Max-Planck-Institut für Mathematik in den Naturwissenschaften Leipzig

Existence of dynamical low-rank approximations to parabolic problems

(revised version: April 2020)

by

Markus Bachmayr, Henrik Eisenmann, Emil Kieri, and André Uschmajew

Preprint no.: 33 2020

EXISTENCE OF DYNAMICAL LOW-RANK APPROXIMATIONS TO PARABOLIC PROBLEMS

MARKUS BACHMAYR, HENRIK EISENMANN, EMIL KIERI, AND ANDRÉ USCHMAJEW

Abstract. The existence and uniqueness of weak solutions of dynamical low-rank evolution for parabolic partial differential equations in two spatial dimensions is shown, covering also non-diagonal diffusion in the elliptic part. The proof is based on a variational time-stepping scheme on the low-rank manifold. Moreover, this scheme is shown to be closely related to practical methods for computing such low-rank evolutions.

1. Introduction

Finding hidden structure in the solutions of partial differential equations has always been a key goal in the study of such equations, whether it is for the sake of modeling or for efficient numerical approximation. In fact, exploiting structures such as low-dimensional parametrizations can be crucial for the numerical treatment of equations on high-dimensional domains to avoid the curse of dimensionality.

It has been observed that under certain conditions on the domain and the data, the solutions of elliptic and parabolic partial differential equations with a dominating "Laplacian part" exhibit low-rank approximability, that is, they can be approximated in certain low-rank tensor formats [19, 42, 11, 3]. If this is the case, then instead of working on full discretization grids, one can impose the low-rank constraint in the design of the solution method in order to take advantage of low-parametric representation. This typically results in a nonlinear approximation algorithm.

A typical approach is to discretize the partial differential equation on possibly huge, but finite grids, and then use numerical linear algebra techniques for solving the resulting linear systems in low-rank formats; see, e.g, [6, 20] for an overview and further references. How the obtained solutions behave with refinement of discretization depends strongly on the details of the considered methods. This point has been considered for methods that adjust solution ranks adaptively in each step [4, 5]. For methods based on a fixed low-rank constraint, this question is more difficult due to the nonlinearity of the resulting constrained problems and has found only limited attention in the literature. Since methods operating on fixed-rank manifolds are important algorithmic building blocks, understanding their robustness under discretization refinement is of high practical interest. A first important requirement is to study the well-posedness of the underlying low-rank problem on function spaces. While it is not so difficult to make an appropriate variational formulation for elliptic problems subject to low-rank constraints that ensure existence of solutions [6, Sec. 4], the parabolic case poses substantial difficulties. In this paper we propose such a formulation for parabolic evolution equations on low-rank manifolds in Hilbert space and prove existence of solutions via a time-stepping scheme.

M.B., E.K., and A.U. acknowledge support by the Hausdorff Center of Mathematics, University of Bonn. M.B. was supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 211504053 - SFB 1060.

Dynamical low-rank approximation is a general technique for approximating time-dependent problems under low-rank constraints by projecting the vector field onto the tangent space of the low-rank manifold. For general initial value problems $\dot{Y} = F(t, Y)$, $Y(0) = Y_0$ for matrices $Y(t)$, the dynamical low-rank approximation on the manifold $\mathcal{M}_r$ of rank-r matrices as considered in [25] is given by

$$\dot{Y}(t) = P_{Y(t)} F\big(t, Y(t)\big), \tag{1.1}$$

where $P_{Y(t)}$ is the orthogonal projector onto the tangent space $T_{Y(t)}\mathcal{M}_r$. Note that (1.1) is equivalent to the variational problem

$$\langle \dot{Y}(t) - F(t, Y(t)), X\rangle = 0 \quad \text{for all } X \in T_{Y(t)}\mathcal{M}_r,$$

in analogy to (1.7); this approach is also known as the Dirac-Frenkel variational principle [13, 29]. It has been adapted to several different classes of evolution problems in scientific computing, see, e.g., [40, 22, 37, 36, 34, 14] as well as [44] for an overview, and the monograph [29] on applications in quantum dynamics.

In this work, we develop a weak formulation of the Dirac-Frenkel principle for low-rank approximation of parabolic problems and prove the existence and uniqueness of solutions in a function space setting. As a model problem one may consider the two-dimensional parabolic equation on the product domain $\Omega = (0, 1)^2$,

$$\begin{aligned} u_t(x,t) - \nabla \cdot \alpha(t) \nabla u(x,t) &= f(x,t) && \text{for } (x,t) \in \Omega \times (0,T), \\ u(x,t) &= 0 && \text{for } (x,t) \in \partial\Omega \times (0,T), \\ u(x,0) &= u_0(x) && \text{for } x \in \Omega. \end{aligned} \tag{1.2}$$

Here we assume that the matrix $\alpha(t) = (\alpha_{ij}(t))_{i,j=1,2}$ is symmetric for every t, uniformly bounded, and uniformly positive definite. The problem (1.2) is typically formulated in weak form as follows: given $f \in L_2(0, T; H^{-1}(\Omega))$ and $u_0 \in L_2(\Omega)$, find

$$u \in W_2^1(0, T; H_0^1(\Omega), L_2(\Omega)) = \{u \in L_2(0, T; H_0^1(\Omega) \colon \exists u' \in L_2(0, T; H^{-1}(\Omega))\}$$

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

$$\begin{aligned} \langle u'(t), v\rangle + a(u(t), v; t) &= \langle f(t), v\rangle \quad \text{for all } v \in H_0^1(\Omega), \\ u(0) &= u_0. \end{aligned} \tag{1.3}$$

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

$$\begin{aligned} a(u, v; t) = \alpha_{11}(t) &\int_\Omega \partial_1 u(x,t)\, \partial_1 v(x,t)\, \mathrm{d}x + \alpha_{22}(t) \int_\Omega \partial_2 u(x,t)\, \partial_2 v(x,t)\, \mathrm{d}x \\ &+ \alpha_{12}(t) \int_\Omega \partial_1 u(x,t)\, \partial_2 v(x,t)\, \mathrm{d}x + \alpha_{21}(t) \int_\Omega \partial_2 u(x,t)\, \partial_1 v(x,t)\, \mathrm{d}x. \end{aligned} \tag{1.4}$$

By classical theory the problem (1.3) admits a unique solution; see, e.g., [46, Thm. 23.A].

Since $\Omega = (0, 1)^2$, we have $L_2(\Omega) = L_2(0, 1) \otimes L_2(0, 1)$ in the sense of tensor products of Hilbert spaces, and $H_0^1(\Omega) = H_0^1(0, 1) \otimes L_2(0, 1) \cap L_2(0, 1) \otimes H_0^1(0, 1)$ with norm

$$\|v\|^2_{H_0^1(\Omega)} = \|v\|^2_{H_0^1(0,1)\otimes L_2(0,1)} + \|v\|^2_{L_2(0,1)\otimes H_0^1(0,1)}.$$

Every function $u \in L_2(\Omega)$ can be written as

$$u(x) = u(x_1, x_2) = \sum_{k=1}^{r} u_k^1(x_1)\, u_k^2(x_2) \quad \text{a.e.,} \tag{1.5}$$

with $u_k^1, u_k^2 \in L_2(0,1)$ for all k, where r can be infinite. By $\operatorname{rank}(u)$ we denote the smallest r such that such a representation exists.

As low-rank representations are convenient for several reasons, one may ask whether the parabolic equation (1.2) admits approximate solutions of low-rank. In dynamical low-rank approximation one assumes this to be the case, and attempts to directly evolve the solution on the set

$$\mathcal{M}_r = \{u \in L_2(\Omega)\colon\ \operatorname{rank}(u) = r\} \tag{1.6}$$

for a certain value of r. One can show that $\mathcal{M}_r$ is a submanifold in $L_2(\Omega)$. The dynamics are then determined by the following problem: find $u \in W_2^1(0,T;H_0^1(\Omega), L_2(\Omega))$ such that

$$u(t) \in \mathcal{M}_r \quad \text{for all } t \in [0,T],$$

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

$$\begin{aligned} \langle u'(t), v\rangle + a(u(t), v; t) = \langle f(t), v\rangle \quad \text{for all } v \in T_{u(t)}\mathcal{M}_r \cap H_0^1(\Omega), \\ u(0) = u_0 \in \mathcal{M}_r, \end{aligned} \tag{1.7}$$

where $T_{u(t)}\mathcal{M}_r$ is the tangent space of the manifold $\mathcal{M}_r$ at $u(t)$. Thus, in contrast to (1.3), in (1.7) we seek a curve $t \mapsto u(t)$ on the manifold $\mathcal{M}_r$ which for almost every $t \in (0,T)$ satisfies the weak parabolic formulation (1.3) on the tangent space only.

Our goal in this paper is to provide an abstract framework for dealing with problems of the type (1.7), and to prove existence of solutions via a time-stepping scheme. In contrast to previous works, we do not require the diffusion matrix α to be diagonal, which means that we allow anisotropic diffusion. If α is diagonal, that is, $\alpha_{12} = \alpha_{21} = 0$, the problem is substantially easier; in particular, in this case the exact solution of the homogeneous equation with $f = 0$ and $u_0 \in \mathcal{M}_r$ satisfies $u(t) = (\exp(t\,\alpha_{11}\,\partial_1^2) \otimes \exp(t\,\alpha_{22}\,\partial_2^2))u_0 \in \mathcal{M}_r$ for all t. In the case of non-diagonal α, the unbounded operator on $L_2(\Omega)$ induced by the bilinear form a no longer maps to the tangent space of the manifold, which means that previously used techniques are no longer applicable in this setting.

Our existence proof is based on a Rothe-type temporal semidiscretization using minimization problems on $\mathcal{M}_r$ in each time step. Off-diagonal parts in the diffusion are treated via bounds on mixed derivatives that are always available for elements in the intersection $\mathcal{M}_r \cap H_0^1(\Omega)$, which is a remarkable aspect of the interplay between low-rank structures and regularity in function spaces. We require slightly more regularity of u_0 and f than necessary for standard parabolic problems in linear spaces like (1.3), but still less than needed for strong solutions. Specifically, applied to the model problem (1.3), our abstract results give solutions to the dynamical low-rank formulation (1.7) under the assumptions $u_0 \in \mathcal{M}_r \cap H_0^1(\Omega)$ and $f \in L_2(0,T;L_2(\Omega))$, as long as the smallest singular values in the low-rank representation of $u(t)$ do not approach zero. Compared to previous works, we do not make use of components in low-rank representations, but treat the problem directly on the manifold. This allows for generalization to evolutions on more general manifolds. We obtain uniqueness of solutions in the same abstract setting under a mild additional integrability assumption, which automatically holds for the model problem (1.7).

Beyond the comparably well-developed analysis of dynamical low-rank approximations in finite-dimensional spaces [25, 1, 24, 17, 39], the available results for low-rank evolution problems in function spaces cover mainly Schrödinger-type equations [29], in particular the closely related higher-dimensional generalization of the multi-configuration time-dependent Hartree method (MCTDH) considered in [35, 26, 8, 7, 28, 16]. An important ingredient in many results is the decomposition of the operators into a Laplacian part, which maps the low-rank manifold to its tangent space, and a potential term satisfying suitable boundedness properties. A very similar decomposition with differential operators mapping to the tangent space is also assumed in the recent work [23] on parameter-dependent parabolic problems, where the separation of variables is done not between spatial variables as considered here, but rather between the spatial and the parametric variables. An error analysis for such an approach was presented in [37].

The paper is organized as follows: in Section 2 we give an abstract formulation of the problem for general evolution equations on manifolds under assumptions that reflect the main features of the model problem (1.7). In Section 3, we introduce the time-stepping scheme that is used to approximate solutions. Then we show in Section 4 that this scheme yields a solution to the continuous problem in the limit, with uniqueness ensured under a minor additional integrability assumption. Section 5 is devoted to questions of numerical approximation. We give an outlook on directions for further work in Section 6.

2. Abstract formulation

Before we switch to an abstract model for our existence proof, we highlight some particular properties of the model problem (1.7) that will motivate the assumptions made in the abstract setting. We believe that the general formulation presented in Section 2.2 will be useful to study parabolic problems on more general low-rank tensor manifolds in tensor product Hilbert spaces of higher order, for instance $L^2((0,1)^d)$, as well. Low-rank tensor formats with suitable properties may include Tucker tensors [12], hierarchical Tucker tensors [21], and tensor trains [38].

2.1. Some features of the model problem on $\Omega = (0,1)^2$.

Let us first inspect the rank-r manifold $\mathcal{M}_r$ defined in (1.6) in more detail. We have already mentioned that it is an embedded Hilbert submanifold of $L_2(\Omega)$, but is not closed. In fact, its closure $\overline{\mathcal{M}_r}$ is the set $\mathcal{M}_{\le r}$ of all $u \in L_2(\Omega)$ with $\operatorname{rank}(u) \le r$ and this closure is even weakly sequentially closed; see, e.g., [20, Lemma 8.6]. In other words,

$$\mathcal{M}_{\le r} = \mathcal{M}_{\le r-1} \cup \mathcal{M}_r = \overline{\mathcal{M}_r} = \overline{\mathcal{M}_r}^w,$$

where the superscript w indicates the weak sequential closure. Another important property of $\mathcal{M}_r$ is that it is a cone, that is, $u \in \mathcal{M}_r$ implies $su \in \mathcal{M}_r$ for all $s > 0$.

2.1.1. *Tangent spaces.* For convenience let us use the notation $u^1 \otimes u^2$ for the tensor product of two $L_2(0,1)$ functions, that is, $(u^1 \otimes u^2)(x_1, x_2) = u^1(x_1)\, u^2(x_2)$ a.e. Every $u \in \mathcal{M}_r$ admits infinitely many representations of the form (1.5), among which the *singular value decomposition* (SVD)

$$u = \sum_{k=1}^{r} \sigma_k u_k^1 \otimes u_k^2 \tag{2.1}$$

is of great importance for the geometric description of the manifold. In (2.1), (u_k^1) and (u_k^2) are both $L_2(0,1)$-orthonormal systems, and $(\sigma_k) = (\sigma_k(u))$ is a non-increasing, positive sequence of *singular values.* The existence of such a decomposition is well known in any tensor product of Hilbert spaces [20, Thm. 4.137].

Given (2.1), the tangent space to $\mathcal{M}_r$ at u can be written as

$$T_u\mathcal{M}_r = \left\{ v = \sum_{k=1}^{r} v_k^1 \otimes u_k^2 + u_k^1 \otimes v_k^2 : \; v_k^1, v_k^2 \in L_2(0,1) \right\}. \tag{2.2}$$

To see this, consider a curve

$$\phi(t) = \sum_{k=1}^{r} \sigma_k \left(u_k^1 + t\sigma_k^{-1} v_k^1\right) \otimes \left(u_k^2 + t\sigma_k^{-1} v_k^2\right) \tag{2.3}$$

in $\mathcal{M}_r$. Then $\phi(0) = u$ and $\phi'(0) = v$ is of the form (2.2). One can show that every admissible curve in $\mathcal{M}_r$ through u is locally of this form, using the orthogonality of the factors in the SVD. Note that $u \in T_u\mathcal{M}_r$, which is also clear due to the cone property.

Without loss of generality, we could add the gauging conditions

$$(v_k^1, u_\ell^1)_{L_2(0,1)} = 0 \quad \text{for all } k, \ell, \qquad (v_k^1, v_\ell^1)_{L_2(0,1)} = 0 \quad \text{for } k \neq \ell \tag{2.4}$$

to the definition of $T_u\mathcal{M}_r$, where $(\cdot,\cdot)_{L_2(0,1)}$ is the inner product in $L_2(0,1)$. Then the representation of tangent vectors becomes unique. With these gauging conditions it is not difficult to show that $T_u\mathcal{M}_r$ is closed in $L_2(\Omega)$ and locally homeomorphic (around zero) to a neighborhood of u in $\mathcal{M}_r$, using essentially the same construction as (2.3). As a result, $\mathcal{M}_r$ is a manifold, e.g. in the sense of [45, Def. 43.10], and in fact it is infinitely smooth.

We will also use the intersection of $\mathcal{M}_r$ with smoothness spaces. As shown below, see (2.13), if $u \in \mathcal{M}_r$ belongs to $H_0^1(\Omega)$, then the factors u_k^1, u_k^2 in the SVD (2.1) all belong to $H_0^1(0,1)$. Likewise, a similar argument shows that if a corresponding tangent vector $v = \sum_{k=1}^r v_k^1 \otimes u_k^2 + u_k^1 \otimes v_k^2$, obeying the gauging conditions (2.4), belongs to $H_0^1(\Omega)$, then the v_k^1, v_k^2 are in $H_0^1(0,1)$ as well. Consequently, in this case the curve (2.3) yielding the tangent vector $v \in T_u\mathcal{M}_r \cap H_0^1(\Omega)$ satisfies

$$\phi(t) \in \mathcal{M}_r \cap H_0^1(\Omega) \tag{2.5}$$

for all t. The same condition will be assumed in the abstract setting as well.

A famous theorem due to Schmidt [41] states that truncating the SVD of u yields best approximations of lower rank in the $L_2(\Omega)$-norm. A particular instance of this result is that the smallest singular value $\sigma_{\min}(u) = \sigma_r(u)$ of $u \in \mathcal{M}_r$ equals the $L_2(\Omega)$-distance of u to the relative boundary $\mathcal{M}_{\leq r-1}$ of $\mathcal{M}_r$:

$$\sigma_{\min}(u) = \operatorname{dist}_{L_2(\Omega)}(u, \mathcal{M}_{\leq r-1}) = \operatorname{dist}_{L_2(\Omega)}(u, \overline{\mathcal{M}_r}^w \setminus \mathcal{M}_r). \tag{2.6}$$

The smallest singular value is also related to curvature bounds for the manifold, specifically to perturbations of tangent spaces. For $u \in \mathcal{M}_r$ we denote by P_u the L_2-orthogonal projection on $T_u\mathcal{M}_r$. It is given as

$$P_u = P_1 \otimes I + I \otimes P_2 - P_1 \otimes P_2 \tag{2.7}$$

where P_1 and P_2 denote the $L_2(\Omega)$-orthogonal projections onto the spans of $u_1^1, \dots, u_r^1$ and $u_1^2, \dots, u_r^2$, respectively. Then one can show the following: for any $\rho > 0$ there exist $M, \varepsilon > 0$ such that for all $u \in \mathcal{M}_r$ with $\sigma_{\min}(u) \geq \rho$ and all $v \in \mathcal{M}_r$ with $\|u - v\|_{L_2(\Omega)} \leq \varepsilon$ we have

$$\|P_u - P_v\|_{L_2(\Omega)\to L_2(\Omega)} \leq \frac{M}{\sigma_{\min}(u)} \|u - v\|_{L_2(\Omega)} \tag{2.8}$$

and

$$\|(I - P_v)(u - v)\|_{L_2(\Omega)\to L_2(\Omega)} \leq \frac{M}{\sigma_{\min}(u)} \|u - v\|_{L_2(\Omega)}^2. \tag{2.9}$$

This behavior of tangent spaces to low-rank manifolds is well known in finite dimension, even for more general tensor formats [1, 32]. In infinite-dimensional Hilbert spaces, a bound like (2.8) and (2.9) was obtained, for instance, for the (more general) Tucker format in [10]. For convenience, we provide a self-contained proof for (2.8) and (2.9) in the appendix (Lemma A.1 and Corollary A.2).

Regarding the estimate (2.8) and (2.9), we note that on every weakly compact subset $\mathcal{M}_r'$ of $\mathcal{M}_r$, the infimum

$$\sigma_* := \inf_{u\in\mathcal{M}_r'} \sigma_{\min}(u) = \inf_{u\in\mathcal{M}_r'} \operatorname{dist}_{L_2(\Omega)}(u, \mathcal{M}_{\leq r-1}) = \operatorname{dist}_{L_2(\Omega)}(\mathcal{M}_r', \overline{\mathcal{M}_r}^w \setminus \mathcal{M}_r)$$

is positive and attained by some $u_* \in \mathcal{M}_r'$. To see this, note first that for Banach spaces, by the Eberlein-Šmulian theorem, weak compactness is equivalent to weak sequential compactness. Consider sequences $(u_n) \subset \mathcal{M}_r'$ and $(v_n) \subset \mathcal{M}_{\leq r-1}$ such that

$$\|u_n - v_n\|_{L_2(\Omega)} \leq \sigma_* + 1/n.$$

Both sequences are bounded, and hence there exists a common weakly converging subsequence. Let u_* and v_* denote the limits. Then $u_* \in \mathcal{M}_r'$ and $v_* \in \mathcal{M}_{\leq r-1}$ since both sets are weakly sequentially closed. Since the norm is weakly sequentially lower semicontinuous, we obtain $\sigma_* \leq \|u_* - v_*\|_{L_2(\Omega)} \leq \sigma_*$, and thus equality. This shows

$$\sigma_* = \operatorname{dist}_{L_2(\Omega)}(u_*, \mathcal{M}_{\leq r-1}) > 0. \tag{2.10}$$

2.1.2. *Elliptic operators and low-rank manifolds.* Let us now discuss the interplay between the elliptic operator and the manifold in the model problem (1.7). Note that from the formulation (1.7), we will only have information on the bilinear form $a(\cdot,\cdot;t)$ on the tangent spaces $T_{u(t)}\mathcal{M}_r$. One can therefore expect that it will not be possible for arbitrary bilinear forms to derive the necessary a priori estimates vital for existence proofs. Obviously, additional structure is required.

In case of the model problem (1.7) we can split the bilinear form into two parts $a = a_1 + a_2$ with

$$a_1 = a_{11} + a_{22}, \quad a_2 = a_{12} + a_{21}.$$

These two parts are generated by the differential operators

$$A_1(t) = -\alpha_{11}(t)\,\partial_1^2 - \alpha_{22}(t)\,\partial_2^2, \quad A_2(t) = -\alpha_{12}(t)\,\partial_1\partial_2 - \alpha_{21}(t)\,\partial_2\partial_1, \tag{2.11}$$

corresponding to divergence and mixed derivatives at time t, respectively. The operator $A_1(t)$ has the remarkable property that it maps sufficiently smooth functions $u \in \mathcal{M}_r$ to the tangent space $T_u\mathcal{M}_r$. Namely, given the SVD representation (2.1), we get

$$(A_1(t)u)(x_1,x_2) = -\sum_{k=1}^r \sigma_k(u)\left(\alpha_{11}(t)\,\partial_1^2 u_k^1(x_1)\,u_k^2(x_2) + u_k^1(x_1)\,\alpha_{22}(t)\,\partial_2^2 u_k^2(x_2)\right), \tag{2.12}$$

which is in $T_u\mathcal{M}_r$ by (2.2) if the second derivatives $\partial_1^2 u_k^1$ and $\partial_2^2 u_k^2$ are in $L_2(0,1)$.

In order to translate this property to the generated bilinear forms $a_1(\cdot,\cdot;t)$, we observe that if $u \in \mathcal{M}_r \cap H_0^1(\Omega)$, then actually $u \in H^1_{\text{mix}}(\Omega) = H_0^1(0,1) \otimes H_0^1(0,1)$. That is, a low-rank function $u \in H_0^1(\Omega)$ automatically possesses mixed derivatives of order one, and all singular vectors u_k^i in the SVD (2.1) are themselves in $H_0^1(0,1)$. To see this, let u have the SVD (2.1), then, by orthogonality

$$u_k^1(x_1) = \frac{1}{\sigma_k}\int_0^1 u(x_1,x_2)\,u_k^2(x_2)\,\mathrm{d}x_2,$$

which gives

$$\|\partial_1 u_k^1\|_{L_2(0,1)} \leq \frac{1}{\sigma_k(u)}\|u\|_{H_0^1(\Omega)}. \tag{2.13}$$

Likewise, $\|\partial_2 u_k^2\|_{L_2(0,1)}$ admits precisely the same bound. Note that these bounds can be refined, since, e.g., in (2.13) only the derivative of u with respect to x_1 is needed, but this will not be required.

Now based on the regularity of the singular vectors one can show that if $u \in \mathcal{M}_r \cap H_0^1(\Omega)$, the tangent space projection P_u given in (2.7) can be bounded in H_0^1-norm as a map from $H_0^1(\Omega)$ to $T_u\mathcal{M}_r \cap H_0^1(\Omega)$ as follows:

$$\|P_u v\|_{H_0^1(\Omega)} \leq \left(1 + \frac{r}{\sigma_r(u)^2}\|u\|^2_{H_0^1(\Omega)}\right)^{1/2} \|v\|_{H_0^1(\Omega)}, \tag{2.14}$$

see Proposition A.3 in the appendix.

As a consequence, requiring only $u \in \mathcal{M}_r \cap H_0^1(\Omega)$, we can generalize the feature that the operator $A_1(t)$ maps to the tangent space to the following property of the induced bilinear form a_1: for every t,

$$a_1(u, v; t) = a_1(u, P_u v; t) \quad \text{for all } u \in \mathcal{M}_r \cap H_0^1(\Omega) \text{ and } v \in H_0^1(\Omega). \tag{2.15}$$

To see this, choose a sequence $(u_n) \subseteq \mathcal{M}_r \cap H^2(\Omega) \cap H_0^1(\Omega)$ converging to u in $H_0^1(\Omega)$-norm. Then for $v \in H_0^1(\Omega)$, 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 (2.12). 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 $L_2(\Omega)$ due to (2.8), and (2.14) implies $\limsup_n \|P_{u_n} v\|_{H_0^1} < \infty$. Since $L_2(\Omega)$ is dense in $H^{-1}(\Omega)$ it follows that $P_{u_n} v \to P_u v$ weakly in $H_0^1(\Omega)$ by a standard argument; see, e.g., [46, Prop. 21.23(g)]. Consequently, $a_1(u_n, P_{u_n} v; t) \to a_1(u, P_u v)$. At the same time, $a_1(u_n, v; t) \to a_1(u, v; t)$, so we have verified (2.15).

For the operator $A_2(t)$ on the other hand, the preceding considerations show that it actually is well defined on $\mathcal{M}_r \cap H_0^1(\Omega)$ in a strong sense: applying $\partial_1\partial_2$ to (2.1) and using the triangle inequality we get from (2.13) that

$$\|\partial_1\partial_2 u\|_{L_2(\Omega)} \leq \sum_{k=1}^{r} \frac{1}{\sigma_k(u)}\|u\|^2_{H_0^1(\Omega)} \leq \frac{r}{\sigma_{\min}(u)}\|u\|^2_{H_0^1(\Omega)}.$$

By (1.4), this implies that for every t, the bilinear form $a_2(\cdot, \cdot; t)$ associated to the operator $A_2(t)$ has the following property: for fixed $u \in \mathcal{M}_r \cap H_0^1(\Omega)$, the linear functional $v \mapsto a_2(u, v; t)$ on $H_0^1(\Omega)$ is actually continuous on $L_2(\Omega)$, with $L_2(\Omega)$ dual norm

$$\|A_2(t)u\|_{L_2(\Omega)} \leq \frac{2r\,|\alpha_{12}(t)|}{\sigma_{\min}(u)}\|u\|^2_{H_0^1(\Omega)}. \tag{2.16}$$

Note that here, the inverse of the smallest singular value of u enters again.

2.2. **Abstract formulation of the problem.** The features of the model problem discussed above are now formalized.

2.2.1. *Standard assumptions on parabolic evolution equations.* We consider a Gelfand triplet

$$V \subseteq H \subseteq V^*$$

where the real Hilbert space V is compactly embedded in the real Hilbert space H. Since the embedding is compact it is also continuous, that is,

$$\|u\|_H^2 \lesssim \|u\|_V^2 \quad \text{for all } u \in V. \tag{2.17}$$

In the case $H = L^2(\Omega)$ and $V = H_0^1(\Omega)$, (2.17) is the Poincaré inequality.

By $\langle\cdot,\cdot\rangle : V^* \times V \to \mathbb{R}$ we denote the dual pairing of V and V^*, and by $(\cdot,\cdot)$ we denote the inner product on H. For every $t \in [0,T]$, let $a(\cdot,\cdot;t) : V \times V \to \mathbb{R}$ be a bilinear form which is assumed to be symmetric,

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

uniformly bounded,

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

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

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

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

$$A(t) : V \to V^* \tag{2.18}$$

such that

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

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

$$|a(u,v;t) - a(u,v;s)| \leq L\beta \|u\|_V \|v\|_V |t-s| \tag{2.19}$$

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

2.2.2. *Manifolds and tangent spaces.* Our aim is to deal with evolution equations on a manifold

$$\mathcal{M} \subseteq H.$$

For present purposes, we do not have to be very strict regarding the notion of a manifold. What we essentially need is a tangent bundle: we assume that for every $u \in \mathcal{M}$ there exists a closed subspace $T_u\mathcal{M} \subset H$ given via a bounded H-orthogonal projection

$$P_u : H \to T_u\mathcal{M},$$

such that T_u contains all tangent vectors to $\mathcal{M}$ at u. Here a tangent vector is any $v \in H$ for which there exists a differentiable curve $\phi : (-\epsilon,\epsilon) \to H$ (for some $\epsilon > 0$) such that $\phi(t) \in \mathcal{M}$ for all t and

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

For our main existence result, we eventually assume that the map $u \mapsto P_u$ is locally Lipschitz continuous on $\mathcal{M}$ as a mapping on H.

It will be tacitly assumed that

- $\mathcal{M} \cap V$ is not empty,
- for every $u \in \mathcal{M} \cap V$, the space $T_u\mathcal{M} \cap V$ is not empty.

Indeed, in the main assumptions below we also require that $\mathcal{M}$ is a cone, as is the case for low-rank manifolds. Then the first property implies the second, because in this case $u \in T_u\mathcal{M}$ for every $u \in \mathcal{M}$.

2.2.3. *Problem formulation and main assumptions.* The abstract problem we are considering is now the following.

Problem 2.1. Given $f \in L_2(0,T;H)$ and $u_0 \in \mathcal{M} \cap V$, find

$$u \in W_2^1(0,T;V,H) = \{u \in L_2(0,T;V)\colon\ u' \in L_2(0,T;H)\}$$

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

$$\begin{aligned} u(t) &\in \mathcal{M}, \\ \langle u'(t), v\rangle + a(u(t), v; t) &= \langle f(t), v\rangle \quad \text{for all } v \in T_{u(t)}\mathcal{M} \cap V, \\ u(0) &= u_0. \end{aligned} \tag{2.20}$$

We emphasize again that the main challenge of this weak formulation is that according to the Dirac-Frenkel principle, the test functions are from the tangent space only. For showing that Problem 2.1 admits solutions we will require several assumptions. These assumptions are abstractions of corresponding properties of the model problem of a low rank manifold as discussed in Section 2.1, and hence the main results of this paper apply to this setting.

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 H, there exist constants $\kappa = \kappa(\mathcal{M}')$ and $\epsilon = \epsilon(\mathcal{M}')$ such that

$$\|P_u - P_v\|_{H\to H} \le \kappa \|u - v\|_H$$

and

$$\|(I - P_u)(u - v)\|_H \le \kappa \|u - v\|_H^2$$

for all $u, v \in \mathcal{M}'$ with $\|u - v\|_H \le \epsilon$.

A3 (Compatibility of tangent space)
(a) For $u \in \mathcal{M} \cap V$ and $v \in T_u\mathcal{M} \cap V$ an admissible curve with $\phi(0) = u$, $\phi'(0) = v$ can be chosen such that

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

for all $|t|$ small enough.
(b) If $u \in \mathcal{M} \cap V$ and $v \in V$ then $P_u v \in T_u\mathcal{M} \cap V$.

A4 (Operator splitting) The associated operator $A(t)$ in (2.18) admits a splitting

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

such that for all $t \in [0,T]$, all $u \in \mathcal{M} \cap V$ and all $v \in V$, 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 $\mathcal{M} \cap V$ to H": For every subset $\mathcal{M}'$ of $\mathcal{M}$ that is weakly compact in H, there exists $\gamma = \gamma(\mathcal{M}') > 0$ such that

$$A_2(t)u \in H \quad \text{and} \quad \|A_2(t)u\|_H \le \gamma \|u\|_V^\eta \quad \text{for all } u \in \mathcal{M}'$$

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

Recall that for the model problem, **A2** is stated in (2.8) and (2.9), taking (2.10) into account. Property **A3**(a) has been discussed in (2.5), and **A3**(b) in (2.14). With the splitting of A according to (2.11), in (2.15) we have shown that **A4**(a) holds, and **A4**(b) follows (with $\eta = 2$ independent of $\mathcal{M}'$) from (2.16), again using (2.10) and the boundedness of α.

Remark 2.2. It is well known that every function $u \in W_2^1(0,T;V,H)$ has a continuous representative $u \in C(0,T;H)$. Yet the notion of solution as defined in Problem 2.1 in principle does not require that $u(t) \in \mathcal{M}$ for all $t \in [0,T)$. Nonetheless, the existence and uniqueness statements for a maximal time interval (Theorems 4.3 and 4.2) will be derived, as expected, by extending continuous local solutions until they may hit the boundary of $\mathcal{M}$, such that $u(t) \in \mathcal{M}$ will be ensured for all t up to this point.

Remark 2.3. In Problem 2.1 we can actually weaken the uniform coercivity assumption to a uniform Gårding inequality

$$\langle A(t)u, u\rangle \leq \mu\|u\|_V^2 - \alpha\|u\|_H^2.$$

To see this suppose v is a solution (in the sense of Problem 2.1) of

$$\begin{aligned}\langle v'(t) + (A(t) + \alpha I)v(t), w\rangle &= \langle e^{-\alpha t} f(t), w\rangle \quad \text{for all } w \in T_{v(t)}\mathcal{M} \cap V,\\ v(0) &= u_0,\end{aligned}$$

which, given the Gårding inequality, has a uniformly coercive operator. Then $u(t) = e^{\alpha t}v(t)$ solves the equation

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

But since $\mathcal{M}$ is a cone, we have $T_{u(t)}\mathcal{M} \cap V = T_{v(t)}\mathcal{M} \cap V$, that is, u is indeed a solution of Problem 2.1 for the initial operator $A(t)$. For convenience we can therefore restrict ourselves to the coercive case.

3. Temporal discretization

Given the main assumptions **A1**–**A4** stated above, we prove existence of solutions for Problem 2.1 by discretizing in time and studying a sequence of approximate solutions with time steps $h \to 0$. A backward Euler method on $\mathcal{M}$ for (2.20) takes the following form: given $u_i \in \mathcal{M} \cap V$ at time step t_i, find $u_{i+1} \in \mathcal{M} \cap V$ at time step $t_{i+1} > t_i$ such that

$$\left(\frac{u_{i+1} - u_i}{t_{i+1} - t_i}, v\right) + a(u_{i+1}, v; t_{i+1}) = \langle f_{i+1}, v\rangle \quad \text{for all } v \in T_{u_{i+1}}\mathcal{M} \cap V. \tag{3.1}$$

Here f_{i+1} are the mean values of f on the interval $[t_i, t_{i+1}]$, that is,

$$f_{i+1} = \frac{1}{t_{i+1} - t_i}\int_{t_i}^{t_{i+1}} f(t)\,\mathrm{d}t. \tag{3.2}$$

As the test space depends on the solution, this equation appears quite difficult to solve. However, when $a(\cdot,\cdot;t_{i+1})$ is symmetric, (3.1) is the first order optimality condition of the optimization problem

$$u_{i+1} = \operatorname*{arg\,min}_{u \in \overline{\mathcal{M}}^w \cap V} F(u) = \frac{1}{2(t_{i+1} - t_i)}\|u - u_i\|_H^2 + \frac{1}{2}a(u, u; t_{i+1}) - \langle f_{i+1}, u\rangle. \tag{3.3}$$

This is stated in the following lemma.

Lemma 3.1. *Let $u_i \in \mathcal{M} \cap V$ and $h > 0$. Then any local minimum $u_{i+1} \in \mathcal{M} \cap V$ of F (defined in* (3.3)*) on $\overline{\mathcal{M}}^w \cap V$ satisfies the conditions* (3.1)*.*

Proof. Let $u_{i+1} \in \mathcal{M} \cap V$ be a local minimum and $v \in T_{u_{i+1}}\mathcal{M} \cap V$. By main assumption **A3**(a) we can find a differentiable curve $\phi(t)$ defined for $|t|$ small enough such that $\phi(0) = u_{i+1}$, $\phi'(0) = v$ and $\phi(t) \in \mathcal{M} \cap V$. Then $t \mapsto F(\phi_v(t))$ has a local minimum at $t = 0$ and hence it's derivative is zero there, which yields (3.1). □

Next, we consider the existence of minima of (3.3) on the set $\overline{\mathcal{M}}^w \cap V$. This asserts that we can generate approximate solutions $u_1, u_2, \dots \in \overline{\mathcal{M}}^w \cap V$ at a sequence of time steps using (3.3), which will serve as the temporal discretization. It will be later ensured that for a small enough time steps, we have $u_i \in \mathcal{M} \cap V$ if $u_0 \in \mathcal{M} \cap V$. Note that in any case the u_i are not uniquely determined from u_0, since in general $\overline{\mathcal{M}}^w \cap V$ is not convex.

Since the function F in (3.3) is convex on V and $\overline{\mathcal{M}}^w$ is weakly sequentially closed in H by definition, the existence of solutions to (3.3) is more or less standard.

Lemma 3.2. *The optimization problem* (3.3) *has at least one solution.*

Proof. Since F is convex and continuous on V it is also weakly sequentially lower semicontinuous on V; see, e.g., [47, Sec. 2.5, Lemma 5]. Note that F has bounded sublevel sets on V since the bilinear form $a(\cdot, \cdot; t_{i+1})$ is coercive by assumptions. It now follows that F attains a minimum on every weakly sequentially closed subset of V by the standard arguments, since the intersection with a sublevel set remains weakly sequentially compact; see, e.g. [45, Prop. 38.12(d)]. It hence remains to verify that $\overline{\mathcal{M}}^w \cap V$ is weakly sequentially closed in V. Consider a sequence $(u_n) \subset \overline{\mathcal{M}}^w \cap V$ converging weakly (in V) to $u \in V$. Obviously, since $H^* \subseteq V^*$, weak convergence in V implies weak convergence in H, and since $\overline{\mathcal{M}}^w$ is weakly sequentially closed in H, we get $u \in \overline{\mathcal{M}}^w \cap V$. This shows that this set is weakly sequentially closed in V. □

4. Existence of solutions

In the previous section we defined a time-stepping scheme through a sequence of optimization problems. Starting from $u_0 \in \mathcal{M} \cap V$ and setting

$$h = T/N, \quad t_i = ih,$$

this generates approximate solutions $u_1, \dots, u_N \in \overline{\mathcal{M}}^w \cap V$ at time points t_i. In this section we will study the properties of these solutions, and use them to prove existence of solutions to Problem 2.1. Specifically, construct a function $\hat{u}_h \colon [0,T] \to V$ by piecewise affine linear interpolation of u_i, and another function $\hat{v}_h \colon [0,T] \to V$ by piecewise constant interpolation of u_i such that $\hat{v}_h(0) = u_0$ and $\hat{v}_h(t) = u_i$ on $t \in (t_{i-1}, t_i]$.

Our main result is then as follows.

Theorem 4.1. *Let the assumptions stated in Section 2.2.3 hold.*

(a) *The functions $\hat{u}_h$ and $\hat{v}_h$ converge, up to subsequences, weakly in $L_2(0,T;V)$ and strongly in $L_2(0,T;H)$, to the same function $\hat{u} \in L_\infty(0,T;V)$ with $\hat{u}(0) = u_0$, while the weak derivatives $\hat{u}_h'$ converge weakly to $\hat{u}'$ in $L_2(0,T;H)$, again up to subsequences. We have $\hat{u}(t) \in \overline{\mathcal{M}}^w \cap V$ for almost all $t \in [0,T]$.*

(b) *Let $\sigma = \operatorname{dist}_H(u_0, \overline{\mathcal{M}}^w \setminus \mathcal{M}) > 0$. There exists a constant $c > 0$ independent of σ such that $\hat{u}$ solves* (2.20) *for almost all $t < (\sigma/c)^2$, where we set $\operatorname{dist}_H(u_0, \emptyset) = \infty$, and $\hat{u}(t) \in \mathcal{M}$ for all $t < (\sigma/c)^2$.*

Note that $\hat{u} \in L_\infty(0,T;V)$ with $\hat{u}' \in L_2(0,T;H)$ implies $\hat{u} \in W_2^1(0,T;V,H)$ in part (a). A possible constant c in statement (b) is provided by the right hand side of (4.2) in the energy estimates below, and thus in particular depends continuously on $\|u_0\|_V$ and $\|f\|_{L_2(0,T;H)}$. In the proof of the theorem, which will be given in Sections 4.1–4.3, we adapt standard techniques for establishing the existence of limits of time discretizations to the abstract manifold setup developed above.

By assuming some minimal integrability of the solution we can also obtain a uniqueness statement that applies to the local solution constructed in Theorem 4.1(b). The proof is given in Section 4.4.

Theorem 4.2. *Let the assumptions stated in Section 2.2.3 hold and let $u \in W_2^1(0,T^*;V,H)$ be a solution of Problem 2.1 on a time interval $[0,T^*]$. Assume that the continuous representative $u \in C(0,T^*;H)$ satisfies $u(t) \in \mathcal{M}$ for all $t \in [0,T^*)$. Moreover, assume that $u \in L_\eta(0,T^*;V)$ with $\eta > 0$ as in* **A4**(b)*. Then u is the only solution of Problem 2.1 in the space $W_2^1(0,T^*;V,H) \cap L_\eta(0,T^*;V)$.*

Note that $W_2^1(0,T^*;V,H) \subset L_\eta(0,T^*;V)$ when $\eta \leq 2$. Hence in this case, the additional assumption is void and we in fact obtain uniqueness in the space $W_2^1(0,T^*;V,H)$.

Combining Theorem 4.1 with a continuation argument and invoking Theorem 4.2, we obtain a unique solution on a maximal time interval.

Theorem 4.3. *Let the assumptions stated in Section 2.2.3 hold. There exist $T^* \in (0,T]$ and $u \in W_2^1(0,T^*;V,H) \cap L_\infty(0,T^*;V)$ such that u solves Problem 2.1 on the time interval $[0,T^*]$, and its continuous representative $u \in C(0,T^*;H)$ satisfies $u(t) \in \mathcal{M}$ for all $t \in [0,T^*)$. If $T^* \neq T$, then*

$$\liminf_{t \to T^*} \operatorname{dist}_H(u(t), \overline{\mathcal{M}}^w \setminus \mathcal{M}) = 0.$$

In either case, u is the unique solution of Problem 2.1 in $W_2^1(0,T^*;V,H) \cap L_\eta(0,T^*;V)$.

Recall that for the model problem (1.7), one has $\eta = 2$ by (2.16), and so this problem has a unique solution u in $W_2^1(0,T^*;H_0^1(\Omega),L_2(\Omega))$, and the $u(t)$ are of full rank r in the time interval $[0,T^*)$.

Proof of Theorem 4.3. The uniqueness of such a solution is immediate from Theorem 4.2. Hence we only need to show existence. Theorem 4.1(b) provides us with a solution u of Problem 2.1 on a time interval $[0,T_1]$ with $0 < T_1 \leq T$ such that $u \in L_\infty(0,T_1;V)$ and either $T_1 = T$ or $T_1 \geq \frac{1}{2}(\sigma_0/c)^2$ with $\sigma_0 = \operatorname{dist}_H(u_0, \overline{\mathcal{M}}^w \setminus \mathcal{M})$ and $c > 0$. In the latter case, we may assume without loss of generality that $u \in C(0,T_1;H)$ and $u(T_1) \in \mathcal{M} \cap V$. Let $\sigma_1 = \operatorname{dist}_H(u(T_1), \overline{\mathcal{M}}^w \setminus \mathcal{M})$. If $T_1 < T$, applying again Theorem 4.1 on $[T_1,T]$ with starting value $u_0 = u(T_1)$, we obtain a continuation of u to an interval $[0,T_2]$ with either $T_2 = T$ or $T_2 \geq T_1 + \frac{1}{2}(\sigma_1/c)^2$. In the latter case, we can again assume $u \in C(0,T_2;H)$ and $u(T_2) \in V$ with corresponding distance $\sigma_2 > 0$. We thus inductively obtain sequences $T_1, T_2, \dots$ and positive distances $\sigma_1, \sigma_2, \dots$ which either terminate with $T_i = T$ for some i, in which case we are done. Otherwise, T_i is defined for all i and $T_i \to T^* \leq T$. Clearly, the constructed $u \in C(0,T^*;H)$ solves (2.20) on $[0,T^*)$. If $\inf_i \sigma_i > 0$, then $T_{i+1} - T_i$ is bounded from below, which contradicts $T_i \leq T^*$. Thus $\liminf_{i\to\infty} \sigma_i = 0$, which implies the assertion. ☐

4.1. **Discrete energy estimates.** First we prove several a priori estimates of the time-discrete solution and its finite differences with respect to time, which are modifications of standard results for time stepping of parabolic PDEs; see, e.g., [18, 15, 9]. As can be seen from the proof, the assumed cone property **A1** of $\mathcal{M}$ is crucial.

Lemma 4.4. *The sequence* $(u_i)_{i=0}^N \subset \overline{\mathcal{M}}^w \cap V$ *generated by* (3.3) *with the time step* $h = T/N$ *satisfies the estimates*

$$\|u_N\|_H^2 + \sum_{i=1}^N \|u_i - u_{i-1}\|_H^2 + \mu h \sum_{i=1}^N \|u_i\|_V^2 \le \|u_0\|_H^2 + C_1 \|f\|_{L_2(0,T;H)}^2, \tag{4.1}$$

$$h \sum_{i=1}^N \left\| \frac{u_i - u_{i-1}}{h} \right\|_H^2 \le C_2 \Big(\|u_0\|_V^2 + \|f\|_{L_2(0,T;H)}^2 \Big), \tag{4.2}$$

$$\|u_i\|_V^2 \le C_3 \Big(\|u_0\|_V^2 + \|f\|_{L_2(0,T;H)}^2 \Big), \qquad i = 1, \dots, N, \tag{4.3}$$

where $C_1, C_2, C_3 > 0$ *depend on* β, μ, L, T, *and on the constant for the continuity of the embedding* $V \subseteq H$ *in* (2.17). *As a result,* $\hat{u}_h$ *and* $\hat{v}_h$ *are bounded in* $L_\infty(0,T;V)$, *uniformly for* $h \to 0$.

Proof. Since $\overline{\mathcal{M}}^w$ is a cone, and $u_{i+1} \in \overline{\mathcal{M}}^w \cap V$ minimizes F in (3.3), it follows directly that u_{i+1} satisfies the optimality condition (3.1) with $v = u_{i+1}$ (even when $u_{i+1} \in \overline{\mathcal{M}}^w \setminus \mathcal{M}$), that is, we have

$$(u_{i+1} - u_i, u_{i+1}) + h a(u_{i+1}, u_{i+1}; t_{i+1}) = h \langle f_{i+1}, u_{i+1} \rangle.$$

Using the identity

$$(u_{i+1} - u_i, u_{i+1}) = \frac{1}{2} \big(\|u_{i+1}\|^2 - \|u_i\|^2 + \|u_{i+1} - u_i\|^2 \big),$$

this reads

$$\|u_{i+1}\|_H^2 - \|u_i\|_H^2 + \|u_{i+1} - u_i\|_H^2 + 2h a(u_{i+1}, u_{i+1}; t_{i+1}) = 2h \langle f_{i+1}, u_{i+1} \rangle$$

The coercivity of a implies

$$\|u_{i+1}\|_H^2 - \|u_i\|_H^2 + \|u_{i+1} - u_i\|_H^2 + 2h\mu \|u_{i+1}\|_V^2 \le 2h \|f_{i+1}\|_{V^*} \|u_{i+1}\|_V,$$

which leads to

$$\|u_{i+1}\|_H^2 - \|u_i\|_H^2 + \|u_{i+1} - u_i\|_H^2 + h\mu \|u_{i+1}\|_V^2 \le \frac{h}{\mu} \|f_{i+1}\|_{V^*}^2,$$

where we have used the geometric mean inequality in the form $2xy \le \mu^{-1} x^2 + \mu y^2$. By summation over i we obtain

$$\|u_N\|_H^2 + \sum_{i=1}^N \|u_i - u_{i-1}\|_H^2 + h\mu \sum_{i=1}^N \|u_i\|_V^2 \le \|u_0\|_H^2 + \frac{h}{\mu} \sum_{i=1}^N \|f_i\|_{V^*}^2.$$

The embedding $V \subseteq H$ is continuous, cf. (2.17), which implies that also the embedding $H \cong H^* \subseteq V^*$ is continuous. Thus

$$\begin{aligned} h \sum_{i=1}^N \|f_i\|_{V^*}^2 \le Ch \sum_{i=1}^N \|f_i\|_H^2 \le Ch \sum_{i=1} \frac{1}{h^2} \left(\int_{t_{i-1}}^{t_i} 1 \mathrm{d}t \right) \left(\int_{t_{i-1}}^{t_i} \|f(t)\|_H^2 \, \mathrm{d}t \right) \\ \le C \|f\|_{L_2(0,T;H)}^2 \end{aligned} \tag{4.4}$$

with a constant $C > 0$ depending only on the one in (2.17), where we have used the Cauchy-Schwarz inequality and the definition (3.2) of f_i. This gives (4.1).

Next we show (4.2). Since u_{i+1} minimizes F,

$$\begin{aligned}2F(u_{i+1}) &= \frac{1}{h}\|u_{i+1}-u_i\|_H^2 + a(u_{i+1},u_{i+1};t_{i+1}) - 2\langle f_{i+1},u_{i+1}\rangle \\ &\le a(u_i,u_i;t_{i+1}) - 2\langle f_{i+1},u_i\rangle = 2F(u_i),\end{aligned}$$

which can be rearranged to

$$\begin{aligned}h\left\|\frac{u_{i+1}-u_i}{h}\right\|_H^2 &\le a(u_i,u_i;t_{i+1}) - a(u_{i+1},u_{i+1};t_{i+1}) + 2h\left\langle f_{i+1},\frac{u_{i+1}-u_i}{h}\right\rangle \\ &\le a(u_i,u_i;t_{i+1}) - a(u_{i+1},u_{i+1};t_{i+1}) + 2h\|f_{i+1}\|_H^2 + \frac{h}{2}\left\|\frac{u_{i+1}-u_i}{h}\right\|_H^2\end{aligned}$$

using a similar trick as above. This yields

$$h\left\|\frac{u_{i+1}-u_i}{h}\right\|_H^2 \le 2a(u_i,u_i;t_{i+1}) - 2a(u_{i+1},u_{i+1};t_{i+1}) + 4h\|f_{i+1}\|_H^2. \tag{4.5}$$

We sum over i, and get

$$\begin{aligned}h\sum_{i=1}^{N}\left\|\frac{u_i-u_{i-1}}{h}\right\|_H^2 \le 2a(u_0,u_0;0) &+ 2\sum_{i=1}^{N}\big(a(u_{i-1},u_{i-1};t_i) - a(u_{i-1},u_{i-1};t_{i-1})\big) \\ &- 2a(u_N,u_N;T) + 4h\sum_{i=1}^{N}\|f_i\|_H^2.\end{aligned}$$

Using the Lipschitz continuity (2.19) in t of the bilinear form then allows the estimates

$$\begin{aligned}h\sum_{i=1}^{N}\left\|\frac{u_i-u_{i-1}}{h}\right\|_H^2 &\le 2\beta\|u_0\|_V^2 + 2\beta Lh\sum_{i=1}^{N}\|u_{i-1}\|_V^2 + 4h\sum_{i=1}^{N}\|f_i\|_H^2 \\ &\le 2\beta(1+Lh)\|u_0\|_V^2 + 2\beta Lh\sum_{i=1}^{N}\|u_i\|_V^2 + 4h\sum_{i=1}^{N}\|f_i\|_H^2.\end{aligned}$$

By (4.1), which we already proved,

$$h\sum_{i=1}^{N}\left\|\frac{u_i-u_{i-1}}{h}\right\|_H^2 \le 2\beta(1+Lh)\|u_0\|_V^2 + 4h\sum_{i=1}^{N}\|f_i\|_H^2 + \frac{2\beta L}{\mu}\Big(\|u_0\|_H^2 + \frac{h}{\mu}\sum_{i=1}^{N}\|f_i\|_{V^*}^2\Big).$$

This allows us to simplify the above expression, and using (4.4) we recover (4.2).

Finally, we prove (4.3). Starting from (4.5), we readily obtain

$$0 \le a(u_{j-1},u_{j-1};t_j) - a(u_j,u_j;t_j) + 2h\|f_j\|_H^2.$$

We sum over $j = 1,\dots,i$ and rearrange:

$$a(u_i,u_i;t_i) \le a(u_0,u_0;0) + \sum_{j=1}^{i}\big(a(u_{j-1},u_{j-1};t_j) - a(u_{j-1},u_{j-1};t_{j-1})\big) + 2h\sum_{j=1}^{i}\|f_j\|_H^2,$$

This implies

$$\begin{aligned}\mu\|u_i\|_V^2 &\le \beta\|u_0\|_V^2 + \beta L h \sum_{j=1}^{i} \|u_{j-1}\|_V^2 + 2h\sum_{j=1}^{i}\|f_j\|_H^2 \\ &\le \beta(1+Lh)\|u_0\|_V^2 + \beta L h\sum_{j=1}^{N}\|u_j\|_V^2 + 2h\sum_{j=1}^{N}\|f_j\|_H^2\end{aligned}$$

for any $i = 1, \dots, N$. Using (4.1) and (4.4) yields (4.3). □

Remark 4.5. In standard estimates of the solution on the full linear space, the difference quotient (4.2) is typically bounded in $L_2(0,T;V^*)$ in terms of $\|u_0\|_H$ and $\|f\|_{L_2(0,T;V^*)}$, cf. [15]. One then uses the boundedness of $a(\cdot,\cdot;t)$ and f to get

$$\left\langle \frac{u_{i+1}-u_i}{h}, v\right\rangle = -a(u_{i+1}, v; t_{i+1}) + \langle f_{i+1}, v\rangle \le \beta\|u_{i+1}\|_V\|v\|_V + \|f_{i+1}\|_{V^*}\|v\|_V.$$

Dividing by $\|v\|_V$ and taking the supremum over $V\backslash\{0\}$ gives

$$\left\|\frac{u_{i+1}-u_i}{h}\right\|_{V^*} \le \beta\|u_{i+1}\|_V + \|f_{i+1}\|_{V^*},$$

and

$$h\sum_{i=1}^{N}\left\|\frac{u_i-u_{i-1}}{h}\right\|_{V^*} \lesssim \Big(\|u_0\|_H + h\sum_{i=1}^{N}\|f_i\|_{V^*}\Big).$$

However, we can not do this for solutions constrained to $\mathcal{M}$. Since the difference quotient is not necessarily in the tangent space, testing only with the tangent space does not give us the supremum and thus not the dual norm. We used a different reasoning in Lemma 4.4, and obtained a bound in the $L_2(0,T;H)$-norm in terms of $\|u_0\|_V$ and $\|f\|_{L_2(0,T;H)}$ instead.

4.2. **Proof of Theorem 4.1(a).** We now prove statement (a) of Theorem 4.1. The argument for showing the existence of the limiting function $\hat{u}$ relies on standard compactness arguments based on the energy estimates in Lemma 4.4. Showing that $\hat{u}(t) \in \overline{\mathcal{M}}^w$ for almost all t is then based on the fact that this set is weakly sequentially closed by definition.

It follows from Lemma 4.4 that $\hat{u}_h$ and $\hat{v}_h$ are bounded in $L_2(0,T;V)$, uniformly with respect to h. Therefore, refinement in time generates sequences in $L_2(0,T;V)$ which, up to subsequences, converge weakly,

$$\hat{u}_h \rightharpoonup \hat{u} \quad \text{and} \quad \hat{v}_h \rightharpoonup \hat{v} \quad \text{in } L_2(0,T;V).$$

In particular, $\hat{u}_h - \hat{v}_h$ converges weakly in $L_2(0,T;H)$ to $\hat{u} - \hat{v}$. Comparing the two sequences in $L_2(0,T;H)$, we get

$$\begin{aligned}\int_0^T \|\hat{u}_h - \hat{v}_h\|_H^2 \,\mathrm{d}t &= \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \|\hat{u}_h - \hat{v}_h\|_H^2 \,\mathrm{d}t \\ &= \sum_{i=1}^N \int_{t_{i-1}}^{t_i} \left\| \left(\frac{t_i - t}{h}\right) u_{i-1} + \left(\frac{t - t_{i-1}}{h}\right) u_i - u_i \right\|_H^2 \mathrm{d}t \\ &= h \sum_{i=1}^N \int_0^1 \|(s-1)(u_i - u_{i-1})\|_H^2 \,\mathrm{d}s \\ &= \frac{h}{3} \sum_{i=1}^N \|u_i - u_{i-1}\|_H^2,\end{aligned}$$

and by Lemma 4.4,

$$\int_0^T \|\hat{u}_h - \hat{v}_h\|_H^2 \,\mathrm{d}t \leq \frac{C_2 h^2}{3} \Big(\|u_0\|_V^2 + \|f\|_{L_2(0,T;H)}^2 \Big), \tag{4.6}$$

which tends to zero as $h \to 0$. We conclude $\hat{u} = \hat{v}$.

Likewise, $\hat{u}_h'$ is uniformly bounded in $L_2(0,T;H)$, and thus up to subsequences, $\hat{u}_h' \rightharpoonup \hat{w}$ for some $\hat{w} \in L_2(0,T;H)$. We want to show that $\hat{w}$ is the weak derivative of $\hat{u}$. For this, we need to show that

$$\int_0^T (\hat{w}(t), v)\, \phi(t) \,\mathrm{d}t + \int_0^T (\hat{u}(t), v)\, \phi'(t) \,\mathrm{d}t = 0$$

for arbitrary $v \in V$ and $\phi \in \mathcal{C}_0^\infty(0,T)$. Adding and subtracting the weak derivative of $\hat{u}_h$, we get

$$\begin{aligned}\int_0^T (\hat{w}(t), v)\, \phi(t) \,\mathrm{d}t &+ \int_0^T (\hat{u}(t), v)\, \phi'(t) \,\mathrm{d}t = \\ &= \int_0^T (\hat{w}(t) - \hat{u}_h'(t), v)\, \phi(t) \,\mathrm{d}t + \int_0^T (\hat{u}(t) - \hat{u}_h(t), v)\, \phi'(t) \,\mathrm{d}t.\end{aligned}$$

Since $\hat{u}_h \rightharpoonup \hat{u}$ and $\hat{u}_h' \rightharpoonup \hat{w}$ in $L_2(0,T;V)$ and $L_2(0,T;H)$, respectively, and since $v\phi$, $v\phi' \in L_2(0,T;V)$, the right hand side converges to zero. Thus, $\hat{w} = \hat{u}'$.

The strong convergence in $L_2(0,T;H)$ of $\hat{u}_h$ follows from the theorem of Lions and Aubin [43, Prop. III.1.3]. It says that when V is compactly embedded in H, then the space $W_2^1(0,T;V,H)$ is compactly embedded in $L_2(0,T;H)$. Thereby, the weak convergence of $\hat{u}_h$ and $\hat{u}_h'$ we just have proven imply the strong convergence $\hat{u}_h \to \hat{u}$ in $L_2(0,T;H)$. This together with (4.6) directly proves that also $\hat{v}_h \to \hat{u}$ in $L_2(0,T;H)$. By (4.3) and lower semicontinuity of the $L_\infty(0,T;V)$-norm with respect to weak convergence in $L_2(0,T;V)$, we even obtain $\hat{u} \in L_\infty(0,T;V)$.

It remains to prove that $\hat{u}(t) \in \overline{\mathcal{M}}^w$ for almost all $t \in [0,T]$. Recall that $h = T/N$, and let $t_i^{(N)} = ih$, $i = 0, \dots, N$. To any fixed $t \in [0,T]$, we associate a sequence $(t_{j_N}^{(N)})_{N=N_0}^\infty \subset [0,T]$ such that $t_{j_N}^{(N)} \to t$ as $N \to \infty$. We construct the sequence such that $t_{j_N}^{(N)}$ is the largest possible $t_i^{(N)} \leq t$, which implies $0 \leq t - t_{j_N}^{(N)} \leq h$. If we can now show that for almost all $t \in [0,T]$ a subsequence of $(\hat{u}_h(t_{j_N}^{(N)})) \subseteq \overline{\mathcal{M}}^w$ converges weakly in H to $\hat{u}(t)$, we get that $\hat{u}(t) \in \overline{\mathcal{M}}^w$ for such t. We will even show that there exists a strongly convergent subsequence based on the inequality

$$\big\|\hat{u}_h(t_{j_N}^{(N)}) - \hat{u}(t)\big\|_H \le \big\|\hat{u}_h(t_{j_N}^{(N)}) - \hat{u}_h(t)\big\|_H + \big\|\hat{u}_h(t) - \hat{u}(t)\big\|_H.$$

Regarding the second term on the right hand side, since $\hat{u}_h \to \hat{u}$ in $L_2(0,T;H)$, and possibly passing to a subsequence, we have $\hat{u}_h(t) \to \hat{u}(t)$ in H for almost all t. In order to show that the first term of the right hand side vanishes in the limit we recall that by construction, $\hat{u}_h$ is linear on the interval $[t_{j_N}^{(N)}, t_{j_N+1}^{(N)}]$ that contains the given t. Therefore, using (4.2),

$$\big\|\hat{u}_h(t_{j_N}^{(N)}) - \hat{u}_h(t)\big\|_H \le \big\|\hat{u}_h(t_{j_N}^{(N)}) - \hat{u}_h(t_{j_N+1}^{(N)})\big\|_H \le \sqrt{C_2 h\Big(\|u_0\|_V^2 + \|f\|_{L_2(0,T;H)}^2\Big)},$$

which vanishes in the limit. This shows $\hat{u}_h(t_{j_N}^{(N)}) \to \hat{u}(t)$ strongly in H for almost all t.

Finally, we show that $\hat{u}(0) = u_0$. By construction, $\hat{u}_h(0) = u_0$. We choose $v \in C^\infty(0,T;V)$ such that $v(T) = 0$ and apply integration by parts,

$$\int_0^T \langle \hat{u}_h'(t), v(t)\rangle \,\mathrm{d}t + \int_0^T \langle \hat{u}_h(t), v'(t)\rangle \,\mathrm{d}t = -(\hat{u}_h(0), v(0)) = -(u_0, v(0)).$$

In the limit $h \to 0$,

$$\begin{aligned} -(u_0, v(0)) = \int_0^T \langle \hat{u}_h'(t), v(t)\rangle \,\mathrm{d}t + \int_0^T \langle \hat{u}_h(t), v'(t)\rangle \,\mathrm{d}t \to \\ \to \int_0^T \langle \hat{u}'(t), v(t)\rangle \,\mathrm{d}t + \int_0^T \langle \hat{u}(t), v'(t)\rangle \,\mathrm{d}t = -(\hat{u}(0), v(0)), \end{aligned}$$

and as $(u_0, v(0))$ is independent of h, $(u_0 - \hat{u}(0), v(0)) = 0$ for all $v(0) \in V$.

This concludes the proof of Theorem 4.1(a). □

4.3. **Proof of Theorem 4.1(b).** Our goal is to show that there exists a constant $c > 0$ such that for all $\theta \in (0,1)$ and

$$T_\theta = \min\left\{\left(\frac{\theta\sigma}{c}\right)^2, T\right\},$$

the limiting function $\hat{u}(t)$ solves Problem 2.1 for almost all $t \in [0, T_\theta]$. Since Problem 2.1 is formulated on $\mathcal{M} \cap V$, we first in particular need to ensure that $\hat{u}(t) \in \mathcal{M}$ for almost all t. We do this by showing next that the $\hat{v}_h(t)$ keep a positive distance in H-norm to $\overline{\mathcal{M}}^w \setminus \mathcal{M}$. For fixed $h = T/N$ the estimate (4.2) in Lemma 4.4 yields for every integer $j \le N$ that

$$\begin{aligned} \|u_j - u_0\|_H \le \sum_{i=1}^j \|u_i - u_{i-1}\|_H &\le \sqrt{j}\left(\sum_{i=1}^j \|u_i - u_{i-1}\|_H^2\right)^{1/2} \\ &= \sqrt{t_j}\left(\sum_{i=1}^j \frac{\|u_i - u_{i-1}\|_H^2}{h}\right)^{1/2} \\ &\le \sqrt{t_j}\, c, \end{aligned}$$

where c is the right hand side of (4.2). Using this c in the definition of T_θ, we have ensured

$$\|u_j - u_0\|_H \le \theta\sigma \quad \text{for all } t_j \le T_\theta \le \frac{\theta^2\sigma^2}{c^2}.$$

Hence, by construction, since $\hat{v}_h$ is the piecewise constant interpolant,

$$\|\hat{v}_h(t) - u_0\|_H \leq \theta\sigma \quad \text{for all } t \leq T_\theta.$$

Since the distance of $u_0 \in \mathcal{M}$ to $\overline{\mathcal{M}}^w \setminus \mathcal{M}$ is larger than σ, this shows that

$$\hat{v}_h(t) \in \mathcal{M}' := \{u \in \mathcal{M} \colon \|u - u_0\|_H \leq \theta\sigma\} \quad \text{for all } t \leq T_\theta. \tag{4.7}$$

The set $\mathcal{M}'$ is indeed a weakly compact subset of $\mathcal{M}$, since it is bounded and every limit point of a weakly converging sequence belongs to $\{u \in \overline{\mathcal{M}}^w \colon \|u - u_0\|_H \leq \theta\sigma\}$, which again implies $u \in \mathcal{M}$, that is, $u \in \mathcal{M}'$. Since, up to subsequences, $\hat{v}_h(t) \to \hat{u}(t)$ strongly in H for almost all t (by part (a)), we get $\hat{u}(t) \in \mathcal{M}' \subset \mathcal{M}$ for almost all $t \in [0, T_\theta]$. Since $\hat{u} \in C(0,T;H) \subset W_2^1(0,T;V,H)$, we obtain

$$\hat{u}(t) \in \mathcal{M} \quad \text{for all } t \in [0, T_\theta].$$

Note next that (4.7) holds independently of h. The main assumptions **A2** and **A3** provide us now with positive constants η, γ, and κ such that

$$\|A_2(t)\hat{v}_h(t)\|_H \leq \gamma \|\hat{v}_h(t)\|_V^\eta$$

and

$$\|P_{\hat{u}(t)} - P_{\hat{v}_h(t)}\|_{H\to H} \leq \kappa \|\hat{u}(t) - \hat{v}_h(t)\|_H \tag{4.8}$$

whenever $\|\hat{u}(t) - \hat{v}_h(t)\|_H \leq \epsilon(\mathcal{M}')$, for all h and almost all $t \in [0, T_\theta]$. These are the crucial estimates in order to show that $\hat{u}(t)$ solves Problem 2.1 for all such t in the remainder of this proof.

Using the interpolant

$$F_h(t) = f_i, \quad t_{i-1} < t \leq t_i, \ i = 1, \dots, N,$$

the Galerkin-type condition (3.1) can be written as

$$\langle \hat{u}_h'(t), v\rangle + a_h(\hat{v}_h(t), v; t) = \langle F_h(t), v\rangle \quad \text{for all } v \in T_{\hat{v}_h(t)}\mathcal{M} \cap V, \tag{4.9}$$

where $a_h(\cdot,\cdot)$ is the piecewise constant in time interpolant of $a(\cdot,\cdot;t)$. By Lemma 3.1, this holds as long as $\hat{v}_h(t) \in \mathcal{M}$, which by our above considerations is ensured for all $t \in [0, T_\theta]$. For these t, we define the spaces

$$\mathcal{V}(t) = T_{\hat{u}(t)}\mathcal{M} \cap V, \quad \mathcal{V}_h(t) = T_{\hat{v}_h(t)}\mathcal{M} \cap V.$$

We need to show that

$$\langle \mathcal{L}(t;\hat{u}), v\rangle = \langle \hat{u}'(t), v\rangle + a(\hat{u}(t), v; t) - \langle f(t), v\rangle = 0, \quad v \in \mathcal{V}(t), \tag{4.10}$$

for almost all $t \in [0, T_\theta]$, and also that $\hat{u}(0) = u_0$. Consider the related expression

$$\langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v\rangle = \langle \hat{u}_h'(t), v\rangle + a_h(\hat{v}_h(t), v; t) - \langle F_h(t), v\rangle \tag{4.11}$$

for an arbitrary $v \in \mathcal{V}(t)$. As v is in the tangent space at $\hat{u}(t)$, and not at $\hat{v}_h(t)$, this expression in general does not equal zero exactly. With test functions in the correct tangent space, however, we do recover (3.1), that is, we have

$$\langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v_h\rangle = 0, \quad v_h \in \mathcal{V}_h(t). \tag{4.12}$$

Our first goal is to show, term by term, that for any $w \in L_2(0, T_\theta; V)$ satisfying $w(t) \in \mathcal{V}(t)$ for almost all t, we have

$$\int_0^{T_\theta} (\mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), w(t))\,\mathrm{d}t \to \int_0^{T_\theta} (\mathcal{L}(t;\hat{u}), w(t))\,\mathrm{d}t\,. \tag{4.13}$$

For the first term in the right hand side of (4.11), we immediately obtain

$$\int_0^{T_\theta} \langle \hat{u}_h'(t), w(t)\rangle \,\mathrm{d}t \to \int_0^{T_\theta} \langle \hat{u}'(t), w(t)\rangle \,\mathrm{d}t$$

as $h \to 0$. Regarding the second term, the bilinear form $a_h(\cdot,\cdot;t)$ defines an operator $A_h(t) : V \to V^*$,

$$a_h(\hat{v}_h(t), w(t); t) = \langle A_h(t)\hat{v}_h(t), w(t)\rangle,$$

and since $a_h(\cdot,\cdot;t)$ is symmetric,

$$a_h(\hat{v}_h(t), w(t); t) = a_h(w(t), \hat{v}_h(t); t) = \langle A_h(t)w(t), \hat{v}_h(t)\rangle.$$

We then get

$$\begin{aligned}\int_0^{T_\theta} a_h(\hat{v}_h(t), w(t); t)\,\mathrm{d}t &= \int_0^{T_\theta} \langle A_h(t)w(t), \hat{v}_h(t)\rangle \,\mathrm{d}t = \\ &= \int_0^{T_\theta} \langle A(t)w(t), \hat{v}_h(t)\rangle \,\mathrm{d}t + \int_0^{T_\theta} \langle (A_h(t) - A(t))w(t), \hat{v}_h(t)\rangle \,\mathrm{d}t. \end{aligned} \tag{4.14}$$

We have $Aw, A_h w \in L_2(0,T;V^*)$, and hence

$$\int_0^{T_\theta} \langle A(t)w(t), \hat{v}_h(t)\rangle \,\mathrm{d}t \to \int_0^{T_\theta} \langle A(t)w(t), \hat{u}(t)\rangle \,\mathrm{d}t = \int_0^{T_\theta} a(\hat{u}(t), w(t); t)\,\mathrm{d}t$$

as $h \to 0$. The second integral in (4.14) vanishes in the limit, since

$$\begin{aligned}\Big|\int_0^{T_\theta} \langle (A_h(t) - A(t))w(t), \hat{v}_h(t)\rangle \,\mathrm{d}t\Big| &\le \int_0^{T_\theta} |\langle (A_h(t) - A(t))w(t), \hat{v}_h(t)\rangle| \,\mathrm{d}t \le \\ \le \int_0^{T_\theta} hL\beta \|w(t)\|_V \|\hat{v}_h(t)\|_V \,\mathrm{d}t &\le hL\beta \|w\|_{L_2(0,T_\theta;V)} \|\hat{v}_h\|_{L_2(0,T_\theta;V)} \to 0.\end{aligned}$$

We thus have shown

$$\int_0^{T_\theta} a_h(\hat{v}_h(t), w(t); t)\,\mathrm{d}t \to \int_0^{T_\theta} a(\hat{u}(t), w(t); t)\,\mathrm{d}t.$$

Finally, as F_h is a piecewise constant interpolant of a given function $f \in L_2(0, T_\theta; H)$, we have $F_h \to f$ strongly in $L_2(0, T_\theta; H)$, and altogether we obtain (4.13).

We next show that for all w as above,

$$\int_0^{T_\theta} \langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), w(t)\rangle \,\mathrm{d}t \to 0. \tag{4.15}$$

By (4.12), for $v \in \mathcal{V}(t)$ and $v_h \in \mathcal{V}_h(t)$, it holds that

$$\langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v\rangle = \langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v - v_h\rangle.$$

We choose $v_h = P_{\hat{v}_h(t)}v$. Note that $v = P_{\hat{u}(t)}v$ for $v \in \mathcal{V}(t)$ for almost all $t \in [0, T_\theta]$. Thus

$$\begin{aligned}\langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v\rangle &= \langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), (P_{\hat{u}(t)} - P_{\hat{v}_h(t)})v\rangle \\ &= \langle \hat{u}_h'(t) + A_h(t)\hat{v}_h(t) - F_h(t), (P_{\hat{u}(t)} - P_{\hat{v}_h(t)})v\rangle \\ &= \langle \hat{u}_h'(t) + A_h(t)\hat{v}_h(t) - F_h(t), (I - P_{\hat{v}_h(t)})(P_{\hat{u}(t)} - P_{\hat{v}_h(t)})v\rangle,\end{aligned}$$

where the last equality holds due to **A3**(b) and (4.9). In light of **A4** we hence have the estimate

$$|\langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v\rangle| \le \big(\|\hat{u}_h'(t)\|_H + \gamma \|\hat{v}_h(t)\|_V^\eta + \|F_h(t)\|_H\big) \|(P_{\hat{u}(t)} - P_{\hat{v}_h(t)})v\|_H$$

for almost all $t \in [0, T_\theta]$. By the curvature bound (4.8),

$$\|(P_{\hat{u}(t)} - P_{\hat{v}_h(t)})v\|_H \le D_{t,h}\|v\|_H, \quad D_{t,h} := \kappa\|\hat{u}(t) - \hat{v}_h(t)\|_H, \tag{4.16}$$

once $\|\hat{u}(t) - \hat{v}_h(t)\|_H \le \epsilon(\mathcal{M}')$. Since $\hat{v}_h$ converges strongly in $L_2(0, T_\theta; H)$, up to passing to a subsequence we have $\|\hat{u}(t) - \hat{v}_h(t)\|_H \to 0$ for almost all $t \in [0, T_\theta]$. Hence for almost all t, (4.16) applies for sufficiently small h. We can thus conclude that

$$\langle \mathcal{L}_h(t; \hat{u}_h, \hat{v}_h), v\rangle \to 0 \quad \text{for almost all } t \in [0, T_\theta]$$

when $v \in \mathcal{V}(t)$. Specifically taking $v = w(t)$ in the above considerations shows

$$|\langle \mathcal{L}(t; \hat{u}_h, \hat{v}_h), w(t)\rangle| \le D_{t,h}(\|\hat{u}_h'(t)\|_H + \gamma\|\hat{v}_h(t)\|_V^\eta + \|F_h(t)\|_H)\|w(t)\|_H,$$

where $D_{t,h}$ and $\|\hat{v}_h(t)\|_V$ are bounded uniformly in h for almost all $t \in [0, T_\theta]$. Hence the right hand side provides an integrable upper bound, and by the dominated convergence theorem we arrive at (4.15).

Combined with (4.13), we conclude

$$\int_0^{T_\theta} \langle \mathcal{L}(t; \hat{u}), w(t)\rangle \,\mathrm{d}t = 0 \quad \text{for all } w \in L_2(0, T_\theta; V),\ w(t) \in \mathcal{V}(t). \tag{4.17}$$

This shows (4.10) as desired, since in the opposite case there would be a subset $S \subseteq [0, T_\theta]$ of positive measure such that for all $t \in S$ we have $\langle \mathcal{L}(t; \hat{u}), v\rangle \ne 0$ for some $v \in \mathcal{V}(t)$. By appropriately scaling these v, we can then choose $w(t) \in \mathcal{V}(t)$ such that $\|w\|_{L^\infty(0,T_\theta;V)} < \infty$ and $\langle \mathcal{L}(t; \hat{u}), w(t)\rangle > 0$ (since $\mathcal{V}(t)$ is a linear space). Hence the left hand side of (4.17) would be positive.

This concludes the proof of Theorem 4.1(b). $\square$

4.4. **Proof of Theorem 4.2.** Let v be another solution of Problem 2.1 in the space $W_2^1(0, T^*; V, H) \cap L_\eta(0, T^*; V)$. For both u and v we take the continuous representative in $C(0, T^*; H)$. Then there exists $0 < t^* \le T^*$ such that for all $t \in [0, t^*]$ both $u(t)$ and $v(t)$ lie in a weakly compact set $\mathcal{M}' \subset \mathcal{M}$ and satisfy $\|u(t) - v(t)\|_H \le \epsilon(\mathcal{M}')$ as in assumption **A2**. For almost all $t \in [0, t^*]$,

$$\begin{aligned}
\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t) - v(t)\|_H^2 &\le \langle u'(t) - v'(t) + A(t)(u(t) - v(t), u(t) - v(t)\rangle \\
&= \langle u'(t) + A(t)u(t) - f(t), u(t) - v(t)\rangle - \\
&\qquad - \langle v'(t) + A(t)v(t) - f(t), u(t) - v(t)\rangle. \\
&= \langle u'(t) + A(t)u(t) - f(t), (I - P_{u(t)})(u(t) - v(t))\rangle - \\
&\qquad - \langle v'(t) + A(t)v(t) - f(t), (I - P_{v(t)})(u(t) - v(t))\rangle \\
&\le \big(\|u'(t)\|_H + \gamma\|u(t)\|_V^\eta + \|f(t)\|_H\big)\, \|(I - P_{u(t)})(u(t) - v(t))\|_H + \\
&\qquad + \big(\|v'(t)\|_H + \gamma\|v(t)\|_V^\eta + \|f(t)\|_H\big)\, \|(I - P_{v(t)})(u(t) - v(t))\|_H,
\end{aligned}$$

where we have made use of the coercivity of $A(t)$, (2.20) and assumption **A4**(b). With **A2** this further yields

$$\begin{aligned}
&\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}t}\|u(t) - v(t)\|_H^2 \le \\
&\qquad \le \kappa\big(\|u'(t)\|_H + \|v'(t)\|_H + \gamma\|u(t)\|_V^\eta + \gamma\|v(t)\|_V^\eta + 2\|f(t)\|_H\big)\|u(t) - v(t)\|_H^2
\end{aligned}$$

for almost all $t \in [0, t^*]$. This upper bound is positive and integrable by assumption. Hence Gronwall's Lemma is applicable and yields $\|u(t) - v(t)\|_H = 0$ for almost all $t \in [0, t^*]$; by continuity, $u(t) = v(t)$ for all $t \in [0, t^*]$.

Now let$[0,T')$ be a largest interval of the given form on which u and v agree. If $T' < T^*$, then by continuity $v(T') = u(T')$, and by assumption $u(T') \in \mathcal{M}$. The above local uniqueness argument can then be repeated from T' on, which yields a contradiction. Hence we must have $T' = T^*$. □

5. Numerical Methods for Low-Rank Matrix Manifolds

In this section, we comment on how the basic variational time stepping scheme (3.3), which we have used to prove the existence of a solution to Problem 2.1, is connected to numerical methods for actually computing the low-rank evolution. A first strategy for solving the general dynamical low-rank problem (1.1), as used in [25], is to extract from (1.1) equations for the components U, S, V in a factorization $Y(t) = U(t)\,S(t)\,V(t)^T$, which can then be solved by standard time stepping schemes.

An alternative scheme was proposed in [30]. For notational convenience, the SVD representation of a low-rank matrix Y can be written in vectorized form as

$$y := \mathrm{vec}(Y) = \mathrm{vec}(USV^T) = (V \otimes U)s, \quad s = \mathrm{vec}(S).$$

With $P_V := VV^T$, (1.1) is then rewritten as

$$y'(t) = (I \otimes P_{U(t)})F\big(t, y(t)\big) - (P_{V(t)} \otimes P_{U(t)})F\big(t, y(t)\big) + (P_{V(t)} \otimes I)F\big(t, y(t)\big), \tag{5.1}$$

based on the formula (2.7) for the tangent space projector. A time stepping scheme is then obtained by applying an operator splitting, that is, by integrating the three terms on the right hand side of (5.1) in time in the given order. As shown in [30], the resulting method has very interesting characteristics; for instance, the splitting is exact if $F\big(t, Y(t)\big)$ is in the tangent space at $Y(t)$ on the considered time interval.

Let $\Phi := (\varphi_n)_{n\in\mathcal{I}}$ with $\mathcal{I} \subseteq \mathbb{N}$ be an orthonormal system in $L_2(0,1)$ with all φ_n sufficiently regular, and let

$$\tilde{A}(t) := \Big(\big\langle A(t)(\varphi_{j_1} \otimes \varphi_{j_2}), (\varphi_{i_1} \otimes \varphi_{i_2})\big\rangle\Big)_{i,j\in\mathcal{I}^2}, \quad \tilde{f}(t) := \Big(\big\langle f(t), (\varphi_{i_1} \otimes \varphi_{i_2})\big\rangle\Big)_{i\in\mathcal{I}^2},$$

so that the initial value problem

$$\tilde{u}'(t) + \tilde{A}(t)\tilde{u}(t) = \tilde{f}(t)$$

for $\tilde{u}(t) \in \ell_2(\mathcal{I})$ is a Galerkin semidiscretization or, if Φ is an orthonormal basis and $\mathcal{I} = \mathbb{N}$, the basis representation of (1.3). The splitting scheme from [30] for this problem, for a time step of length h with low-rank initial data $u_0 = (V_0 \otimes U_0)s_0$, formally reads as follows:

– Determine $U_1 = U(h)$, $s_1^+ = s(h)$ as solutions of

$$\frac{\mathrm{d}}{\mathrm{d}t}(V_0 \otimes U)s = -(V_0V_0^T \otimes I)\tilde{A}(V_0 \otimes U)s + (V_0V_0^T \otimes I)\tilde{f}, \quad U(0) = U_0,\ s(0) = s_0. \tag{5.2a}$$

– Determine $s_0^+ = s(h)$ as solution of

$$\frac{\mathrm{d}}{\mathrm{d}t}(V_0 \otimes U_1)s = (V_0V_0^T \otimes U_1U_1^T)\tilde{A}(V_0 \otimes U_1)s - (V_0V_0^T \otimes U_1U_1^T)\tilde{f}, \quad s(0) = s_1^+. \tag{5.2b}$$

– Determine $V_1 = V(h)$, $s_1 = s(h)$ as solutions of

$$\frac{\mathrm{d}}{\mathrm{d}t}(V \otimes U_1)s = -(I \otimes U_1U_1^T)\tilde{A}(V \otimes U_1)s + (I \otimes U_1U_1^T)\tilde{f}, \quad V(0) = V_0,\ s(0) = s_0^+ \tag{5.2c}$$

Altogether, this yields $u_1 = (V_1 \otimes U_1)s_1$. Note that while (5.2a) and (5.2c) are parabolic problems projected to a subspace, (5.2b) is a *backwards* parabolic problem (projected to a finite-dimensional space) that can in principle be extremely ill-conditioned.

As we show next, a suitable adaptation of (5.2) can mitigate this issue. This adaptation turns out to be closely related to the Alternating Least Squares (ALS) low-rank minimization method applied to (3.3). The ALS method for (3.3), with $\tilde{A}_i = \tilde{A}(t_i)$ and $\tilde{f}_i = \tilde{f}(t_i)$, consists in the following iteration.

Given $y_0 = (V_0 \otimes U_0)s_0$, repeat for $i = 0, 1, 2, \dots$:

– Solve

$$(I + h(V_i^T \otimes I)\tilde{A}_{i+1}(V_i \otimes I))k_{i+1} = (I \otimes U_i)s_i + h(V_i^T \otimes I)\tilde{f}_{i+1} \tag{5.3a}$$

for k_{i+1}.

– Factorize

$$k_{i+1} = (I \otimes U_{i+1})s_{i+1}^+ \tag{5.3b}$$

– Solve

$$(I + h(I \otimes U_{i+1}^T)\tilde{A}_{i+1}(I \otimes U_{i+1}))\ell_{i+1} = (V_i \otimes U_{i+1}^T U_i)s_i + h(I \otimes U_{i+1}^T)\tilde{f}_{i+1} \tag{5.3c}$$

for ℓ_{i+1}.

– Factorize

$$(V_{i+1} \otimes I)s_{i+1} \tag{5.3d}$$

to obtain $y_{i+1} = (V_{i+1} \otimes U_{i+1})s_{i+1}$.

Theorem 5.1. *The approximation of the splitting scheme* (5.2) *provided by solving* (5.2a) *and* (5.2c) *with backward Euler and* (5.2b) *with forward Euler is equivalent to one sweep* (5.3) *of ALS for* (3.3).

Proof. Given is $u_0 = (V_0 \otimes U_0)s_0$. Approximating the first step (5.2a) by the implicit Euler method, having factored out $(V_0 \otimes I)$, amounts to

$$(I \otimes U_1)s_1^+ + h(V_0^T \otimes I)\tilde{A}_1(V_0 \otimes U_1)s_1^+ = (I \otimes U_0)s_0 + h(V_0^T \otimes I)\tilde{f}_1, \tag{5.4}$$

which is precisely (5.3a) combined with (5.3b). In the second step we solve (5.2b) with the explicit Euler method and with $\tilde{A}$ and $\tilde{f}$ evaluated at the final time, that is,

$$s_0^+ = s_1^+ + h(V_0^T \otimes U_1^T)\tilde{A}_1(V_0 \otimes U_1)s_1^+ - h(V_0^T \otimes U_1^T)\tilde{f}_1.$$

Multiplying (5.4) by $(I \otimes U_1^T)$ we thus obtain

$$s_0^+ = (I \otimes U_1^T U_0)s_0. \tag{5.5}$$

In the third step we solve (5.2c) with the implicit Euler method. This reads

$$(V_1 \otimes I)s_1 + h(I \otimes U_1^T)\tilde{A}_1(V_1 \otimes U_1)s_1 = (V_0 \otimes I)s_0^+ + h(I \otimes U_1^T)\tilde{f}_1.$$

With (5.5) this is equivalent to

$$(I + h(I \otimes U_1^T)A(I \otimes U_1))(V_1 \otimes I)s_1 = (V_0 \otimes U_1^T U_0)s_0 + h(I \otimes U_1^T)f_1,$$

which is precisely the combination of (5.3c) and (5.3d). $\square$

6. Outlook

We expect that the obtained existence and uniqueness result is applicable to dynamical low-rank tensor approximations [27, 32, 31] of higher-dimensional parabolic problems in suitable low-rank formats. Beyond the intrinsic interest of parabolic evolution equations under low-rank constraints (or on more general manifolds), the dynamical low-rank approach can also be of interest as an algorithmic component in approximation schemes involving rank adaptivity. For instance, with constant right-hand side f, performing the low-rank evolution for $u' + Au = f$ to sufficiently large times yields an approximation of $A^{-1}f$. The approach considered here can thus be used in the construction of preconditioners for low-rank approximations of elliptic problems with strongly anisotropic diffusion, where the existing methods for Laplacian-type operators are less efficient.

Appendix A. A curvature bound in Hilbert space

In this section we generalize known curvature bounds for finite-dimensional fixed-rank matrix manifolds to the Hilbert space case. We first consider

$$\mathcal{M} = \{X \in \ell_2(\mathbb{N}^2) : \operatorname{rank} X = r\},$$

the manifold of fixed rank-r infinite matrices in the real tensor product Hilbert space $\ell_2(\mathbb{N}^2) = \ell_2(\mathbb{N}) \otimes \ell_2(\mathbb{N})$. For $X \in \mathcal{M}$ we have the singular value decomposition

$$X = \sum_{i=1}^{r} \sigma_i u_i^1 \otimes u_i^2, \quad \sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0, \ (u_i^1, u_j^1)_{\ell_2} = (u_i^2, u_j^2)_{\ell_2} = \delta_{i,j}.$$

The induced norm in the tensor product space is the Hilbert–Schmidt norm,

$$\begin{aligned}\|X\|^2_{\ell_2(\mathbb{N}^2)} = \|X\|^2_{\ell_2(\mathbb{N})\otimes\ell_2(\mathbb{N})} &= \Big(\sum_i \sigma_i u_i^1 \otimes u_i^2, \sum_j \sigma_j u_j^1 \otimes u_j^2\Big)_{\ell_2(\mathbb{N})\otimes\ell_2(\mathbb{N})} \\ &= \sum_{i,j} \sigma_i\sigma_j (u_i^1 \otimes u_i^2, u_j^1 \otimes u_j^2)_{\ell_2(\mathbb{N})\otimes\ell_2(\mathbb{N})} \\ &= \sum_{i,j} \sigma_i\sigma_j (u_i^1, u_j^1)_{\ell_2(\mathbb{N})}(u_i^2, u_j^2)_{\ell_2(\mathbb{N})} = \sum_i \sigma_i^2.\end{aligned}$$

Interpreting X as a linear operator on $\ell_2(\mathbb{N})$ we can also define the spectral norm

$$\|X\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})} = \sup_{\|w\|_{\ell_2(\mathbb{N})}\leq 1} \|Xw\|_{\ell_2(\mathbb{N})} = \sup_{\|w\|_{\ell_2(\mathbb{N})}\leq 1} \Big\| \sum_i \sigma_i (u_i^2, w)_{\ell_2(\mathbb{N})} u_i^1 \Big\|_{\ell_2(\mathbb{N})} = \sigma_1.$$

Then the inequality

$$\|XY\|_{\ell_2(\mathbb{N}^2)} \leq \|X\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}\|Y\|_{\ell_2(\mathbb{N}^2)}$$

holds for all $X, Y \in \ell_2(\mathbb{N}^2)$.

For a symmetric X, i.e., $u_i^1 = \pm u_i^2$ for all i, and $\kappa > 0$, we define the operator

$$P_{X,\kappa} = -\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_\kappa} (X - zI)^{-1}\,\mathrm{d}z,$$

where γ_κ is a circular path of radius κ around the origin which may not intersect the spectrum of X. This operator is the orthogonal projection onto the space spanned by all eigenvalues of X of modulus less than κ. Since X is symmetric this space is spanned by the singular vectors $u_{r+1}^1, u_{r+2}^1, \ldots$ in the SVD of X, where r is the largest i such that $\sigma_i > \kappa$.

We are now ready to state the curvature bound. The proof is an adaptation of the same result in [2, Prop. 16] for (finite-dimensional) matrices.

Lemma A.1. *Let $X \in \mathcal{M} \subset \ell_2(\mathbb{N}^2)$, with smallest nonzero singular value $\sigma_r \geq \rho > 0$, and let $Y = X + \Delta \in \mathcal{M}$ such that $\|\Delta\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} < \rho/4$. Then the tangent space projections satisfy the Lipschitz-like bound*

$$\|P_Y(Z) - P_X(Z)\|_{\ell_2(\mathbb{N}^2)} \leq \frac{4}{\rho}\|Y - X\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})}\|Z\|_{\ell_2(\mathbb{N}^2)} \tag{A.1}$$

for all $Z \in \ell_2(\mathbb{N}^2)$ and

$$\|(I - P_Y)(X - Y)\|_{\ell_2(\mathbb{N}^2)} \leq \frac{4\sqrt{2}}{\rho}\|Y - X\|^2_{\ell_2(\mathbb{N}^2)}. \tag{A.2}$$

Proof. We write the singular value decompositions of X and Y in matrix form, $X = U^1 S (U^2)^T$ and $Y = \hat{U}^1 \hat{S} (\hat{U}^2)^T$, where $S, \hat{S} \in \mathbb{R}^{r \times r}$ are diagonal, not necessarily positive, and $U^1, U^2, \hat{U}^1, \hat{U}^2 \in \ell_2(\mathbb{N} \times \{1, \dots, r\})$ have orthonormal columns. For the moment, we assume that X and Y are symmetric. Then we can assume $U^1 = U^2$ and $\hat{U}^1 = \hat{U}^2$.

If $X = U^1 S (U^1)^T$ and $Z \in \ell_2(\mathbb{N}^2)$, then by (2.7) the tangent space projection is

$$P_X(Z) = P_1 Z + Z P_1 - P_1 Z P_1 = (I - P_X^\perp)(Z) = Z - P_1^\perp Z P_1^\perp,$$

where $P_1 = U^1 (U^1)^T$ is the projection onto the space spanned by the singular vectors corresponding to non-zero singular values, and $P_1^\perp = I - P_1$. In the following we just write $P = P_1$. Also let $\hat{P} = \hat{U}^1 (\hat{U}^1)^T$. By a direct calculation one can verify that

$$P_Y(Z) - P_X(Z) = (\hat{P} - P) Z P^\perp + \hat{P}^\perp Z (\hat{P} - P).$$

Furthermore, for any $0 < \kappa < \rho$ we have

$$\hat{P} - P = P^\perp - \hat{P}^\perp = P_{X,\kappa} - P_{Y,\kappa}$$

as defined above. In the following we fix $\kappa = \rho/2$. We then have

$$\begin{aligned}\|P_Y(Z) - P_X(Z)\|_{\ell_2(\mathbb{N}^2)} &= \|(P_{X,\kappa} - P_{Y,\kappa}) Z P^\perp + \hat{P}^\perp Z (P_{X,\kappa} - P_{Y,\kappa})\|_{\ell_2(\mathbb{N}^2)} \\ &\leq \|(P_{X,\kappa} - P_{Y,\kappa}) Z\|_{\ell_2(\mathbb{N}^2)} + \|Z (P_{X,\kappa} - P_{Y,\kappa})\|_{\ell_2(\mathbb{N}^2)}.\end{aligned}$$

By the second resolvent identity,

$$\begin{aligned}(P_{X,\kappa} - P_{Y,\kappa}) Z &= -\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_\kappa} \big((X - zI)^{-1} - (Y - zI)^{-1}\big) Z \,\mathrm{d}z \\ &= -\frac{1}{2\pi \mathrm{i}} \oint_{\gamma_\kappa} (X - zI)^{-1} \Delta (Y - zI)^{-1} Z \,\mathrm{d}z,\end{aligned}$$

and therefore

$$\begin{aligned}&\|(P_{X,\kappa} - P_{Y,\kappa}) Z\|_{\ell_2(\mathbb{N}^2)} \leq \frac{1}{2\pi} \oint_{\gamma_\kappa} \|(X - zI)^{-1} \Delta (Y - zI)^{-1} Z\|_{\ell_2(\mathbb{N}^2)} \,\mathrm{d}z \leq \\ &\leq \frac{1}{2\pi} \oint_{\gamma_\kappa} \|(X - zI)^{-1}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} \|(Y - zI)^{-1}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} \,\mathrm{d}z \|\Delta\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} \|Z\|_{\ell_2(\mathbb{N}^2)}.\end{aligned}$$

Here $\|(X - zI)^{-1}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})}$ equals $1/|\lambda|$, where λ is the smallest in modulus eigenvalue of $X - zI$. Hence, since the eigenvalues of X are zero or have modulus σ_i, it is not difficult to see that

$$\max_{z \in \gamma_\kappa} \|(X - zI)^{-1}\|_{\ell_2(\mathbb{N}) \to \ell_2(\mathbb{N})} = \frac{1}{\min(\kappa, \sigma_r - \kappa)} = \frac{2}{\rho},$$

due to $\kappa = \rho/2$. The r-th singular value of Y can bounded below by $\sigma_r - \|\Delta\|_{\ell_2\to\ell_2} > \frac{3}{4}\rho$ by standard perturbation results, see, e.g. [33, Cor. 5.3]. As a result we have

$$\max_{z\in\gamma_\kappa} \|(Y - zI)^{-1}\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})} \le \frac{4}{\rho}.$$

In summary we obtain the estimate

$$\|(P_{X,\kappa} - P_{Y,\kappa})Z\|_{\ell_2(\mathbb{N}^2)} \le \frac{2\pi\kappa}{2\pi}\frac{2}{\rho}\frac{4}{\rho}\|\Delta\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}\|Z\|_{\ell_2(\mathbb{N}^2)} = \frac{4}{\rho}\|\Delta\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}\|Z\|_{\ell_2(\mathbb{N}^2)}.$$

This proves (A.1) the result for symmetric X and Y.

For the non-symmetric case we construct the formally symmetric matrix

$$\bar{X} = \begin{pmatrix} 0 & X \\ X^T & 0 \end{pmatrix} \in (\ell_2(\mathbb{N}^2))^{2\times 2}.$$

Then $\bar{X}$ has rank $2r$ and an eigendecomposition $\bar{X} = \bar{U}\bar{S}\bar{U}^T$, with

$$\bar{U} = \frac{1}{\sqrt{2}}\begin{pmatrix} U^1 & U^1 \\ U^2 & -U^2 \end{pmatrix} \in (\ell_2(\mathbb{N}))^{2\times 2r}, \quad \bar{S} = \begin{pmatrix} S & 0 \\ 0 & -S \end{pmatrix} \in \mathbb{R}^{2r\times 2r},$$

and the projection onto the space spanned by the eigenvectors corresponding to non-zero eigenvalues is

$$\bar{P} = \bar{U}\bar{U}^T = \begin{pmatrix} U^1(U^1)^T & 0 \\ 0 & U^2(U^2)^T \end{pmatrix} = \begin{pmatrix} P_1 & 0 \\ 0 & P_2 \end{pmatrix}.$$

We also construct the symmetric $\bar{Y}$, $\bar{Z}$, and $\bar{\Delta}$ accordingly. By a direct calculation, one can verify that

$$P_{\bar{X}}(\bar{Z}) = \begin{pmatrix} 0 & P_X(Z) \\ (P_X(Z))^T & 0 \end{pmatrix},$$

and thus,

$$\|P_{\bar{Y}}(\bar{Z}) - P_{\bar{X}}(\bar{Z})\|_{\ell_2(\mathbb{N}^2)^{2\times 2}} = \sqrt{2}\|P_Y(Z) - P_X(Z)\|_{\ell_2(\mathbb{N}^2)}.$$

As $\bar{X}$, $\bar{Y}$ and $\bar{Z}$ are symmetric the derivation in the first part of the proof can be applied to the left difference. Since $\|\bar{\Delta}\|_{(\ell_2(\mathbb{N}))^2\to(\ell_2(\mathbb{N}))^2} = \|\Delta\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}$ this yields

$$\|P_Y(Z) - P_X(Z)\|_{\ell_2(\mathbb{N}^2)} \le \frac{1}{\sqrt{2}}\frac{4}{\rho}\|\bar{\Delta}\|_{(\ell_2(\mathbb{N}))^2\to(\ell_2(\mathbb{N}))^2}\|\bar{Z}\|_{\ell_2(\mathbb{N}^2)} = \frac{4}{\rho}\|\Delta\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}\|Z\|_{\ell_2(\mathbb{N}^2)},$$

as desired.

For the inequality (A.2) we first note that

$$(I - P_Y)(X - Y) = (I - \hat{P}_1)(X - Y)(I - \hat{P}_2) = (I - \hat{P}_1)(X - Y)(P_2 - \hat{P}_2)$$

as $(I - \hat{P}_1)Y = 0$ and $X = XP_2$. Here we now use the notation $\hat{P}_1 = \hat{U}^1(\hat{U}^1)^T$ and $\hat{P}_2 = \hat{V}^1(\hat{V}^1)^T$. Since

$$\|P_2 - \hat{P}_2\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})} \le \|\bar{P} - \hat{\bar{P}}\|_{(\ell_2(\mathbb{N}))^2\to(\ell_2(\mathbb{N}))^2} \le \frac{4\sqrt{2}}{\rho}\|X - Y\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}$$

(where $\hat{\bar{P}}$ is defined analogously to $\bar{P}$) this implies

$$\begin{aligned}\|(I - P_Y)(X - Y)\|_{\ell_2(\mathbb{N}^2)} &\le \|I - \hat{P}_1\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})}\|X - Y\|_{\ell_2(\mathbb{N}^2)}\|P_2 - \hat{P}_2\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})} \\ &\le \|X - Y\|_{\ell_2(\mathbb{N}^2)}\frac{4\sqrt{2}}{\rho}\|X - Y\|_{\ell_2(\mathbb{N})\to\ell_2(\mathbb{N})} \le \frac{4\sqrt{2}}{\rho}\|X - Y\|^2_{\ell_2(\mathbb{N}^2)},\end{aligned}$$

concluding the proof. $\square$

Applying Lemma A.1 to sequence representations with respect to a tensor product orthonormal basis of $L_2(\Omega)$, $\Omega = (0,1)^2$, immediately gives the following result.

Corollary A.2. *For $u, v, w \in \mathcal{M}_r \subset H = L_2(\Omega)$, let ρ be a lower bound on the smallest singular values of u and v in $H = L_2(0,1) \otimes L_2(0,1)$, and let $\|u - v\|_H \leq \rho/4$. Then*

$$\|(P_u - P_v)w\|_H \leq \frac{4}{\rho}\|u - v\|_H \|w\|_H$$

and

$$\|(I - P_v)(u - v)\|_H \leq \frac{4\sqrt{2}}{\rho}\|u - v\|_H^2.$$

We further provide a proof for the estimate (2.14).

Proposition A.3. For $u \in \mathcal{M}_r \cap V$ with $V = H_0^1(\Omega)$, $\Omega = (0,1)^2$, we have

$$\|P_u\|_{V \to V} \leq \left(1 + \frac{r}{\sigma_r(u)^2}\|u\|_V\right)^{\frac{1}{2}}.$$

Proof. Let $\phi \in H_0^1(0,1)$. For the L_2-orthogonal projection P_1 on the span of the left singular vectors $u_1^1, \dots, u_r^1$, the estimate (2.13) yields

$$\begin{aligned} \|(\partial_1 \circ P_1)\phi)\|_{L_2(0,1)}^2 = \left\| \sum_{k=1}^r \langle u_k^1, \phi \rangle_{L_2(0,1)} \partial_1 u_k^1 \right\|_{L_2(0,1)}^2 &\leq \left(\sum_{k=1}^r \left|\langle \phi, u_k^1 \rangle\right|^2 \right) \left(\sum_{k=1}^r \|\partial_1 u_k^1\|_{L_2(0,1)}^2 \right) \\ &\leq r \frac{1}{\sigma_r(u)^2} \|u\|_V \|\phi\|_{L_2(0,1)}^2. \end{aligned} \tag{A.3}$$

Since $\|\phi\|_{L_2(0,1)} \leq \|\phi\|_{H_0^1(0,1)}$ by the Poincaré inequality, this shows

$$\|P_1\|_{H_0^1(0,1) \to H_0^1(0,1)} \leq \frac{\sqrt{r}}{\sigma_r(u)}\|u\|_V.$$

Using (2.7) we can write $P_u = I \otimes P_2 + P_1 \otimes (I - P_2)$, which due to (A.3) gives

$$\begin{aligned} \|P_u v\|_{H_0^1(0,1) \otimes L_2(0,1)}^2 &= \|(I \otimes P_2)v\|_{H_0^1(0,1) \otimes L_2(0,1)}^2 + \|(P_1 \otimes (I - P_2))v\|_{H_0^1(0,1) \otimes L_2(0,1)}^2 \\ &\leq \|v\|_{H_0^1(0,1) \otimes L_2(0,1)}^2 + \frac{r}{\sigma_r(u)^2}\|u\|_{H_0^1(\Omega)}^2 \|v\|_{H_0^1(0,1) \otimes L_2(0,1)}^2 \end{aligned}$$

for any $v \in V$, where we use that the operator norm of a tensor product operator equals the product of operator norms; see, e.g., [20, Prop. 4.150]. The norm $\|P_u v\|_{L_2(0,1) \otimes H_0^1(0,1)}^2$ can be estimated in the same way, so in summary we have

$$\|P_u v\|_V^2 \leq \left(1 + \frac{r}{\sigma_r(u)^2}\|u\|_V^2\right)\|v\|_V^2,$$

as asserted. □

In the proof we have used the Poincaré inequality $\|\phi_k\|_{L_2(0,1)} \leq \|\phi_k\|_{H_0^1(0,1)}$ on the interval $(0,1)$. When a general domain $\Omega = (a_1, b_1) \times (a_2, b_2)$ is considered, one can obtain a similar estimate $\|P_u\|_{V \to V}^2 \leq 1 + \bar{c}^2 \frac{r}{\sigma_r(u)^2}\|u\|_V^2$ where $\bar{c}$ is maximum of the Poincaré constants of the intervals.

References

[1] A. Arnold and T. Jahnke. On the approximation of high-dimensional differential equations in the hierarchical Tucker format. *BIT*, 54(2):305–341, 2014.

[2] F. R. Bach. Consistency of trace norm minimization. *J. Mach. Learn. Res.*, 9:1019–1048, 2008.

[3] M. Bachmayr and A. Cohen. Kolmogorov widths and low-rank approximations of parametric elliptic PDEs. *Math. Comp.*, 86(304):701–724, 2017.

[4] M. Bachmayr and W. Dahmen. Adaptive near-optimal rank tensor approximation for high-dimensional operator equations. *Found. Comput. Math.*, 15(4):839–898, 2015.

[5] M. Bachmayr and W. Dahmen. Adaptive low-rank methods: problems on Sobolev spaces. *SIAM J. Numer. Anal.*, 54(2):744–796, 2016.

[6] M. Bachmayr, R. Schneider, and A. Uschmajew. Tensor networks and hierarchical tensors for the solution of high-dimensional partial differential equations. *Found. Comput. Math.*, 16(6):1423–1472, 2016.

[7] C. Bardos, I. Catto, N. Mauser, and S. Trabelsi. Setting and analysis of the multi-configuration time-dependent Hartree-Fock equations. *Arch. Ration. Mech. Anal.*, 198(1):273–330, 2010.

[8] C. Bardos, I. Catto, N. J. Mauser, and S. Trabelsi. Global-in-time existence of solutions to the multiconfiguration time-dependent Hartree-Fock equations: a sufficient condition. *Appl. Math. Lett.*, 22(2):147–152, 2009.

[9] S. Bartels. *Numerical methods for nonlinear partial differential equations.* Springer, Cham, 2015.

[10] D. Conte and C. Lubich. An error analysis of the multi-configuration time-dependent Hartree method of quantum dynamics. *M2AN Math. Model. Numer. Anal.*, 44(4):759–780, 2010.

[11] W. Dahmen, R. DeVore, L. Grasedyck, and E. Süli. Tensor-sparsity of solutions to high-dimensional elliptic partial differential equations. *Found. Comput. Math.*, 16(4):813–874, 2016.

[12] L. De Lathauwer, B. De Moor, and J. Vandewalle. A multilinear singular value decomposition. *SIAM J. Matrix Anal. Appl.*, 21(4):1253–1278, 2000.

[13] P. A. M. Dirac. Note on exchange phenomena in the Thomas atom. *Math. Proc. Cambridge Philos. Soc.*, 26(3):376–385, 1930.

[14] L. Einkemmer and C. Lubich. A low-rank projector-splitting integrator for the Vlasov-Poisson equation. *SIAM J. Sci. Comput.*, 40(5):B1330–B1360, 2018.

[15] E. Emmrich. *Gewöhnliche und Operator-Differentialgleichungen.* Vieweg, Wiesbaden, 2004.

[16] A. Falcó, W. Hackbusch, and A. Nouy. On the Dirac-Frenkel variational principle on tensor Banach spaces. *Found. Comput. Math.*, 19(1):159–204, 2019.

[17] F. Feppon and P. F. J. Lermusiaux. A geometric approach to dynamical model order reduction. *SIAM J. Matrix Anal. Appl.*, 39(1):510–538, 2018.

[18] V. Girault and P.-A. Raviart. *Finite element approximation of the Navier-Stokes equations.* Springer-Verlag, Berlin-New York, 1979.

[19] L. Grasedyck. Existence and computation of low Kronecker-rank approximations for large linear systems of tensor product structure. *Computing*, 72(3-4):247–265, 2004.

[20] W. Hackbusch. *Tensor spaces and numerical tensor calculus.* Springer, Cham, second edition, 2019.

[21] W. Hackbusch and S. Kühn. A new scheme for the tensor representation. *J. Fourier Anal. Appl.*, 15(5):706–722, 2009.

[22] J. Haegeman, J. I. Cirac, T. J. Osborne, I. Pižorn, H. Verschelde, and F. Verstraete. Time-dependent variational principle for quantum lattices. *Phys. Rev. Lett.*, 107(7):070601, 2011.

[23] Y. Kazashi and F. Nobile. Existence of dynamical low rank approximations for random semi-linear evolutionary equations on the maximal interval. arXiv preprint arXiv:2002.02356, 2020.

[24] E. Kieri, C. Lubich, and H. Walach. Discretized dynamical low-rank approximation in the presence of small singular values. *SIAM J. Numer. Anal.*, 54(2):1020–1038, 2016.

[25] O. Koch and C. Lubich. Dynamical low-rank approximation. *SIAM J. Matrix Anal. Appl.*, 29(2):434–454, 2007.

[26] O. Koch and C. Lubich. Regularity of the multi-configuration time-dependent Hartree approximation in quantum molecular dynamics. *M2AN Math. Model. Numer. Anal.*, 41(2):315–331, 2007.

[27] O. Koch and C. Lubich. Dynamical tensor approximation. *SIAM J. Matrix Anal. Appl.*, 31(5):2360–2375, 2010.

[28] O. Koch and C. Lubich. Variational-splitting time integration of the multi-configuration time-dependent Hartree-Fock equations in electron dynamics. *IMA J. Numer. Anal.*, 31(2):379–395, 2011.

[29] C. Lubich. *From quantum to classical molecular dynamics: reduced models and numerical analysis*. European Mathematical Society (EMS), Zürich, 2008.

[30] C. Lubich and I. V. Oseledets. A projector-splitting integrator for dynamical low-rank approximation. *BIT*, 54(1):171–188, 2014.

[31] C. Lubich, I. V. Oseledets, and B. Vandereycken. Time integration of tensor trains. *SIAM J. Numer. Anal.*, 53(2):917–941, 2015.

[32] C. Lubich, T. Rohwedder, R. Schneider, and B. Vandereycken. Dynamical approximation by hierarchical Tucker and tensor-train tensors. *SIAM J. Matrix Anal. Appl.*, 34(2):470–494, 2013.

[33] A. S. Markus. Eigenvalues and singular values of the sum and product of linear operators. *Uspehi Mat. Nauk*, 19(4 (118)):93–123, 1964. In Russian, English translation in *Russ. Math. Surv.*, 19(91), 1964.

[34] H. Mena, A. Ostermann, L.-M. Pfurtscheller, and C. Piazzola. Numerical low-rank approximation of matrix differential equations. *J. Comput. Appl. Math.*, 340:602–614, 2018.

[35] H.-D. Meyer, U. Manthe, and L. S. Cederbaum. The multi-configurational time-dependent hartree approach. *Chem. Phys. Lett.*, 165(1):73–78, 1990.

[36] E. Musharbash and F. Nobile. Dual dynamically orthogonal approximation of incompressible Navier Stokes equations with random boundary conditions. *J. Comput. Phys.*, 354:135–162, 2018.

[37] E. Musharbash, F. Nobile, and T. Zhou. Error analysis of the dynamically orthogonal approximation of time dependent random PDEs. *SIAM J. Sci. Comput.*, 37(2):A776–A810, 2015.

[38] I. V. Oseledets. Tensor-train decomposition. *SIAM J. Sci. Comput.*, 33(5):2295–2317, 2011.

[39] A. Ostermann, C. Piazzola, and H. Walach. Convergence of a low-rank Lie-Trotter splitting for stiff matrix differential equations. *SIAM J. Numer. Anal.*, 57(4):1947–1966, 2019.

[40] T. P. Sapsis and P. F. J. Lermusiaux. Dynamically orthogonal field equations for continuous stochastic dynamical systems. *Phys. D*, 238(23-24):2347–2360, 2009.

[41] E. Schmidt. Zur Theorie der linearen und nichtlinearen Integralgleichungen. *Math. Ann.*, 63(4):433–476, 1907.

[42] R. Schneider and A. Uschmajew. Approximation rates for the hierarchical tensor format in periodic Sobolev spaces. *J. Complexity*, 30(2):56–71, 2014.

[43] R. E. Showalter. *Monotone operators in Banach space and nonlinear partial differential equations*. American Mathematical Society, Providence, RI, 1997.

[44] A. Uschmajew and B. Vandereycken. Geometric methods on low-rank matrix and tensor manifolds. In P. Grohs, M. Holler, and A. Weinmann, editors, *Handbook of variational methods for nonlinear geometric data*, pages 261–313. Springer, Cham, 2020.

[45] E. Zeidler. *Nonlinear functional analysis and its applications. III. Variational methods and optimization.* Springer-Verlag, New York, 1985.

[46] E. Zeidler. *Nonlinear functional analysis and its applications. II/A. Linear monotone operators.* Springer-Verlag, New York, 1990.

[47] E. Zeidler. *Applied functional analysis. Main principles and their applications*. Springer-Verlag, New York, 1995.