omega)} \le \frac{c_\Omega}{\sigma} \|v\|_{H_0^1(\Omega)} \|u\|_{H_0^1(\Omega)}.$$

Next we consider P_u^1; the estimates for the projections $P_u^2, \ldots, P_u^d$ follow in analogy. The action of P_u^1 is given by the tensor product of L_2-orthogonal projections

$$P_u^1(v) = (\mathrm{id} - P_1) \otimes P_2 v,$$

where

$$(P_1 w)(x_1) = \sum_{k=1}^{r_1} u_{1,k}(x_1) \int_{\Omega_1} u_{1,k}(y_1) w(y_1)\, dy_1$$

and

$$\begin{aligned} (P_2 w)(x_2, \ldots, x_d) \; &= \sum_{k=1}^{r_1} u_{2,k}(x_2, \ldots, x_d) \int_{\times_{\mu=2}^d \Omega_\mu} u_{2,k} \\ &\qquad (y_2, \ldots, y_d) w(y_2, \ldots, y_d)\, d(y_2, \ldots, y_d). \end{aligned}$$

We again use the decomposition (5.3) with L_2-orthonormal and H_0^1-orthogonal sets of vectors $\{u_{1,k} : k = 1, \ldots, r_1\}$ and $\{u_{2,k} : k = 1, \ldots, r_1\}$. Using orthogonality, we get the estimate

$$\begin{aligned} \|P_1 w\|_{H_0^1(\Omega_1)}^2 &= \sum_{k=1}^{r_1} \|u_{1,k}\|_{H_0^1(\Omega_1)}^2 \left(\int_0^1 u_{1,k}(y_1) w(y_1)\, dy_1 \right)^2 \\ &\le \frac{1}{\sigma^2} \|u\|_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)}^2 \|w\|_{L_2(\Omega_1)}^2 \end{aligned}$$

and similarly

$$\|P_2 w\|_{H_0^1(\times_{\mu=2}^d \Omega_\mu)} \leq \frac{1}{\sigma} \|u\|_{L_2(\Omega_1)\otimes H_0^1(\times_{\mu=2}^d \Omega_\mu)} \|w\|_{L_2(\times_{\mu=2}^d \Omega_\mu)}.$$

We finally obtain

$$\begin{aligned}
\|P_u^1 v\|_{H_0^1(\Omega)}^2 &= \|P_u^1 v\|^2_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} + \|P_u^1 v\|^2_{L_2(\Omega_1)\otimes H_0^1(\times_{\mu=2}^d \Omega_\mu)} \\
&= \|((\mathrm{id} - P_1) \otimes P_2) v\|^2_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} \\
&\quad + \|((\mathrm{id} - P_1) \otimes P_2) v\|^2_{L_2(\Omega_1)\otimes H_0^1(\times_{\mu=2}^d \Omega_\mu)} \\
&\leq \left(\frac{1}{\sigma} \|u\|_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} \|v\|_{L_2(\Omega)} + \|v\|_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)} \right)^2 \\
&\quad + \frac{1}{\sigma^2} \|u\|^2_{L_2(\Omega_1)\otimes H_0^1(\times_{\mu=2}^d \Omega_\mu)} \|v\|^2_{L_2(\Omega)} \\
&\leq 2 \left(1 + \frac{c_{\Omega_1}}{\sigma^2} \|u\|^2_{H_0^1(\Omega)} \right) \|v\|^2_{H_0^1(\Omega_1)\otimes L_2(\times_{\mu=2}^d \Omega_\mu)},
\end{aligned}$$

where we have again used Poincaré's inequality. □

We have thus verified Assumption **A3**.

5.1.2 Assumption A4: Operator decomposition

The bilinear form $a(\cdot, \cdot; t)$ can be written as $a(\cdot, \cdot; t) = a_1(\cdot, \cdot; t) + a_2(\cdot, \cdot; t)$ with

$$a_1(u, v; t) = \int_\Omega \sum_{\mu=1}^{d} \big(b_{\mu\mu}(t) \nabla_{x_\mu} u(x)\big) \cdot \nabla_{x_\mu} v(x)\, dx$$

and

$$a_2(u, v; t) = \int_\Omega \sum_{\substack{\mu,\nu=1 \\ \mu\neq\nu}}^{d} \big(b_{\mu\nu}(t) \nabla_{x_\mu} u(x)\big) \cdot \nabla_{x_\nu} v(x)\, dx,$$

where $b_{\mu\nu}(t)$ are the corresponding blocks of the matrix $B(t)$. These bilinear forms in turn define linear operators A_1 and A_2. For **A4**(a), we can use a similar technique as in [3], albeit on the more complicated manifold $\mathcal{M}$. First, we show that the strong version of A_1, when defined, maps to the tangent space.

Lemma 5.5 *Let $\mathcal{M}$ be of the form (4.1) and $u \in H^2(\Omega) \cap H_0^1(\Omega) \cap \mathcal{M}$. Then for matrices a_μ of the corresponding size, we have $\sum_{\mu=1}^d \nabla_{x_\mu} \cdot (a_\mu \nabla_{x_\mu} u) \in T_u\mathcal{M}$.*

Proof Since $u \in \mathcal{M}$, we can write u as in (5.2) and

$$\nabla_{x_1} \cdot (a_1 \nabla_{x_1} u) = \sum_{k_1=1}^{r_1} \cdots \sum_{k_d=1}^{r_d} C(k_1, \ldots, k_d) \nabla_{x_1} \cdot (a_1 \nabla_{x_1} u^1_{k_1}) \otimes u^2_{k_2} \otimes \cdots \otimes u^d_{k_d}.$$

We define $\varphi(t) = u + t\nabla_{x_1} \cdot (a_1 \nabla_{x_1} u)$. For sufficiently small $|t|$, the Gramian of the system $\{u^1_1 + t\nabla_{x_1} \cdot (a_1 \nabla_{x_1} u^1_1), u^1_2 + t\nabla_{x_1} \cdot (a_1 \nabla_{x_1} u^1_2), \ldots, u^1_{r_1} + t\nabla_{x_1} \cdot (a_1 \nabla_{x_1} u^1_{r_1})\}$ is invertible and hence $\varphi(t) \in \mathcal{M}$ with $\varphi'(t) = \nabla_{x_1} \cdot (a_1 \nabla_{x_1} u)$. Thus $\nabla_{x_1} \cdot (a_1 \nabla_{x_1} u) \in T_u\mathcal{M}$. Analogously, $\nabla_{x_\mu} \cdot (a_\mu \nabla_{x_\mu} u) \in T_u\mathcal{M}$ for $\mu = 1, \ldots, d$ and by linearity, $\sum_{\mu=1}^{d} \nabla_{x_\mu} \cdot (a_\mu \nabla_{x_\mu} u) \in T_u\mathcal{M}$. □

The assumption **A4**(a) now follows by a density argument. For a proof, choose a sequence $(u_n) \subset \mathcal{M} \cap H^2(\Omega) \cap H^1_0(\Omega)$ converging to u in $H^1_0(\Omega)$-norm. Then for $v \in H^1_0(\Omega)$, we have

$$a_1(u_n, v; t) = \langle A_1(t)u_n, v\rangle = \langle A_1(t)u_n, P_{u_n} v\rangle = a_1(u_n, P_{u_n} v; t)$$

since $A_1(t)u_n \in T_{u_n}\mathcal{M}$ by Lemma 5.5. Moreover,

$$a_1(u_n, P_{u_n} v; t) = a_1(u, P_u v) + a_1(u, (P_{u_n} - P_u)v) + a_1(u_n - u, P_{u_n} v).$$

We have $P_{u_n} v \to P_u v$ strongly in $L_2(\Omega)$ by Proposition 4.7, and Proposition 5.4 yields $\limsup_n \|P_{u_n} v\|_{H^1_0} < \infty$. Since $L_2(\Omega)$ is dense in $H^{-1}(\Omega)$ it follows that $P_{u_n} v \to P_u v$ weakly in $H^1_0(\Omega)$ by a standard argument; see for example [37, Prop. 21.23(g)]. Consequently, $a_1(u_n, P_{u_n} v; t) \to a_1(u, P_u v; t)$. At the same time, $a_1(u_n, v; t) \to a_1(u, v; t)$, so we have verified Assumption **A4**(a).

For Assumption **A4**(b), it suffices to verify mixed smoothness for $u \in H^1_0(\Omega) \cap \mathcal{M}$.

Lemma 5.6 *Let $\mathcal{M}$ be of the form (4.1), $u \in H^1_0(\Omega) \cap \mathcal{M}$, and $\sigma = \operatorname{dist}_{L_2}(u, \overline{\mathcal{M}}^w \setminus \mathcal{M})$. Then for $\nu \neq \mu$, we have $\|\nabla_{x_\mu} \nabla_{x_\nu} u\|_{L_2(\Omega)} \leq \frac{1}{2\sigma} \|u\|^2_{H^1_0(\Omega)}$.*

Proof Let $u(x) = \sum_{k=1}^{r_\mu} \sigma_k u_{1,k}(x_\mu) u_{2,k}(x_{\{1,\ldots,d\}\setminus\{\mu\}})$ be a singular value decomposition of u separating the μ-th variable. Then $\sigma_k \geq \sigma$ for $k = 1, \ldots, r_\nu$. We will use the abbreviation $\Omega_{\mu^c} = \times_{\lambda \neq \mu} \Omega_\lambda$. On the one hand, by the triangle inequality, Young's inequality, and $\sigma \leq \sigma_k$, we have

$$\begin{aligned}
\|\nabla_{x_\mu} \nabla_{x_\nu} u\|_{L_2(\Omega)} &\leq \sum_{k=1}^{r_\mu} \sigma_k \|\nabla_{x_\mu} \nabla_{x_\nu} u_{1,k} \otimes u_{2,k}\|_{L_2(\Omega)} \\
&= \sum_{k=1}^{r_\mu} \sigma_k \|\nabla_{x_\mu} u_{1,k}\|_{L_2(\Omega_\mu)} \|\nabla_{x_\nu} u_{2,k}\|_{L_2(\Omega_{\mu^c})} \\
&\leq \sum_{k=1}^{r_\mu} \frac{\sigma_k^2}{2\sigma} \left(\|\nabla_{x_\mu} u_{1,k}\|^2_{L_2(\Omega_\mu)} + \|\nabla_{x_\nu} u_{2,k}\|^2_{L_2(\Omega_{\mu^c})} \right).
\end{aligned}$$

On the other hand, by L_2-orthogonality of the singular vectors, we have

$$\|u\|^2_{H^1_0(\Omega)} = \int_\Omega \sum_{\lambda=1}^{d} |\nabla_{x_\lambda} u(x)|^2 \, dx = \sum_{k=1}^{r_\mu} \sigma_k^2 \left(\|\nabla_{x_\mu} u_{1,k}\|^2_{L_2(\Omega_\mu)} + \sum_{\lambda \neq \mu} \|\nabla_{x_\lambda} u_{2,k}\|^2_{L_2(\Omega_{\mu^c})} \right),$$

and hence $\|\nabla_{x_\mu} \nabla_{x_\nu} u\|_{L_2(\Omega)} \le \frac{1}{2\sigma} \|u\|^2_{H^1_0(\Omega)}$ as asserted. □

Assumption **A4**(b) now follows directly by integration by parts.

5.1.3 Assumptions B1 and B2: Spatial discretizations

We now exhibit space discretizations that can be used to achieve convergence to the infinite dimensional solution as in Theorem 3.2. It turns out to be sufficient for the discretization to allow the format (4.1). This is a natural requirement, since otherwise one does not have the required product structure for using the low-rank approximation in practice.

For short, let us denote $H^1_0(\Omega) = \mathcal{V} = V^1 \otimes H^2 \otimes \cdots \otimes H^d \cap \cdots \cap H^1 \otimes \cdots \otimes H^{d-1} \otimes V^d$, where $V^\mu = H^1_0(\Omega_\mu)$ and $H^\mu = L_2(\Omega_\mu)$ for $\mu = 1, \ldots, d$. Then $V^1 \otimes \cdots \otimes V^d$ is a continuously and densely embedded subspace of $\mathcal{V}$. As the finite-dimensional subspaces, we choose

$$\mathcal{V}_h = V^1_h \otimes \cdots \otimes V^d_h \quad \text{satisfying} \quad \|P_{V^\mu_h} v^\mu - v^\mu\|_{V^\mu} \to 0 \quad \text{as} \quad h \to 0 \quad \text{for all } v^\mu \in V^\mu. \tag{5.5}$$

This can be a finite element space, but also any other discretization suitable for Ω_μ. As a consequence, for $v \in V^1 \otimes \cdots \otimes V^d$, we have

$$\|v - P_{V^1_h \otimes \cdots \otimes V^d_h} v\|_{\mathcal{V}} \le C \|v - P_{V^1_h \otimes \cdots \otimes V^d_h} v\|_{V^1 \otimes \cdots \otimes V^d} \to 0 \quad \text{for} \quad h \to 0.$$

Since $V^1 \otimes \cdots \otimes V^d$ is a dense subspace of $\mathcal{V}$, the $\mathcal{V}$-orthogonal projection onto $V^1_h \otimes \cdots \otimes V^d_h$ satisfies **B1**(a). For assumption **B1**(b), let $u \in \mathcal{V} \cap \mathcal{M}$. Then $u \in V^1 \otimes \cdots \otimes V^d$ by Lemma 5.2. By (4.2), we have that $P_{V^1_h \otimes \cdots \otimes V^d_h} u \in \overline{\mathcal{M}}^w$. Hence there exists $u_h \in V^1_h \otimes \cdots \otimes V^d_h \cap \mathcal{M}$ such that $\|u_h - P_{V^1_h \otimes \cdots \otimes V^d_h} u\|_{\mathcal{V}} \le \epsilon$ for any $\epsilon > 0$. Thus $\|u_h - u\|_{\mathcal{V}} \to 0$ as $h \to 0$. Possibly after rescaling, we can thus construct a sequence (u_h) that converges to u in $\mathcal{V}$ as $h \searrow 0$ with $\|u_h\|_{\mathcal{V}} \le \|u\|_{\mathcal{V}}$.

Assumption **B2** follows immediately by noting that $\mathcal{M} \cap \mathcal{V}_h$ is of the same form as $\mathcal{M}$ and Theorem 4.1 can be applied.

Remark 5.7 In practical numerical realizations, the elements of the discretization subspaces $\mathcal{V}_h$ in (5.5) need to be represented in terms of suitable basis functions, usually obtained as tensor products of bases of each V^ν_h. While the results given here refer to the represented functions, the properties of the problem in terms of basis coefficients depend also on the condition number of the chosen basis. For tensor product bases, this condition number is the product of the condition number of the univariate bases, and thus in general depends exponentially on d unless orthonormal bases are used for each V^ν_h. A possible remedy for large d is the use of nonstandard basis functions (such as wavelets) for finite element spaces. We refer to [2] for further details.

5.2 Main results

The main results for the model problem are the following specific versions of Theorems 2.2, 3.2, and 2.5. They follow directly by applying the results of Sects. 4 and 5.1.

Theorem 5.8 (Existence and uniqueness of solutions) *Let u_0 have positive L_2-distance from $\overline{\mathcal{M}}^w \setminus \mathcal{M}$. There exist $T^* \in (0, T]$ and $u \in W(0, T^*; H_0^1(\Omega), L_2(\Omega)) \cap L_\infty(0, T^*; H_0^1(\Omega))$ such that u solves Problem 5.1 on the time interval $[0, T^*]$, and its continuous representative $u \in C(0, T^*; L_2(\Omega))$ satisfies $u(t) \in \mathcal{M}$ for all $t \in [0, T^*)$. Here T^* is maximal for the evolution on $\mathcal{M}$ in the sense that if $T^* < T$, then*

$$\liminf_{t \to T^*} \inf_{v \in \overline{\mathcal{M}}^w \setminus \mathcal{M}} \|u(t) - v\|_{L_2(\Omega)} = 0.$$

In either case, u is the unique solution of Problem 5.1 in $W(0, T^*; H_0^1(\Omega), L_2(\Omega))$.

In particular, with $\sigma = \operatorname{dist}_{L_2(\Omega)}(u_0, \overline{\mathcal{M}}^w \setminus \mathcal{M})$, there exists a constant $c > 0$ such that $T^* \geq \min(\sigma^2/c, T)$.

The solution satisfies the following estimates:

$$\|u\|^2_{L_2(0,T^*;H_0^1(\Omega))} \leq \|u_0\|^2_{L_2(\Omega)} + C_1 \|f\|^2_{L_2(0,T^*;L_2(\Omega))},$$
$$\|u'\|^2_{L_2(0,T^*;L_2(\Omega))} \leq C_2 \left(\|u_0\|^2_{H_0^1(\Omega)} + \|f\|^2_{L_2(0,T^*;L_2(\Omega))} \right),$$
$$\|u\|^2_{L^\infty(0,T^*;H_0^1(\Omega))} \leq C_3 \left(\|u_0\|^2_{H_0^1(\Omega)} + \|f\|^2_{L_2(0,T^*;L_2(\Omega))} \right),$$

where C_1, C_2, and C_3 are the constants from [3, Lemma 4.4].

Theorem 5.9 (Convergence of spatial discretizations) *Let $\mathcal{V}_h$ be of the form (5.5). Let $u_{0,h} \in \mathcal{M} \cap \mathcal{V}_h$ define a sequence that converges to u_0 in $H_0^1(\Omega)$ as $h \searrow 0$ and let u_0 have positive $L_2(\Omega)$-distance σ to the relative boundary $\overline{\mathcal{M}}^{\mathrm{w}} \setminus \mathcal{M}$. Then there exists a constant $c > 0$ independent of σ and a constant $h_0 > 0$ such that there is a unique u_h in $W(0, T^*; H_0^1(\Omega), L_2(\Omega)) \cap L_\eta(0, T^*; H_0^1(\Omega))$ that solves Problem 3.1 on the time interval $[0, T^*]$ when $T^* < \sigma^2/c$ for all $h \leq h_0$. Furthermore, u_h converges to the unique solution u of Problem 5.1 in $W(0, T^*; H_0^1(\Omega), L_2(\Omega)) \cap L_\eta(0, T^*; H_0^1(\Omega))$ weakly in $L_2(0, T^*; H_0^1(\Omega))$ and strongly in $C(0, T^*; L_2(\Omega))$, while the weak derivatives u_h' converge weakly to u' in $L_2(0, T^*, L_2(\Omega))$.*

Theorem 5.10 (Stability) *Let $u, v \in W(0, T^*; H_0^1(\Omega), L_2(\Omega))$ be two solutions of Problem 5.1 on a time interval $[0, T^*]$ corresponding to right-hand sides $f, g \in L_2(0, T; L_2(\Omega))$ and initial values $u_0, v_0 \in \mathcal{M}$, respectively. Assume that the continuous representatives $u, v \in C(0, T^*; L_2(\Omega))$ have pointwise positive $L_2(\Omega)$-distance to $\overline{\mathcal{M}}^w \setminus \mathcal{M}$ of at least σ. Then for any $\varepsilon > 0$,*

$$\|u(t) - v(t)\|^2_{L_2(\Omega)} \leq \left(\|u_0 - v_0\|^2_{L_2(\Omega)} + \frac{1}{\varepsilon} \int_0^t \|f(s) - g(s)\|^2_{L_2(\Omega)}\, ds \right) \exp(\Lambda(t) + \varepsilon t),$$

where

$$\Lambda(t) := 2\kappa \int_0^t \|u'(s)\|_{L_2(\Omega)} + \|v'(s)\|_{L_2(\Omega)} + \gamma \left(\|u(s)\|^{\eta}_{H^1_0(\Omega)} + \|v(s)\|^{\eta}_{H^1_0(\Omega)} \right) \\ + \|f(s)\|_{L_2(\Omega)} + \|g(s)\|_{L_2(\Omega)} \, ds < \infty$$

with $\kappa = \kappa(\sigma) = \frac{\sqrt{d+c^2}}{\sigma}$ *from Proposition 4.9.*

Appendix A. Proofs of Theorem 4.1 and Proposition 4.11

Proof of Theorem 4.1 Ad (i). We fix a particular $X_* = C_* \times_1 U^1_* \times_2 \cdots \times_d U^d_*$. The construction of the submersion follows [30], where this has been done for manifolds of fixed multilinear rank in finite-dimensional tensor spaces. Here we additionally have to take the constraint $C \in \mathcal{M}_{\mathrm{c}}$ for the coefficient tensor into account. In the following, $\mathcal{O}$ is an open neighborhood of X_* in $\ell_2(\mathbb{N}^d)$ that can always be chosen sufficiently small to ensure that all maps are well defined.

Since $X_* \in \mathcal{M}$, the matricizations $M^{\mu}_{X_*}$ admit low-rank decompositions (4.4), which can be written in matrix product form as

$$M^{\mu}_{X_*} = U^{\mu}_* (V^{\mu}_*)^{\mathsf{T}}.$$

Here the columns of U^{μ}_* and V^{μ}_* are bases of the minimal subspaces $\mathcal{U}^{\mu}_* \subset \ell_2(\mathbb{N})$ and $\mathcal{V}^{\mu}_* \subset \ell_2(\mathbb{N}^{d-1})$, respectively. For $X \in \mathcal{M}$ sufficiently close to X_*, it will be useful to define a particular basis U^{μ}_X for the μ-th minimal subspace $\mathcal{U}^{\mu}$, $\mu = 1, \ldots, d$, as a continuous function of X. To this end, we choose

$$U^{\mu}_X = M^{\mu}_X (V^{\mu}_*)^{\mathsf{T}}_+, \tag{A.1}$$

where $Y_+ = [Y^{\mathsf{T}} Y]^{-1} Y^{\mathsf{T}}$ denotes the pseudoinverse of a matrix with full column rank. Then U^{μ}_X has full column rank for X close enough to X_*, which follows from $U^{\mu}_X \to U^{\mu}_*$ for all μ for $X \to X_*$ and the lower semicontinuity of the rank. As a result, every X in the neighborhood of X_* can be written as

$$X = C_X \times_1 U^1_X \times_2 \cdots \times_d U^d_X.$$

Moreover, we can assume that for $\mu = 1, \ldots, d$, the $r_\mu \times r_\mu$ matrices $(U^{\mu}_*)_+ U^{\mu}_X$ are invertible (again, since $U^{\mu}_X \to U^{\mu}_*$ for all μ for $X \to X_*$). We then also consider

$$\bar{C}_X = X \times_1 (U^1_*)_+ \times_2 \cdots \times_d (U^d_*)_+ = C_X \times_1 (U^1_*)_+ U^1_X \times_2 \cdots \times_d (U^d_*)_+ U^d_X. \tag{A.2}$$

Clearly, $\bar{C}_{X_*} = C_*$. Noting that by (4.2) the condition $C_X \in \mathcal{M}_{\mathrm{c}}$ in (4.1) is independent under invertible changes of basis, we arrive at the following local description of $\mathcal{M}$:

$$\mathcal{M} \cap \mathcal{O} = \left\{ X \in \mathcal{O} \colon \bar{C}_X \in \mathcal{M}_{\mathrm{c}},\ \operatorname{rank}(M^{\mu}_X) = r_\mu \text{for}\ \mu = 1, \ldots, d \right\}. \tag{A.3}$$

We next describe the constraints as preimages of smooth maps. We begin with the constraint $\bar{C}_X \in \mathcal{M}_c$. Since $\mathcal{M}_c$ is assumed to be an embedded submanifold of $\mathbb{R}_*^{r_1\times\cdots\times r_d}$, there exists a submersion ϕ from an open neighborhood of $C_* \in \mathcal{M}_c$ to $\mathbb{R}^q$ (here q is the co-dimension of $\mathcal{M}_c$) such that the conditions $C \in \mathcal{M}_c$ and $\phi(C) = 0$ are equivalent in this neighborhood. Considering

$$g_0 : \mathcal{O} \to \mathbb{R}^q, \quad X \mapsto \phi(\bar{C}_X),$$

we then locally have $\bar{C}_X \in \mathcal{M}_c$ if and only if $g_0(X) = 0$.

We now consider the rank constraints in (A.3), which first will be reformulated. Let $P_{\mathcal{U}_*^\mu} = U_*^\mu (U_*^\mu)_+$ denote the orthogonal projections on the μ-th minimal subspaces of X_*, and let

$$\mathbb{P}_\mu = P_{\mathcal{U}_*^1} \otimes \cdots \otimes P_{\mathcal{U}_*^\mu} \otimes \mathrm{id}_{\mu+1} \otimes \cdots \otimes \mathrm{id}_d,$$

with the convention $\mathbb{P}_0 = \mathrm{id}$. In the following we consider tensors $\mathbb{P}_{\mu-1}(X)$, that is, orthogonal projections of X onto the subspaces $\mathcal{U}_1^* \otimes \cdots \otimes \mathcal{U}_*^{\mu-1} \otimes \ell_2(\mathbb{N}^{d-\mu+1})$. In particular, we claim that for any fixed $1 \le \nu \le d$ there exists a neighborhood of X_* in which the condition $\mathrm{rank}(M_X^\mu) = r_\mu$ for $\mu = 1, \ldots, \nu$ is equivalent with $\mathrm{rank}(M^\mu_{\mathbb{P}_{\mu-1}(X)}) = r_\mu$ for all $\mu = 1, \ldots, \nu$. This is shown by induction over ν. For $\nu = 1$ the statement is trivial since $\mathbb{P}_0 = \mathrm{id}$. In the induction step $\nu - 1 \to \nu$, it suffices to show that in some neighborhood of X_*, any X satisfying $\mathrm{rank}(M_X^\mu) = r_\mu$ for $\mu = 1, \ldots, \nu-1$ also satisfies $\mathrm{rank}(M_X^\nu) = \mathrm{rank}(M^\nu_{\mathbb{P}_{\nu-1}(X)})$. Any such X lies in a subspace $\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^{\nu-1} \otimes \ell_2(\mathbb{N}^{d-\nu+1})$, where $\mathcal{U}^\mu$ are the minimal r_μ-dimensional subspaces of X. We choose a neighborhood of X_* in which the restrictions of all $P_{\mathcal{U}_*^\mu}$ to $\mathcal{U}^\mu$ are necessarily invertible maps between $\mathcal{U}^\mu$ and $\mathcal{U}_*^\mu$ (which is equivalent with $(U_*^\mu)_+ U_X^\mu$ being invertible). Hence in this neighborhood the projection $\mathbb{P}_{\nu-1}$ is a tensor product of invertible operators between $\mathcal{U}^1 \otimes \cdots \otimes \mathcal{U}^{\nu-1} \otimes \ell_2(\mathbb{N}^{d-\nu+1})$ and $\mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^{\nu-1} \otimes \ell_2(\mathbb{N}^{d-\nu+1})$, which hence leaves all matricization ranks invariant. Hence, for such X, we obtain that $\mathrm{rank}(M_X^\nu) = \mathrm{rank}(M^\nu_{\mathbb{P}_{\nu-1}X})$, which completes the induction.

Applying the above equivalence with $\nu = d$ allows us to replace the conditions $\mathrm{rank}(M_X^\mu) = r_\mu$ in (A.3) with $\mathrm{rank}(M^\mu_{\mathbb{P}_{\mu-1}(X)}) = r_\mu$ for $\mu = 1, \ldots, d$. These latter conditions are now handled via Schur complements as follows. Note that

$$M^\mu_{\mathbb{P}_{\mu-1}(X)} \in \ell_2(\mathbb{N}) \otimes [\mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^{\mu-1} \otimes \ell_2(\mathbb{N}^{d-\mu})].$$

We consider orthogonal decompositions

$$\ell_2(\mathbb{N}) = \mathcal{U}_*^\mu \oplus (\mathcal{U}_*^\mu)^\perp$$

and

$$\mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^{\mu-1} \otimes \ell_2(\mathbb{N}^{d-\mu}) = \mathcal{V}_*^\mu \oplus \mathcal{W}_*^\mu,$$

which is possible since $\mathcal{V}_*^\mu$ is even contained in the smaller subspace of $\mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^{\mu-1} \otimes \mathcal{U}_*^{\mu+1} \otimes \cdots \otimes \mathcal{U}_*^d$ (which in turn follows from $X_* \in \mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^d$). Hence in this notation

$$M_{\mathbb{P}_{\mu-1}(X)} \in [\mathcal{U}_*^\mu \oplus (\mathcal{U}_*)^\perp] \otimes [\mathcal{V}_*^\mu \oplus \mathcal{W}_*^\mu].$$

By applying a block decomposition of $M^\mu_{\mathbb{P}_{\mu-1}(X)}$ into the four corresponding parts,

$$\begin{aligned} Q_X^\mu &= P_{\mathcal{U}_*^\mu} M^\mu_{\mathbb{P}_{\mu-1}(X)} P_{\mathcal{V}_*^\mu} \in \mathcal{U}_*^\mu \otimes \mathcal{V}_*^\mu, \\ R_X^\mu &= P_{\mathcal{U}_*^\mu} M^\mu_{\mathbb{P}_{\mu-1}(X)} P_{\mathcal{W}_*^\mu} \in \mathcal{U}_*^\mu \otimes \mathcal{W}_*^\mu, \\ S_X^\mu &= (\mathrm{id} - P_{\mathcal{U}_*^\mu}) M^\mu_{\mathbb{P}_{\mu-1}(X)} P_{\mathcal{V}_*^\mu} \in (\mathcal{U}_*^\mu)^\perp \otimes \mathcal{V}_*^\mu, \\ T_X^\mu &= (\mathrm{id} - P_{\mathcal{U}_*^\mu}) M^\mu_{\mathbb{P}_{\mu-1}(X)} P_{\mathcal{W}_*^\mu} \in (\mathcal{U}_*^\mu)^\perp \otimes \mathcal{W}_*^\mu, \end{aligned}$$

we can consider the Schur complement functions

$$g_\mu : \mathcal{O} \to (\mathcal{U}_*^\mu)^\perp \otimes \mathcal{W}_*^\mu, \quad X \mapsto T_X^\mu - S_X^\mu (Q_X^\mu)^{-1} R_X^\mu. \tag{A.4}$$

Note that Q_X^μ is indeed invertible (as an $r_\mu \times r_\mu$ matrix in $\mathcal{U}_*^\mu \otimes \mathcal{V}_*^\mu$) for X close enough to X_*, since $Q_{X_*}^\mu = P_{\mathcal{U}_*^\mu} M_{X_*}^\mu P_{\mathcal{V}_*^\mu}$ is invertible. As for finite matrices, we then have $g_\mu(X) = 0$ if and only if $\operatorname{rank}(M^\mu_{\mathbb{P}_{\mu-1}(X)}) = r_\mu$.

Defining

$$g = (g_0, g_1, \ldots, g_d) \colon \mathcal{O} \to \mathbb{R}^q \times [(\mathcal{U}_*^1)^\perp \otimes \mathcal{W}_*^1] \times \cdots \times [(\mathcal{U}_*^d)^\perp \otimes \mathcal{W}_*^d]$$

we conclude from all the previous considerations that (A.3) can be written as

$$\mathcal{M} \cap \mathcal{O} = g^{-1}(0).$$

We need to show that g is a submersion in X_*, that is, $g'(X_*)$ is surjective. First note that for $\mu = 1, \ldots, d$ we have

$$g_\mu'(X_*)[H] = T_H^\mu = (\mathrm{id} - P_{\mathcal{U}_*^\mu}) M^\mu_{\mathbb{P}_{\mu-1}(H)} P_{\mathcal{W}_*^\mu}, \tag{A.5}$$

that is, $g_\mu'(X_*)$ is the orthogonal projection (of the μ-th matricization) onto the subspace $(\mathcal{U}_*^\mu)^\perp \otimes \mathcal{W}_*^\mu$. This follows by applying a product rule to (A.4) and noting that $R_{X_*}^\mu = 0$ and $S_{X_*}^\mu = 0$. When viewed as subspaces of $\ell_2(\mathbb{N}^d)$, the subspaces $(\mathcal{U}_*^\mu)^\perp \otimes \mathcal{W}_*^\mu$ are mutually orthogonal to each other, since they are contained in the pairwise orthogonal subspaces $\mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^{\mu-1} \otimes (\mathcal{U}_*^\mu)^\perp \otimes \ell_2(\mathbb{N}^{d-\mu})$, respectively. Moreover, all of them are orthogonal to the subspace $\mathcal{U}_*^1 \otimes \cdots \otimes \mathcal{U}_*^d$. Regarding g_0, note that

$$g_0(C \times_1 U_*^1 \times_2 \cdots \times_d U_*^d) = \phi(C) \tag{A.6}$$

(again with $U^\mu_{X_*} = U^\mu_*$), which shows that already the restriction of g_0 to $\mathcal{U}^1_* \otimes \cdots \otimes \mathcal{U}^d_*$ is a submersion in X_*, since $\phi'(C_*)$ is surjective to $\mathbb{R}^q$. It is now easy to conclude from these facts that $g'(X_*)$ is altogether surjective.

By the local submersion theorem in Hilbert space, see [36, Thm. 73.C], $\mathcal{M} \cap \mathcal{O} = g^{-1}(0)$ is a smooth submanifold of $\mathcal{H}$. The tangent space $T_{X_*}\mathcal{M}$ at X_* is the null space of the $g'(X_*)$. The proof of part (i) is therefore completed.

Ad (ii). The existence of the continuously Fréchet-differentiable homeomorphism φ of the asserted form is a consequence of Ljusternik's submersion theorem as stated in [35, Thm. 43.C]. To show that (in a possibly smaller neighborhood around zero) φ is also an immersion, that is, $\varphi'(\xi) \colon T_{X_*}\mathcal{M} \to T_{\varphi(\xi)}\mathcal{M}$ is injective and its range splits, it suffices to show that there exists $c > 0$ such that $\|\varphi'(\xi)h\| \geq c\|h\|$ for all $h \in T_{X_*}\mathcal{M}$. This, however, follows immediately from the continuity of φ' and $\varphi'(0)h = h$. By definition, φ is therefore a local embedding [36, Def. 73.43].

Ad (iii). Let ξ be an element of the form (4.5) (but at $X_* = C_* \times_1 U^1_* \times_2 \cdots \times_d U^d_*$). Since $\dot{C} \in T_{C_*}\mathcal{M}_{\mathrm{c}}$, there exists a curve $C(t)$ in $\mathcal{M}_{\mathrm{c}}$ such that $C(0) = C_*$ and $C'(0) = \dot{C}$. For small enough t,

$$X(t) = C(t) \times_1 (U^1_* + t\dot{U}^1) \times_2 \cdots \times_d (U^d_* + t\dot{U}^d)$$

then defines a curve in $\mathcal{M}$ because $U^\mu_* + t\dot{U}^\mu$ has full column rank for $\mu = 1, \dots, d$. Obviously $X(0) = X_*$, and by multilinearity it is easily seen that $X'(0) = \xi$, which shows $\xi \in T_{X_*}\mathcal{M}$.

In order to show that all tangent vectors are of the form (4.5), let $\xi \in T_{X_*}\mathcal{M}$ and a corresponding curve $X(t) \in \mathcal{M}$ with $X(0) = X_*$ and $X'(0) = \xi$ be given. For small enough t we represent $X(t)$ in the particular bases $U^\mu_{X(t)}$ defined in (A.1) as

$$X(t) = C_{X(t)} \times_1 U^1_{X(t)} \times_2 \cdots \times_d U^d_{X(t)},$$

where $C_{X(t)}$ is in $\mathcal{M}_{\mathrm{c}}$. Clearly, the curves $t \mapsto U^\mu_{X(t)}$ are smooth. It implies that for small enough t the pseudoinverses $t \mapsto (U^\mu_{X(t)})_+$ are also smooth functions. It then follows from (A.2), by applying an inverse transformation, that also $t \mapsto C_{X(t)}$ is a smooth curve, since $\bar{C}_{X(t)}$ is. By the product rule we then get that

$$\xi = X'(0) = \tilde{C} \times_1 U^1_* \times_2 \cdots \times_d U^d_* + C_* \times_1 \tilde{U}^1 \times_2 \cdots \times_d U^d_* + \cdots + C_* \times_1 U^1_* \times_2 \cdots \times_d \tilde{U}^d, \tag{A.7}$$

where $\tilde{C} \in T_{C_*}\mathcal{M}_{\mathrm{c}}$ is the derivative of $t \mapsto C_{X(t)}$ in $t = 0$, and $\tilde{U}^\mu \in (\ell_2(\mathbb{N}))^{r_\mu}$ is the derivative of $t \mapsto U^\mu_{X(t)}$ in $t = 0$ for $\mu = 1, \dots, d$. By decomposing every column of $\tilde{U}^\mu$ into the span of U^μ_* and its orthogonal complement, we can write

$$\tilde{U}^\mu = U^\mu_* S_\mu + \dot{U}^\mu_*,$$

where S_μ is some $r_\mu \times r_\mu$ matrix and $(U^\mu_*)^{\mathsf{T}} \dot{U}^\mu_* = 0$. Expanding the expression (A.7) we then have

$$\xi \ = K \times_1 U^1_* \times_2 \cdots \times_d U^d_* + C_* \times_1 \dot{U}^1_* \times_2 \cdots \times_d U^d_* + \cdots + C_*$$

$$\times_1 U_*^1 \times_2 \cdots \times_d \dot{U}_*^d$$

where

$$K = \tilde{C} + C_* \times_1 S_1 \times_2 \mathrm{id} \times_3 \cdots \times_d \mathrm{id} + \cdots + C_* \times_1 \mathrm{id} \times_2 \cdots \times_d S_d.$$

It remains to show that $K \in T_{C_*}\mathcal{M}_\mathrm{c}$ to conclude that ξ is of the asserted form (4.5). Since $T_{C_*}\mathcal{M}_\mathrm{c}$ is a linear space it suffices to show this for every term. For $\tilde{C}$ there is nothing to show. The sum of the remaining terms equals the derivative of the curve

$$C(t) = C_* \times_1 (\mathrm{id} + tS_1) \times_2 \cdots \times_d (\mathrm{id} + tS_d)$$

at $t = 0$, which for small enough t lies in $\mathcal{M}_\mathrm{c}$ by the invariance condition (4.2). Hence $C'(0) \in T_{C_*}\mathcal{M}_\mathrm{c}$. □

Proof of Proposition 4.11 Let $X^{\{1,\ldots,\mu\}} \in \mathbb{R}^{N_1\cdots N_\mu \times B_{\mu+1}\cdots N_d}$ be a matricization of X. We can decompose

$$X^{\{1,\ldots,\mu\}} = (U_1 \otimes \mathrm{id}_{N_2\cdots N_\mu}) \cdots (U_{\mu-1} \otimes \mathrm{id}_{N_\mu}) U_\mu \Sigma_\mu V_{\mu+1}^\mathsf{T} \cdots (\mathrm{id}_{N_{\mu+1}\cdots N_{d-1}} \otimes V_d^\mathsf{T}),$$

with $U_\nu^\mathsf{T} U_\nu = \mathrm{id}_{r_\nu}$ and $V_\nu^\mathsf{T} V_\nu = \mathrm{id}_{r_\nu}$. Furthermore, we define the spaces $\mathcal{U}_{\{1,\ldots,\mu\}}$ and $\mathcal{V}_{\{\mu+1,\ldots,d\}}$ via their respective orthonormal bases

$$U_{\{1,\ldots,\mu\}} = (U_1 \otimes \mathrm{id}_{N_2\cdots N_\mu}) \cdots (U_{\mu-1} \otimes \mathrm{id}_{N_\mu}) U_\mu$$

and

$$V_{\mu+1,\ldots,d}^\mathsf{T} = V_{\mu+1}^\mathsf{T} \cdots (\mathrm{id}_{N_{\mu+1}\cdots N_{d-1}} \otimes V_d^\mathsf{T}).$$

The projection P_X can be decomposed as

$$P_X = P_1^X + \ldots + P_d^X,$$

where

$$P_\mu^X = (P_{\mathcal{U}_{\{1,\ldots,\mu-1\}}} \otimes \mathrm{id}_{N_\mu} - P_{\mathcal{U}_{\{1,\ldots,\mu\}}}) \otimes P_{\mathcal{V}_{\{\mu+1,\ldots,d\}}}$$

for $\mu = 1, \ldots, d-1$ and

$$P_d^X = P_{\mathcal{U}_{\{1,\ldots,d-1\}}} \otimes \mathrm{id}_{N_d};$$

see, e.g. [25] or [34, Section 9.3.4]. Let $\tilde{\mathcal{U}}_{\{1,\ldots,\mu\}}$ and $\tilde{\mathcal{V}}_{\{\mu+1,\ldots,d\}}$ be the analogous spaces for Y. Then by (4.13),

$$\|P_\mu^X - P_\mu^Y\| = \|(P_{\mathcal{U}_{\{1,\ldots,\mu-1\}}} \otimes \mathrm{id}_{N_\mu} - P_{\mathcal{U}_{\{1,\ldots,\mu\}}}) \otimes P_{\mathcal{V}_{\{\mu+1,\ldots,d\}}}$$

$$
\begin{aligned}
&- (P_{\tilde{\mathcal{U}}_{\{1,\ldots,\mu-1\}}} \otimes \mathrm{id}_{N_\mu} - P_{\tilde{\mathcal{U}}_{\{1,\ldots,\mu\}}}) \otimes P_{\tilde{\mathcal{V}}_{\{\mu+1,\ldots,d\}}} \| \\
&\leq \| (P_{\mathcal{U}_{\{1,\ldots,\mu-1\}}} - P_{\tilde{\mathcal{U}}_{\{1,\ldots,\mu-1\}}}) \otimes \mathrm{id}_{N_\mu} \otimes P_{\mathcal{V}_{\{\mu+1,\ldots,d\}}} \| \\
&\quad + \| P_{\tilde{\mathcal{U}}_{\{1,\ldots,\mu-1\}}} \otimes \mathrm{id}_{N_\mu} \otimes (P_{\mathcal{V}_{\{\mu+1,\ldots,d\}}} - P_{\tilde{\mathcal{V}}_{\{\mu+1,\ldots,d\}}}) \| \\
&\quad + \| (P_{\mathcal{U}_{\{1,\ldots,\mu\}}} - P_{\tilde{\mathcal{U}}_{\{1,\ldots,\mu\}}}) \otimes P_{\mathcal{V}_{\{\mu+1,\ldots,d\}}} \| \\
&\quad + \| P_{\tilde{\mathcal{U}}_{\{1,\ldots,\mu\}}} \otimes (P_{\mathcal{V}_{\{\mu+1,\ldots,d\}}} - P_{\tilde{\mathcal{V}}_{\{\mu+1,\ldots,d\}}}) \| \\
&\leq \frac{4}{\sigma} \| X - Y \|
\end{aligned}
$$

holds and the first desired inequality readily follows. For the other inequalities, we use the decomposition of the identity matrix

$$
\begin{aligned}
\mathrm{id} = {} & P_{\mathcal{U}_{\{1,\ldots,d-1\}}} \otimes \mathrm{id}_{N_d} + (P_{\mathcal{U}_{\{1,\ldots,d-2\}}} \otimes \mathrm{id}_{N_{d-1}} - P_{\mathcal{U}_{\{1,\ldots,d-1\}}}) \otimes \mathrm{id}_{N_d} \\
& + \ldots + (\mathrm{id}_{N_1} - P_{\mathcal{U}_{\{1\}}}) \otimes \mathrm{id}_{N_2 \ldots N_d}
\end{aligned}
$$

into orthogonal projections onto mutually orthogonal spaces. Then

$$
\begin{aligned}
(\mathrm{id} - P_X)(X - Y) &= (P_{\mathcal{U}_{\{1,\ldots,d-2\}}} \otimes \mathrm{id}_{N_{d-1}} - P_{\mathcal{U}_{\{1,\ldots,d-1\}}}) \otimes (\mathrm{id}_{N_{d-1} N_d} - P_{\mathcal{V}_{\{d\}}})(X - Y) \\
&\quad + \ldots \\
&\quad + (\mathrm{id}_{N_1} - P_{\mathcal{U}_{\{1\}}}) \otimes (\mathrm{id}_{N_2 \ldots N_d} - P_{\mathcal{V}_{\{2,\ldots,d\}}})(X - Y) \\
&= (P_{\mathcal{U}_{\{1,\ldots,d-2\}}} \otimes \mathrm{id}_{N_{d-1}} - P_{\mathcal{U}_{\{1,\ldots,d-1\}}}) \otimes (P_{\tilde{\mathcal{V}}_{\{d\}}} - P_{\mathcal{V}_{\{d\}}})(X - Y) \\
&\quad + \ldots \\
&\quad + (\mathrm{id}_{N_1} - P_{\mathcal{U}_{\{1\}}}) \otimes (P_{\tilde{\mathcal{V}}_{\{2,\ldots,d\}}} - P_{\mathcal{V}_{\{2,\ldots,d\}}})(X - Y)
\end{aligned}
$$

holds. Note that the operators map onto orthogonal subspaces. Hence, we get the desired estimate

$$
\| (\mathrm{id} - P_X)(X - Y) \| \leq \frac{\sqrt{d-1}}{\sigma} \| X - Y \|^2 .
$$

By continuity of the projection and taking limits, the estimate also holds for $Y \in \overline{\mathcal{M}_{\mathrm{c}}}$. In a similar way, the second inequality follows. □

Acknowledgements M.B. acknowledges funding by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – Projektnummern 442047500, 501389786. The work of H.E. was funded by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 501389786. The work of A.U. was supported by Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 506561557.

Funding Open Access funding enabled and organized by Projekt DEAL.

References

1. Arnold, A., Jahnke, T.: On the approximation of high-dimensional differential equations in the hierarchical Tucker format. BIT **54**(2), 305–341 (2014)
2. Bachmayr, M.: Low-rank tensor methods for partial differential equations. Acta Numer. **32**, 1–121 (2023)
3. Bachmayr, M., Eisenmann, H., Kieri, E., Uschmajew, A.: Existence of dynamical low-rank approximations to parabolic problems. Math. Comp. **90**(330), 1799–1830 (2021)
4. Bachmayr, M., Schneider, R., Uschmajew, A.: Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. Found. Comput. Math. **16**(6), 1423–1472 (2016)
5. Bardos, C., Catto, I., Mauser, N., Trabelsi, S.: Setting and analysis of the multi-configuration time-dependent Hartree-Fock equations. Arch. Ration. Mech. Anal. **198**(1), 273–330 (2010)
6. Bardos, C., Catto, I., Mauser, N.J., Trabelsi, S.: Global-in-time existence of solutions to the multi-configuration time-dependent Hartree-Fock equations: a sufficient condition. Appl. Math. Lett. **22**(2), 147–152 (2009)
7. Ceruti, G., Lubich, C., Walach, H.: Time integration of tree tensor networks. SIAM J. Numer. Anal. **59**(1), 289–313 (2021)
8. Conte, D., Lubich, C.: An error analysis of the multi-configuration time-dependent Hartree method of quantum dynamics. M2AN Math. Model. Numer. Anal. **44**(4), 759–780 (2010)
9. Dirac, P.A.M.: Note on exchange phenomena in the Thomas atom. Math. Proc. Cambridge Philos. Soc. **26**(3), 376–385 (1930)
10. Einkemmer, L., Lubich, C.: A low-rank projector-splitting integrator for the Vlasov-Poisson equation. SIAM J. Sci. Comput. **40**(5), B1330–B1360 (2018)
11. Einkemmer, L., Ostermann, A., Piazzola, C.: A low-rank projector-splitting integrator for the Vlasov-Maxwell equations with divergence correction. J. Comput. Phys. **403**, 109063 (2020)
12. Eisenmann, H.: Multilinear Optimization in Low-rank Models. PhD thesis, Universität Leipzig (2023)
13. Falcó, A., Hackbusch, W., Nouy, A.: On the Dirac-Frenkel variational principle on tensor Banach spaces. Found. Comput. Math. **19**(1), 159–204 (2019)
14. Falcó, A., Hackbusch, W., Nouy, A.: Tree-based tensor formats. SeMA J. **78**(2), 159–173 (2021)
15. Frenkel, J.: Advanced general theory. In: Wave Mechanics. Clarendon Press, Oxford (1934)
16. Guo, W., Qiu, J.-M.: A low rank tensor representation of linear transport and nonlinear Vlasov solutions and their associated flow maps. J. Comput. Phys. **458**, 111089, 24 (2022)
17. Hackbusch, W.: Tensor Spaces and Numerical Tensor Calculus, 2nd edn. Springer, Cham (2019)
18. Holtz, S., Rohwedder, T., Schneider, R.: On manifolds of tensors of fixed TT-rank. Numer. Math. **120**(4), 701–731 (2012)
19. Kazashi, Y., Nobile, F.: Existence of dynamical low rank approximations for random semi-linear evolutionary equations on the maximal interval. Stoch. Part. Differ. Equ. Anal. Comput. **9**(3), 603–629 (2021)
20. Koch, O., Lubich, C.: Dynamical low-rank approximation. SIAM J. Matrix Anal. Appl. **29**(2), 434–454 (2007)
21. Koch, O., Lubich, C.: Dynamical tensor approximation. SIAM J. Matrix Anal. Appl. **31**(5), 2360–2375 (2010)
22. Kolda, T.G., Bader, B.W.: Tensor decompositions and applications. SIAM Rev. **51**(3), 455–500 (2009)
23. Lubich, C.: Time integration in the multiconfiguration time-dependent Hartree method of molecular quantum dynamics. Appl. Math. Res. Express. AMRX **2**, 311–328 (2015)
24. Lubich, C., Oseledets, I.V.: A projector-splitting integrator for dynamical low-rank approximation. BIT **54**(1), 171–188 (2014)
25. Lubich, C., Oseledets, I.V., Vandereycken, B.: Time integration of tensor trains. SIAM J. Numer. Anal. **53**(2), 917–941 (2015)

26. Lubich, C., Rohwedder, T., Schneider, R., Vandereycken, B.: Dynamical approximation by hierarchical Tucker and tensor-train tensors. SIAM J. Matrix Anal. Appl. **34**(2), 470–494 (2013)
27. Meyer, H.-D., Manthe, U., Cederbaum, L.S.: The multi-configurational time-dependent Hartree approach. Chem. Phys. Lett. **165**(1), 73–78 (1990)
28. Nonnenmacher, Achim, Lubich, Christian: Dynamical low-rank approximation: applications and numerical experiments. Math. Comput. Simul. **79**(4), 1346–1357 (2008)
29. Oseledets, I.V.: Tensor-train decomposition. SIAM J. Sci. Comput. **33**(5), 2295–2317 (2011)
30. Steinlechner, M.M.: Riemannian Optimization for Solving High-Dimensional Problems with Low-Rank Tensor Structure. PhD thesis, MATHICSE, EPFL, Lausanne (2016)
31. Teschl, G.: Ordinary Differential Equations and Dynamical Systems. American Mathematical Society, Providence, RI (2012)
32. Uschmajew, A.: Regularity of tensor product approximations to square integrable functions. Constr. Approx. **34**(3), 371–391 (2011)
33. Uschmajew, A., Vandereycken, B.: The geometry of algorithms using hierarchical tensors. Linear Algebra Appl. **439**(1), 133–166 (2013)
34. Uschmajew, A., Vandereycken, B.: Geometric methods on low-rank matrix and tensor manifolds. In: Handbook of variational methods for nonlinear geometric data, pp. 261–313. Springer, Cham (2020)
35. Zeidler, E.: Nonlinear Functional Analysis and Its Applications. III. Variational Methods and Optimization. Springer-Verlag, New York (1985)
36. Zeidler, E.: Nonlinear Functional Analysis and Its Applications. IV. Applications to Mathematical Physics. Springer-Verlag, New York (1988)
37. Zeidler, E.: Nonlinear Functional Analysis and Its Applications. II/A. Linear monotone operators. Springer-Verlag, New York (1990)
38. Zeidler, E.: Nonlinear Functional Analysis and Its Applications. II/B. Nonlinear Monotone Operators. Springer-Verlag, New York (1990)