SIAM J. CONTROL OPTIM.
Vol. 49, No. 5, pp. 2183–2211

A PRIORI ERROR ANALYSIS OF THE PETROV–GALERKIN CRANK–NICOLSON SCHEME FOR PARABOLIC OPTIMAL CONTROL PROBLEMS*

DOMINIK MEIDNER† AND BORIS VEXLER†

Abstract. In this paper, a finite element discretization of an optimal control problem governed by the heat equation is considered. The temporal discretization is based on a Petrov–Galerkin variant of the Crank–Nicolson scheme, whereas the spatial discretization employs usual conforming finite elements. With a suitable postprocessing step, a discrete solution is obtained for which error estimates of optimal order are proven. A numerical result is presented for illustrating the theoretical findings.

Key words. optimal control, heat equation, control constraints, finite elements, Crank–Nicolson scheme, error estimates

AMS subject classifications. 49J20, 35K20, 49M05, 49M25, 49M29, 65M12, 65M60

DOI. 10.1137/100809611

1. Introduction.

In this paper, we develop an a priori error analysis for finite element discretizations of an optimal control problem which is governed by the heat equation and is subject to pointwise control constraints. The temporal discretization is a Petrov–Galerkin scheme with continuous piecewise linear ansatz functions and piecewise constant (discontinuous) test functions. This scheme is often referred to as the continuous Galerkin cG(1) method and is a variant of the Crank–Nicolson scheme; see, e.g., [2, 11, 30]. For the spatial discretization we use the usual conforming finite element methods.

The model problem under consideration is formulated as follows:

$$\text{(1.1a)} \qquad \text{Minimize } J(q,u) = \frac{1}{2}\int_0^T \int_\Omega (u(t,x) - \hat{u}(t,x))^2 \, dx \, dt + \frac{\alpha}{2}\int_0^T \sum_{i=1}^{D} q_i(t)^2 \, dt$$

subject to the state equation

$$\text{(1.1b)} \qquad \begin{aligned} \partial_t u - \Delta u &= f + Gq && \text{in } (0,T)\times\Omega, \\ u &= 0 && \text{in } (0,T)\times\partial\Omega, \\ u(0) &= u_0 && \text{in } \Omega, \end{aligned}$$

and to the control constraints

$$\text{(1.1c)} \qquad q_i^a \le q_i(t) \le q_i^b \quad \text{for } i = 1,2,\dots,D \text{ and a.e. in } (0,T),$$

where $u = u(t,x)$ denotes the state variable, $q = q(t) = (q_i(t))_{i=1}^D$ is the control variable with D time-dependent components, and G is a linear control operator to be specified later. A precise formulation of this problem including a functional analytic setting is given in the next section.

*Received by the editors September 23, 2010; accepted for publication (in revised form) August 2, 2011; published electronically October 27, 2011.
http://www.siam.org/journals/sicon/49-5/80961.html

†Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik, Boltzmannstraße 3, 85748 Garching b. München, Germany (meidner@ma.tum.de, vexler@ma.tum.de).

There are lots of publications on a priori error analysis for discretizations of optimal control problems governed by elliptic equations. First, a priori error estimates for problems with parabolic state equations can be found in [19, 32, 15, 18, 28]. In [22, 23], we developed an a priori error analysis for space-time finite element discretization of parabolic optimal control problems, where we applied discontinuous Galerkin dG(r) schemes for temporal discretization; see also [21] for the corresponding a posteriori error estimates. For parabolic problems with semilinear state equation, we refer to [7, 25], and for problems with state constraints to [20, 10].

In all articles mentioned above, the maximum order of convergence with respect to the time step k is not better than $\mathcal{O}(k)$ for problems with control constraints. The sole exception is the paper [28], where $\mathcal{O}(k^{3/2})$ for a problem in 1 space dimension with discretization of only the control variable is shown.

The temporal discretization scheme we use in this paper corresponds to the Crank–Nicolson method and therefore has the capability to be of second order. However, for a direct application of this cG(1) scheme, one can expect only the order $\mathcal{O}(k)$, even if the control variable is not discretized. This is due to the fact that the test functions for the discretization of the state equation are piecewise constant in time. Thus, a consistent discretization of the adjoint equation has to employ piecewise constant ansatz functions, and therefore only first order convergence can be achieved.

The main contribution of this paper is to show that a suitable postprocessing strategy allows for second order convergence with respect to k. For the error between the exact optimal solution $\bar{q}$ and a postprocessed discrete solution $\tilde{q}_{kh}$ (see (6.1) for its definition), we will prove the estimate

$$\|\bar{q} - \tilde{q}_{kh}\|_{L^2(0,T;\mathbb{R}^D)} = \mathcal{O}(k^2 + h^2), \tag{1.2}$$

where k and h denote the temporal and the spatial discretization parameters; see Theorem 6.2.

In a recent preprint [1], the authors consider an optimal control problem without control constraints. For a Crank–Nicolson discretization in time they obtain the second order of convergence in the midpoints of time intervals (cf. the definition of the "dual" time partition in section 3), under stronger regularity assumptions than Assumption 1 below. Another recent result on a Crank–Nicolson discretization of optimal control problems is presented in [6]. Therein an optimal control problem governed by a convection-diffusion equation and without inequality constraints is considered. For the spatial discretization, finite elements with a symmetric stabilization are applied. The temporal discretization of both the state and the adjoint equation is done by the Crank–Nicolson scheme in an optimize-then-discretize fashion. This leads—in contrast to our approach—to a nonsymmetric discrete optimality system. Second order error estimates are shown again under stronger assumptions than Assumption 1 below.

The outline of the paper is as follows. In the next section, we provide a functional analytic setting for the problem under consideration, and we discuss the optimality conditions and the regularity properties of the optimal solution. In section 3, the complete discretization is introduced. Section 4 is devoted to stability estimates for continuous, semidiscrete, and discrete equations. In section 5, we prove some technical error estimates for solutions of auxiliary equations which form the basis of our main result (1.2) proved in section 6. In section 7, we discuss a numerical example illustrating our theoretical results.

2. Continuous problem. In this section, we briefly discuss the precise formulation of the optimization problem under consideration. Furthermore, we recall theoretical results on existence, uniqueness, and regularity of optimal solutions as well as optimality conditions.

To set up a weak formulation of the state equation (1.1b), we introduce the following notation. For a convex polygonal domain $\Omega \subset \mathbb{R}^n$, $n = 2, 3$, we denote V to be $H_0^1(\Omega)$. Together with $H := L^2(\Omega)$, the Hilbert space V and its dual $V^* = H^{-1}(\Omega)$ build a Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$. Here and in what follows, we employ the usual notion for Lebesgue and Sobolev spaces.

For a time interval $I = (0, T)$ we introduce the state space

$$X := W(0,T) := \{\ v \mid v \in L^2(I,V) \text{ and } \partial_t v \in L^2(I,V^*)\ \}$$

and the control space

$$Q := L^2(I, \mathbb{R}^D)$$

with some $D \in \mathbb{N}$. The operator G defined from $\mathbb{R}^D$ to H is by means of given functions $g_i \in V$, $i = 1, 2, \dots, D$, for $p \in \mathbb{R}^D$ defined by

$$Gp := \sum_{i=1}^{D} p_i g_i.$$

Clearly, G is linear and continuous, and it can be extended to time-dependent functions via the setting $(Gq)(t) := Gq(t)$.

We use the following notation for the inner products and norms on $L^2(\Omega)$, $L^2(I, \mathbb{R}^D)$, and $L^2(I, H)$:

$$\begin{aligned} (v,w) &:= (v,w)_{L^2(\Omega)}, & \langle v,w\rangle_I &:= (v,w)_{L^2(I,\mathbb{R}^D)}, & (v,w)_I &:= (v,w)_{L^2(I,H)},\\ \|v\| &:= \|v\|_{L^2(\Omega)}, & |v|_I &:= \|v\|_{L^2(I,\mathbb{R}^D)}, & \|v\|_I &:= \|v\|_{L^2(I,H)}. \end{aligned}$$

In this setting, a standard weak formulation of the state equation (1.1b) for given control $q \in Q$, $f \in L^2(I, H)$, and $u_0 \in H$ reads as follows: Find a state $u \in X$ satisfying

$$\begin{aligned} (\partial_t u, \varphi)_I + (\nabla u, \nabla \varphi)_I &= (f + Gq, \varphi)_I \quad \forall \varphi \in X,\\ u(0) &= u_0. \end{aligned} \tag{2.1}$$

ASSUMPTION 1. *For our analysis, we will assume the following regularity properties of the data: $f, \hat{u} \in H^1(I, H)$ with $f(0), \hat{u}(T) \in V$, and $u_0 \in V$ with $\Delta u_0 \in V$.*

Using this assumption, the following result on existence and regularity can be proved.

PROPOSITION 2.1. *Under Assumption* 1 *and for fixed control $q \in Q$, there exists a unique solution $u \in X$ of problem* (2.1). *Moreover, the solution exhibits the regularity*

$$u \in L^2(I, H^2(\Omega) \cap V) \cap H^1(I, H) \hookrightarrow C(\bar{I}, V)$$

with the estimate

$$\|\partial_t u\|_I + \|\Delta u\|_I + \|\nabla u(T)\| \le C\{\|f\|_I + |q|_I + \|\nabla u_0\|\}.$$

If additionally the fixed control q is in $H^1(I,\mathbb{R}^D) \subset Q$, the state u exhibits the improved regularity

$$u \in H^1(I, H^2(\Omega) \cap V) \cap H^2(I, H),$$

and the stability estimate

$$\|\partial_t \Delta u\|_I + \|\partial_t^2 u\|_I \leq C\big\{\|f\|_{H^1(I,H)} + |q|_{H^1(I,\mathbb{R}^D)} + \|\nabla f(0)\| + \|\nabla \Delta u_0\|\big\}$$

holds.

Proof. The proof of existence and uniqueness is given, e.g., in [16] and [33].

(i) The regularity $u \in L^2(I, H^2(\Omega) \cap V) \cap H^1(I, H)$ relies on the fact that Ω is polygonal and convex and that $f + Gq \in L^2(I, H)$. It is proved, e.g., in [12]. The embedding of $L^2(I, H^2(\Omega) \cap V) \cap H^1(I, L^2(\Omega))$ into $C(\bar{I}, V)$ can be found, for instance, in [9].

(ii) The improved regularity $u \in H^1(I, H^2(\Omega) \cap V) \cap H^2(I, H)$ can be proved as in [12] provided that the right-hand side $f + Gq$ exhibits the regularity $f + Gq \in H^1(I, H)$ with $(f + Gq)(0) \in V$. This is ensured by Assumption 1, the assumed regularity $q \in H^1(I, \mathbb{R}^D)$, the concrete form of G, and the embedding $H^1(I, V) \hookrightarrow C(\bar{I}, V)$. ∎

To formulate the optimal control problem, we introduce the admissible set Q_{ad} collecting the inequality constraints (1.1c) as

$$Q_{\text{ad}} := \left\{\, q \in Q \;\middle|\; q_i^a \leq q_i(t) \leq q_i^b \text{ for } i = 1, 2, \ldots, D \text{ and a.a. } t \in I \,\right\},$$

where the bounds $q^a, q^b \in \mathbb{R}^D$ fulfill $q_i^a < q_i^b$ for $i = 1, 2, \ldots, D$.

The weak formulation of the optimal control problem (1.1) is given as follows:

$$\text{(2.2)}\quad \text{Minimize } J(q,u) := \frac{1}{2}\|u - \hat{u}\|_I^2 + \frac{\alpha}{2}|q|_I^2 \text{ subject to (2.1) and } (q,u) \in Q_{\text{ad}} \times X,$$

where $\alpha > 0$ is the regularization parameter.

Proposition 2.2. *For $\alpha > 0$ the optimal control problem* (2.2) *admits a unique solution $(\bar{q}, \bar{u}) \in Q_{ad} \times X$.*

Proof. For the standard proof we refer, e.g., to [16]. ∎

The existence result for the state equation in Proposition 2.1 ensures the existence of a control-to-state mapping $q \mapsto u = u(q)$ defined through (2.1). By means of this mapping, we introduce the reduced cost functional $j\colon Q \to \mathbb{R}$ as

$$j(q) := J(q, u(q)).$$

The optimal control problem (2.2) can then be equivalently reformulated as follows:

$$\text{Minimize } j(q) \text{ subject to } q \in Q_{\text{ad}}. \tag{2.3}$$

The first order necessary optimality condition for (2.3) reads as

$$j'(\bar{q})(\delta q - \bar{q}) \geq 0 \quad \forall \delta q \in Q_{\text{ad}}. \tag{2.4}$$

Due to the linear-quadratic structure of the optimal control problem, this condition is also sufficient for optimality.

Utilizing the adjoint state equation for $z = z(q) \in X$ given by

$$\begin{aligned} -(\varphi, \partial_t z)_I + (\nabla \varphi, \nabla z)_I &= (\varphi, u(q) - \hat{u})_I \quad \forall \varphi \in X, \\ z(T) &= 0, \end{aligned} \tag{2.5}$$

the first derivative of the reduced cost functional can be expressed as

$$j'(q)(\delta q) = \langle \alpha q + G^* z(q), \delta q \rangle_I, \tag{2.6}$$

where $G^*\colon H \to \mathbb{R}^D$ denotes the adjoint of G.

Using a pointwise projection on the admissible set Q_{ad},

$$P_{Q_{\text{ad}}}\colon Q \to Q_{\text{ad}}, \quad P_{Q_{\text{ad}}}(q)_i(t) = \max(q_i^a, \min(q_i^b, q_i(t))), \quad i = 1, 2, \dots, D, \tag{2.7}$$

the optimality condition (2.4) can be expressed as

$$\bar{q} = P_{Q_{\text{ad}}}(-\alpha^{-1} G^* z(\bar{q})). \tag{2.8}$$

It is well known that for $1 \le p \le \infty$ and $r \in W^{1,p}(I, \mathbb{R}^D)$ the projection $P_{Q_{\text{ad}}}$ possesses the property

$$\|P_{Q_{\text{ad}}}(r)\|_{W^{1,p}(I,\mathbb{R}^D)} \le \|r\|_{W^{1,p}(I,\mathbb{R}^D)}. \tag{2.9}$$

Employing formulation (2.8) of the optimality condition, we obtain the following regularity result.

Proposition 2.3. *Let $(\bar{q}, \bar{u})$ be the solution of the optimization problem* (2.2) *and $\bar{z} = z(\bar{q})$ be the corresponding adjoint state. Then, there holds*

$$\bar{u}, \bar{z} \in H^1(I, H^2(\Omega) \cap V) \cap H^2(I, H), \qquad \bar{q} \in W^{1,\infty}(I, \mathbb{R}^D).$$

Furthermore, the following stability estimates are fulfilled:

$$\|\partial_t \Delta \bar{u}\|_I + \|\partial_t^2 \bar{u}\|_I \le C\big\{\|f\|_{H^1(I,H)} + \|\bar{q}\|_{H^1(I,\mathbb{R}^D)} + \|\nabla f(0)\| + \|\nabla \Delta u_0\|\big\},$$
$$\|\partial_t \Delta \bar{z}\|_I + \|\partial_t^2 \bar{z}\|_I \le C\big\{\|\hat{u}\|_{H^1(I,H)} + \|\nabla \hat{u}(T)\| + \|f\|_I + |q|_I + \|\nabla u_0\|\big\}.$$

Proof. For $\bar{q} \in Q$ Proposition 2.1 implies that $\bar{u} \in L^2(I, H^2(\Omega) \cap V) \cap H^1(I, H)$. This implies that the right-hand side of the adjoint equation (2.5) fulfills $\bar{u} - \hat{u} \in H^1(I, H)$, and since $L^2(I, H^2(\Omega) \cap V) \cap H^1(I, H) \hookrightarrow C(\bar{I}, V)$, we have that $\bar{u}(T) - \hat{u}(T) \in V$. Consequently, we obtain by Proposition 2.1 that $\bar{z} \in H^1(I, H^2(\Omega) \cap V) \cap H^2(I, H)$ and $G^* \bar{z} \in H^2(I, \mathbb{R}^D) \hookrightarrow W^{1,\infty}(I, \mathbb{R}^D)$. By property (2.9), this implies the stated regularity of $\bar{q}$.

The stability estimates for $\bar{u}$ follow directly from Proposition 2.1. For $\bar{z}$, Proposition 2.1 applied to the adjoint equation (2.5) implies

$$\|\partial_t \Delta \bar{z}\|_I + \|\partial_t^2 \bar{z}\|_I \le C\big\{\|\bar{u}\|_{H^1(I,H)} + \|\hat{u}\|_{H^1(I,H)} + \|\nabla u(T)\| + \|\nabla \hat{u}(T)\|\big\}$$

and the estimate from Proposition 2.1 yields the assertion. ◻

3. Discretization. In this section, we describe the space-time finite element discretization of the optimal control problem (2.2).

3.1. Semidiscretization in time. At first, we present the semidiscretization in time of the state equation by continuous Galerkin methods. We consider a partitioning of the time interval $\bar{I} = [0, T]$ as

$$\bar{I} = \{\, 0 \,\} \cup I_1 \cup I_2 \cup \cdots \cup I_M \tag{3.1}$$

with subintervals $I_m = (t_{m-1}, t_m]$ of size k_m and time points

$$0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T.$$

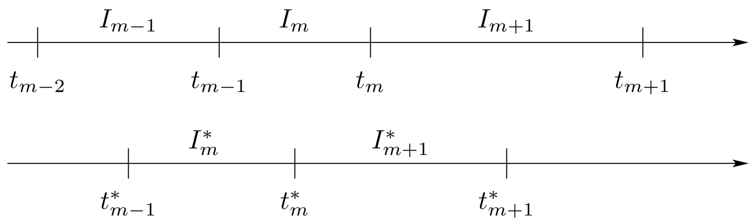


FIG. 3.1. *Partitions* $\{ I_m \mid m = 1, 2, \dots, M \}$ *(above) and* $\{ I_m^* \mid m = 1, 2, \dots, M+1 \}$ *(below) of the time interval* I.

We define the discretization parameter k as a piecewise constant function by setting $k\big|_{I_m} = k_m$ for $m = 1, 2, \dots, M$. Moreover, we denote by k the maximal size of the time steps, i.e., $k = \max_{m=1,2,\dots,M} k_m$. We impose the following conditions on the time mesh:

(i) There is a constant $\kappa > 0$ (independent of k) such that for all $m = 1, 2, \dots, M-1$,

$$\kappa^{-1} \le \frac{k_m}{k_{m+1}} \le \kappa$$

holds.

(ii) There is a constant $\gamma > 0$ (independent of k) such that

$$k \le \gamma \min_{m=1,2,\dots,M} k_m.$$

The semidiscrete trial space is given as

$$X_k^r = \left\{ v_k \in C(\bar{I}, V) \;\middle|\; v_k\big|_{I_m} \in \mathcal{P}_r(I_m, V),\ m = 1, 2, \dots, M \right\},$$

while the test space consisting of discontinuous piecewise polynomials of order $r-1$ is defined as

$$\tilde{X}_k^{r-1} = \left\{ v_k \in L^2(I, V) \;\middle|\; v_k\big|_{I_m} \in \mathcal{P}_{r-1}(I_m, V),\ m = 1, 2, \dots, M,\ v_k(0) \in V \right\}.$$

Here, $\mathcal{P}_r(I_m, V)$ denotes the space of polynomials up to order r defined on I_m with values in V. We use the notation

$$(v, w)_{I_m} := (v, w)_{L^2(I_m, H)} \quad \text{and} \quad \|v\|_{I_m} := \|v\|_{L^2(I_m, H)}.$$

Additionally later on we will need the "dual" partition of the time interval $\bar{I}$ (cf. also [1]) defined by

$$\bar{I} = \{ 0 \} \cup I_1^* \cup I_2^* \cup \dots \cup I_{M+1}^*$$

with $I_m^* := (t_{m-1}^*, t_m^*]$ for $m = 1, 2, \dots, M+1$ and

$$t_0^* := t_0, \quad t_m^* := \frac{t_{m-1} + t_m}{2} \text{ for } m = 1, 2, \dots, M, \quad \text{and} \quad t_{M+1}^* := t_M; \tag{3.2}$$

see Figure 3.1. On this partition, we define the space Q_k^r by

$$Q_k^r := \left\{ w_k \in C(\bar{I}, \mathbb{R}^D) \;\middle|\; w_k\big|_{I_m^*} \in \mathcal{P}_r(I_m^*, \mathbb{R}^D),\ m = 1, 2, \dots, M+1 \right\},$$

and on the "usual" partition we define the space $\tilde{Q}_k^{r-1}$ by

$$\tilde{Q}_k^{r-1} := \left\{ w_k \in L^2(I, \mathbb{R}^D) \,\middle|\, w_k\big|_{I_m} \in \mathcal{P}_{r-1}(I_m, \mathbb{R}^D),\ m = 1, 2, \ldots, M,\ w_k(0) \in \mathbb{R}^D \right\}.$$

To define the continuous Galerkin (cG(r)) approximation using the spaces X_k^r and $\tilde{X}_k^r$, we use for $w_k \in X_k^r$ the abbreviation $w_{k,m} := w_k(t_m)$, and employ the following definitions for functions $w_k \in \tilde{X}_k^{r-1}$:

$$w_{k,m}^+ := \lim_{t \to 0^+} w_k(t_m + t), \quad w_{k,m}^- := \lim_{t \to 0^+} w_k(t_m - t), \quad [w_k]_m := w_{k,m}^+ - w_{k,m}^-.$$

The bilinear form $B(\cdot, \cdot)$ for $u_k \in X_k^r$ and $\varphi \in \tilde{X}_k^{r-1}$ is then defined by

$$B(u_k, \varphi) := (\partial_t u_k, \varphi)_I + (\nabla u_k, \nabla \varphi)_I + (u_{k,0}, \varphi_0^-). \tag{3.3}$$

Throughout the paper we restrict ourselves to the case $r = 1$. The resulting cG(1) scheme is a variant of the Crank–Nicolson method; cf. [3]. The cG(1) semidiscretization of the state equation (2.1) for a given control $q \in Q$ reads as follows: Find a state $u_k = u_k(q) \in X_k^1$ such that

$$B(u_k, \varphi) = (f + Gq, \varphi)_I + (u_0, \varphi_0^-) \quad \forall \varphi \in \tilde{X}_k^0. \tag{3.4}$$

The existence and uniqueness of solutions to (3.4) with $r = 1$ can be directly shown by "elliptic" arguments. For the general cG(r) case we refer to [30].

Remark 3.1. Due to Proposition 2.1, we have that the exact solution $u = u(q) \in X$ lies in $C(\bar{I}, V)$. Using a density argument, it is possible to show that u also satisfies the identity

$$B(u, \varphi) = (f + Gq, \varphi)_I + (u_0, \varphi_0^-) \quad \forall \varphi \in \tilde{X}_k^0.$$

Thus, we have here the property of Galerkin orthogonality

$$B(u - u_k, \varphi) = 0 \quad \forall \varphi \in \tilde{X}_k^0.$$

The semidiscrete optimization problem for the cG(1) time discretization has the following form:

$$\text{Minimize } J(q_k, u_k) \text{ subject to (3.4) and } (q_k, u_k) \in Q_{\text{ad}} \times X_k^1. \tag{3.5}$$

As in the continuous case, the following result holds.

PROPOSITION 3.2. *For $\alpha > 0$, the semidiscrete optimal control problem* (3.5) *admits a unique solution $(\bar{q}_k, \bar{u}_k) \in Q_{ad} \times X_k^1$.*

Note that the optimal control $\bar{q}_k$ is searched for in the subset Q_{ad} of the continuous space Q, and the subscript k indicates only the usage of the semidiscretized state equation.

Similarly to the continuous case, we introduce the semidiscrete reduced cost functional $j_k \colon Q \to \mathbb{R}$ by

$$j_k(q) := J(q, u_k(q))$$

and reformulate the semidiscrete optimal control problem (3.5) as follows:

$$\text{Minimize } j_k(q_k) \text{ subject to } q_k \in Q_{\text{ad}}.$$

The first order necessary optimality condition reads as

$$j_k'(\bar q_k)(\delta q - \bar q_k) \ge 0 \quad \forall \delta q \in Q_{\mathrm{ad}}, \tag{3.6}$$

and the derivative of j_k can be expressed as

$$j_k'(q)(\delta q) = \langle \alpha q + G^* z_k(q), \delta q\rangle_I. \tag{3.7}$$

Here, $z_k = z_k(q) \in \tilde X_k^0$ denotes the solution of the semidiscrete adjoint equation

$$B(\varphi, z_k) = (\varphi, u_k(q) - \hat u)_I \quad \forall \varphi \in X_k^1. \tag{3.8}$$

Remark 3.3. Note that in contrast to the state equation (3.4), in the formulation of the adjoint equation (3.8) the space $\tilde X_k^0$ is used as trial space and X_k^1 acts as test space. For the formal derivation of the adjoint equation and its formulation as a time stepping scheme, we refer to [3].

Similarly to (2.8), the optimality condition (3.6) can be rewritten as

$$\bar q_k = P_{Q_{\mathrm{ad}}}(-\alpha^{-1} G^* z_k(\bar q_k)). \tag{3.9}$$

This yields that $\bar q_k$ is piecewise constant in time, i.e., that $\bar q_k \in \tilde Q_k^0$.

For the error analysis derived later, we will make use of the following projections into the time-discrete spaces $\tilde X_k^0$, $\tilde Q_k^0$, and Q_k^1:

1. L^2 projection $P_k \colon L^2(I,V) \to \tilde X_k^0$ given by

$$P_k v\big|_{I_m} = \frac{1}{k_m}\int_{I_m} v(t)\,dt \quad \text{for } m = 1,2,\dots,M \quad \text{and} \quad (P_k v)_0^- = 0. \tag{3.10}$$

2. Interpolation $\Pi_k \colon \begin{cases} C(\bar I, V) \to \tilde X_k^0, \\ C(\bar I, \mathbb{R}^D) \to \tilde Q_k^0 \end{cases}$ given by

$$\Pi_k v\big|_{I_m} \coloneqq v(t_m^*) \quad \text{for } m = 1,2,\dots,M \quad \text{and} \quad (\Pi_k v)_0^- \coloneqq v(0), \tag{3.11}$$

where t_m^* is given as in (3.2).

3. Interpolation $\pi_k \colon C(\bar I, \mathbb{R}^D) \cup \tilde Q_k^0 \to Q_k^1$ defined by

$$\begin{aligned} \pi_k v(t)\big|_{I_1^* \cup I_2^*} &\coloneqq v(t_1^*) + \frac{t - t_1^*}{t_2^* - t_1^*}(v(t_2^*) - v(t_1^*)), \\ \pi_k v(t)\big|_{I_m^*} &\coloneqq v(t_{m-1}^*) + \frac{t - t_{m-1}^*}{t_m^* - t_{m-1}^*}(v(t_m^*) - v(t_{m-1}^*)),\ m = 3,4,\dots,M-1, \\ \pi_k v(t)\big|_{I_M^* \cup I_{M+1}^*} &\coloneqq v(t_{M-1}^*) + \frac{t - t_{M-1}^*}{t_M^* - t_{M-1}^*}(v(t_M^*) - v(t_{M-1}^*)). \end{aligned} \tag{3.12}$$

The action of the interpolation operator π_k is depicted in Figure 3.2.

For the interpolation error $y - \Pi_k y$, the well-known estimates

$$\|y - \Pi_k y\|_{I_m} \le C k_m \|\partial_t y\|_{I_m} \quad \text{and} \quad \|y - \Pi_k y\|_{L^\infty(I_m, H)} \le C k_m^{\frac12} \|\partial_t y\|_{I_m} \tag{3.13}$$

hold for all $y \in H^1(I_m, H) \cap C(\bar I, V)$; see, e.g., [31].

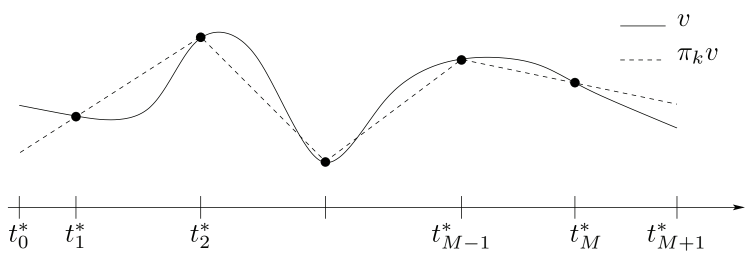


FIG. 3.2. *Interpolation* π_k.

3.2. Discretization in space. To define the finite element discretization in space, we consider two or three dimensional shape-regular meshes; see, e.g., [8]. A mesh consists of quadrilateral or hexahedral cells K, which constitute a nonoverlapping cover of the computational domain Ω. The corresponding mesh is denoted by $\mathcal{T}_h = \{K\}$, where we define the discretization parameter h as a cellwise constant function by setting $h\big|_K = h_K$ with the diameter h_K of the cell K. We use the symbol h also for the maximal cell size, i.e., $h = \max h_K$.

On the mesh $\mathcal{T}_h$ we construct a conforming finite element space $V_h \subset V$ in a standard way:

$$V_h^s = \left\{ \, v \in V \;\middle|\; v\big|_K \in \mathcal{Q}_s(K) \text{ for } K \in \mathcal{T}_h \, \right\}.$$

Here, $\mathcal{Q}_s(K)$ consists of shape functions obtained via (bi-/tri-)linear transformations of polynomials in $\widehat{\mathcal{Q}_s}(\widehat{K})$ defined on the reference cell $\widehat{K} = (0,1)^n$, where

$$\widehat{\mathcal{Q}_s}(\widehat{K}) = \operatorname{span} \left\{ \prod_{j=1}^{n} x_j^{\alpha_j} \;\middle|\; \alpha_j \in \mathbb{N}_0,\ \alpha_j \le s \right\}.$$

Remark 3.4. The definition of V_h^s can be adapted for the case of triangular meshes in the obvious way.

To obtain the fully discretized versions of the time discretized state equation (3.4), we utilize the space-time finite element spaces

$$X_{k,h}^{r,s} = \left\{ v_{kh} \in C(\bar{I}, V_h^s) \;\middle|\; v_{kh}\big|_{I_m} \in \mathcal{P}_r(I_m, V_h^s) \right\} \subset X_k^r$$

and

$$\tilde{X}_{k,h}^{r,s} = \left\{ v_{kh} \in L^2(I, V_h^s) \;\middle|\; v_{kh}\big|_{I_m} \in \mathcal{P}_r(I_m, V_h^s) \text{ and } v_{kh}(0) \in V_h^s \right\} \subset \tilde{X}_k^r.$$

Throughout, we will restrict ourselves to the consideration of (bi-/tri-)linear elements; i.e., we set $s = 1$ and consider the so-called cG(1)cG(1) scheme. The cG(1)cG(1) discretization of the state equation for given control $q \in Q$ has the following form: Find a state $u_{kh} = u_{kh}(q) \in X_{k,h}^{1,1}$ such that

$$B(u_{kh}, \varphi) = (f + Gq, \varphi)_I + (u_0, \varphi_0^-) \quad \forall \varphi \in \tilde{X}_{k,h}^{0,1}. \tag{3.14}$$

Remark 3.5. The notation cG(s)cG(r) is taken from [11] and describes a method with conforming (continuous) discretization in the space of order s and continuous discretization in time of order r.

Then, the corresponding optimal control problem is given as follows:

$$\text{Minimize } J(q_{kh}, u_{kh}) \text{ subject to (3.14) and } (q_{kh}, u_{kh}) \in Q_{\text{ad}} \times X_{k,h}^{1,1}, \tag{3.15}$$

and by means of the discrete reduced cost functional $j_{kh} \colon Q \to \mathbb{R}$,

$$j_{kh}(q) := J(q, u_{kh}(q)),$$

it can be reformulated as follows:

$$\text{Minimize } j_{kh}(q_{kh}) \text{ subject to } q_{kh} \in Q_{\text{ad}}.$$

The uniquely determined optimal solution of (3.15) is denoted by $(\bar{q}_{kh}, \bar{u}_{kh}) \in Q_{\text{ad}} \times X_{k,h}^{1,1}$.

The optimal control $\bar{q}_{kh} \in Q_{\text{ad}}$ fulfills the first order optimality condition

$$j'_{kh}(\bar{q}_{kh})(\delta q - \bar{q}_{kh}) \geq 0 \quad \forall \delta q \in Q_{\text{ad}}, \tag{3.16}$$

where $j'_{kh}(q)(\delta q)$ is given by

$$j'_{kh}(q)(\delta q) = \langle \alpha q + G^* z_{kh}(q), \delta q \rangle_I \tag{3.17}$$

with the discrete adjoint solution $z_{kh} = z_{kh}(q) \in \tilde{X}_{k,h}^{0,1}$ of

$$B(\varphi, z_{kh}) = (\varphi, u_{kh}(q) - \hat{u})_I \quad \forall \varphi \in X_{k,h}^{1,1}. \tag{3.18}$$

As before, the optimality condition (3.16) can be rewritten in terms of the projection $P_{Q_{\text{ad}}}$ as

$$\bar{q}_{kh} = P_{Q_{\text{ad}}}(-\alpha^{-1} G^* z_{kh}(\bar{q}_{kh})). \tag{3.19}$$

For the error analysis derived later, we will make use of the Ritz projection $R_h \colon \tilde{X}_k^0 \to \tilde{X}_{k,h}^{0,1}$ defined by

$$(\nabla R_h v_k, \nabla \varphi)_I = (\nabla v_k, \nabla \varphi)_I \quad \forall \varphi \in \tilde{X}_{k,h}^{0,1}. \tag{3.20}$$

3.3. Discretization of the controls. From the definition of G, we have that $G^* \colon H \to \mathbb{R}^D$ and consequently that $G^* z_{kh}$ is piecewise constant in time with values in $\mathbb{R}^D$. Hence, by inspection of the optimality condition (3.19), we obtain that $\bar{q}_{kh} \in \tilde{Q}_k^0$ and does not need to be discretized explicitly.

4. Stability estimates. In this section we consider the auxiliary solutions $v \in X$ and $y \in X$ of

$$\begin{aligned} \partial_t v - \Delta v &= g && \text{in } (0,T) \times \Omega, \\ v &= 0 && \text{in } (0,T) \times \partial\Omega, \\ v(0) &= 0 && \text{in } \Omega \end{aligned} \tag{4.1}$$

and

$$\begin{aligned} -\partial_t y - \Delta y &= g && \text{in } (0,T) \times \Omega, \\ y &= 0 && \text{in } (0,T) \times \partial\Omega, \\ y(T) &= 0 && \text{in } \Omega \end{aligned} \tag{4.2}$$

with the right-hand side $g \in L^2(I,H)$. Furthermore, we consider their semidiscrete and discrete analogues, i.e., the solutions $v_k \in X_k^1$ and $y_k \in \tilde{X}_k^0$ of

$$B(v_k,\varphi) = (g,\varphi)_I \quad \forall \varphi \in \tilde{X}_k^0 \tag{4.3}$$

and

$$B(\varphi,y_k) = (\varphi,g)_I \quad \forall \varphi \in X_k^1, \tag{4.4}$$

as well as the solutions $v_{kh} \in X_{k,h}^{1,1}$ and $y_{kh} \in \tilde{X}_{k,h}^{0,1}$ of

$$B(v_{kh},\varphi) = (g,\varphi)_I \quad \forall \varphi \in \tilde{X}_{k,h}^{0,1} \tag{4.5}$$

and

$$B(\varphi,y_{kh}) = (\varphi,g)_I \quad \forall \varphi \in X_{k,h}^{1,1}. \tag{4.6}$$

At first, we recall a standard stability result for the continuous solution $y \in X$ of (4.2).

PROPOSITION 4.1. *For the solution $y \in X$ of* (4.2)*, it holds that*

$$\|\partial_t y\|_I + \|\Delta y\|_I + \|\nabla y\|_I + \max_{t\in I}\|y(t)\| \le C\|g\|_I.$$

Proof. The proof can be found, e.g., in [12]. ∎

In what follows, we present stability estimates for the semidiscrete and discrete auxiliary solutions v_k, y_k, v_{kh}, and y_{kh} which are needed in the following sections.

LEMMA 4.2. *For the solution $v_k \in X_k^1$ of* (4.3) *and P_k defined by* (3.10)*, it holds that*

$$\|\partial_t v_k\|_I + \|\nabla v_{k,M}\| \le \|g\|_I \quad \textit{and} \quad \|P_k \Delta v_k\|_I \le C\|g\|_I.$$

Proof. First, we observe that by choosing $\varphi \in \tilde{X}_k^0$ in (4.3) such that $\varphi\big|_{I_m} = 0$ for $m = 1,2,\dots,M$ and $\varphi_0^- = v_{k,0}$, we obtain that $v_{k,0} = 0$.

(i) To estimate $\partial_t v_k$, we test (4.3) with $\varphi \in \tilde{X}_k^0$ defined by $\varphi\big|_{I_m} = \partial_t v_k\big|_{I_m}$ for $m = 1,2,\dots,M$ with $\varphi_0^- = 0$ to get

$$\|\partial_t v_k\|_I^2 + (\nabla v_k, \partial_t \nabla v_k)_I = (g, \partial_t v_k)_I,$$

and hence

$$\|\partial_t v_k\|_I^2 + \|\nabla v_{k,M}\|^2 \le \|g\|_I^2 + \|\nabla v_{k,0}\|^2, \tag{4.7}$$

which implies the first assertion since $v_{k,0} = 0$.

(ii) Next, we want to choose $\varphi = -P_k \Delta v_k$. For applying integration by parts in space to (4.3), it is necessary to prove $\Delta v_k\big|_{I_m} \in \mathcal{P}_1(I_m,H)$. This assertion follows immediately from applying elliptic regularity theory (cf. [12]) to

$$(\nabla v_k, \nabla\varphi)_{I_m} = (g - \partial_t v_k, \varphi)_{I_m} \quad \forall \varphi \in \mathcal{P}_0(I_m,V),$$

which is obtained from (4.3) by choosing $\varphi\big|_{I_l} = 0$ for all $l \ne m$. The fact that $v_k\big|_{I_m}$ is polynomial in time with values in $V \subset H$ implies that the right-hand side is in H for almost all $t \in I_m$. Thus, $\Delta v_k\big|_{I_m}$ is also in H for almost all $t \in I_m$, and since $v_k\big|_{I_m}$ is polynomial with respect to time, this yields $\Delta v_k\big|_{I_m} \in \mathcal{P}_1(I_m, H)$.

Consequently, it is feasible to integrate (4.3) by parts in space to obtain the formulation

$$(\partial_t v_k, \varphi)_I - (\Delta v_k, \varphi)_I = (g, \varphi)_I \quad \forall \varphi \in \tilde{X}_k^0. \tag{4.8}$$

The arising boundary terms vanish due to the prescribed homogeneous Dirichlet boundary conditions.

Since there are no spatial derivatives on the test function φ anymore, formulation (4.8) holds not only for all $\varphi \in \tilde{X}_k^0$ but by the density of V in H holds also for all φ with $\varphi\big|_{I_m} \in \mathcal{P}_0(I_m, H)$ for $m = 1, 2, \ldots, M$. Hence, we may choose $\varphi = -P_k \Delta v_k$ as a test function and get

$$-(\partial_t v_k, P_k \Delta v_k)_I + (\Delta v_k, P_k \Delta v_k)_I = -(g, P_k \Delta v_k)_I,$$

which implies by orthogonality and (4.7) that

$$\|P_k \Delta v_k\|_I \le \|\partial_t v_k\|_I + \|g\|_I \le 2\|g\|_I.$$

This completes the proof. ∎

COROLLARY 4.3. *For the solution $v_{kh} \in X_{k,h}^{1,1}$ of* (4.5)*, it holds that*

$$\|\partial_t v_{kh}\|_I \le \|g\|_I.$$

Proof. The assertion can be proved similarly to how the corresponding assertion for v_k in Lemma 4.2 was proved. ∎

LEMMA 4.4. *For the solution $y_k \in \tilde{X}_k^0$ of* (4.4)*, it holds that*

$$\|y_k\|_I \le C\|g\|_I.$$

Moreover, provided that $g \in \tilde{X}_k^0$, it holds that

$$\|\Delta y_k\|_I \le C\|g\|_I.$$

Proof.

(i) To prove the first assertion, we choose in (4.4) $\varphi = w_k \in X_k^1$ such that $\partial_t w_k\big|_{I_m} = y_k\big|_{I_m}$ for $m = 1, 2, \ldots, M$ and $w_k(0) = 0$. Then, we obtain

$$\|y_k\|_I^2 + (\nabla w_k, \partial_t \nabla w_k)_I = (w_k, g)_I.$$

It follows by means of the Poincaré inequality in time that

$$\|y_k\|_I^2 + \frac{1}{2}\|\nabla w_{k,M}\|^2 \le \|w_k\|_I \|g\|_I \le C\|y_k\|_I \|g\|_I,$$

which implies the first assertion.

(ii) For the second assertion, we note that $\Delta y_k \in L^2(I, H)$ as in the proof of Lemma 4.2. Hence, we may consider the solution $v_k \in X_k^1$ of (4.3) with $g = -\Delta y_k$. Furthermore, after integration by parts in space, $\varphi = -\Delta y_k$ is a feasible test function in (4.4). Since $g \in \tilde{X}_k^0$, we get

$$\|\Delta y_k\|_I^2 = B(v_k, -\Delta y_k) = B(-\Delta v_k, y_k) = -(\Delta v_k, g)_I = -(P_k \Delta v_k, g)_I.$$

By means of Lemma 4.2 we finally obtain

$$\|\Delta y_k\|_I^2 \le \|P_k \Delta v_k\|_I \|g\|_I \le C\|\Delta y_k\|_I \|g\|_I,$$

which completes the proof. ∎

COROLLARY 4.5. *For the solution* $y_{kh} \in \tilde{X}^{0,1}_{k,h}$ *of* (4.6), *it holds that*

$$\|y_{kh}\|_I \le C\|g\|_I.$$

Proof. The assertion can be proved similarly to how the corresponding assertion for y_k in Lemma 4.4 was proved. ∎

LEMMA 4.6. *For the solution* $y_k \in \tilde{X}^0_k$ *of* (4.4), *it holds that*

$$\max_{m=1,2,\dots,M} \|y_k\|_{L^\infty(I_m,H)} \le C\|g\|_I.$$

Proof. By means of the solution $y \in X$ of (4.2), we have by Proposition 4.1 for $m = 1, 2, \dots, M$

$$\|y_k\|_{L^\infty(I_m,H)} \le \|y\|_{L^\infty(I_m,H)} + \|y - y_k\|_{L^\infty(I_m,H)} \le \|g\|_I + \|y - y_k\|_{L^\infty(I_m,H)}.$$

For $\|y - y_k\|_{L^\infty(I_m,H)}$, we proceed using the interpolation Π_k defined by (3.11):

$$\begin{aligned}
\|y - y_k\|_{L^\infty(I_m,H)} &\le \|y - \Pi_k y\|_{L^\infty(I_m,H)} + \|\Pi_k y - y_k\|_{L^\infty(I_m,H)} \\
&\le \|y - \Pi_k y\|_{L^\infty(I_m,H)} + k_m^{-1}\|\Pi_k y - y_k\|_{I_m} \\
&\le \|y - \Pi_k y\|_{L^\infty(I_m,H)} + k_m^{-1}\big\{\|\Pi_k y - y\|_{I_m} + \|y - y_k\|_{I_m}\big\}.
\end{aligned}$$

The estimates (3.13) for $y - \Pi_k y$ and Lemma 5.2 in the following section yield for $\|y - y_k\|_{I_m}$ the estimate

$$\|y - y_k\|_{I_m} \le \|y - y_k\|_I \le Ck\big\{\|\partial_t y\|_I + \|\Delta y\|_I\big\}.$$

Together with the estimate for $\|\partial_t y\|_I$ and $\|\Delta y\|_I$ from Proposition 4.1 and the assumed regularity of the time grid, this completes the proof. ∎

Finally, we derive a stability result for the solution u_k of the semidiscrete state equation (3.4).

LEMMA 4.7. *For the solution* $u_k = u_k(q) \in X^1_k$ *of the semidiscrete state equation* (3.4) *for given control* $q \in Q$ *with* $u_0 \in V$, *it holds that*

$$\|\partial_t u_k\|_I + \|\nabla u_{k,M}\| + \|u_k\|_I \le C\big\{\|f\|_I + |q|_I + \|\nabla u_0\|\big\}.$$

Proof.

(i) We first choose in (2.1) $\varphi \in \tilde{X}^0_k$ defined by $\varphi\big|_{I_m} = \partial_t u_k\big|_{I_m}$ for $m = 1, 2, \dots, M$ with $\varphi_0^- = u_{k,0}$ in (3.4) to obtain

$$\|\partial_t u_k\|_I^2 + (\nabla u_k, \partial_t \nabla u_k)_I + \|u_{k,0}\|^2 = (f + Gq, \partial_t u_k)_I + (u_0, u_{k,0}).$$

This implies

$$\begin{aligned}
&\|\partial_t u_k\|_I^2 + \frac{1}{2}\|\nabla u_{k,M}\|^2 + \|u_{k,0}\|^2 \\
&\qquad\qquad \le \|f + Gq\|_I \|\partial_t u_k\|_I + \|u_0\|\|u_{k,0}\| + \frac{1}{2}\|\nabla u_0\|^2,
\end{aligned}$$

and consequently, by Young's and Poincaré's inequalities we get the assertion on $\|\partial_t u_k\|_I$ and $\|\nabla u_{k,M}\|$ that

$$\|\partial_t u_k\|_I^2 + \|\nabla u_{k,M}\|^2 \le C\big\{\|f\|_I^2 + |q|_I^2 + \|\nabla u_0\|^2\big\}.$$

(ii) For the remaining estimate on $\|u_k\|_I$ we proceed as follows by means of Poincaré's inequality:

$$\|u_k\|_I \le C\|\nabla u_k\|_I \le C\sqrt{T} \max_{m=1,2,\dots,M} \|\nabla u_{k,m}\|.$$

Then, the previous estimate on $\|\nabla u_{k,M}\|$ which also holds for $\|\nabla u_{k,M}\|$ with $m = 1, 2, \dots, M-1$ implies the second assertion.

This completes the proof. ∎

COROLLARY 4.8. *For the solution $u_{kh} = u_{kh}(q) \in X_k^1$ of the semidiscrete state equation* (3.14) *for given control $q \in Q$ with $u_0 \in V$, it holds that*

$$\|\partial_t u_{kh}\|_I + \|\nabla u_{kh,M}\| + \|u_{kh}\|_I \le C\big\{\|f\|_I + |q|_I + \|\nabla u_0\|\big\}.$$

Proof. The assertion can be proved similarly to how the corresponding assertion for u_k in Lemma 4.7 was proved. ∎

5. Analysis of the discretization error for auxiliary equations. The aim of this section is the derivation of error estimates for the auxiliary solutions y, y_k, y_{kh} introduced in section 4.

5.1. Analysis of the temporal discretization error.

LEMMA 5.1. *For $m \in \{1, 2, \dots, M\}$, $w \in H^2(I_m, H) \cap C(\bar{I}, V)$, and $\psi_k \in \mathcal{P}_0(I_m, H)$, it holds for the projection Π_k given by* (3.11) *that*

$$(\psi_k, w - \Pi_k w)_{I_m} \le C k_m^{\frac{5}{2}} \|\psi_{k,m}\| \|\partial_t^2 w\|_{I_m}.$$

Proof. Since $\psi_k \in \mathcal{P}_0(I_m, H)$, we have

$$(\psi_k, w - \Pi_k w)_{I_m} \le \|\psi_{k,m}\| \int_{I_m} \|w(t) - \Pi_k w(t)\|\, dt. \tag{5.1}$$

By means of a Taylor expansion, the integral on the right-hand side can be estimated as

$$\begin{aligned}
\int_{I_m} \|w(t) - \Pi_k w(t)\|\, dt &= \int_{I_m} \|w(t) - w(t_m^*)\|\, dt \\
&\le \int_{I_m} (t - t_m^*) \|\partial_t w(t_m^*)\|\, dt + \int_{I_m} \int_{t_m^*}^{t} (s - t) \|\partial_t^2 w(t)\|\, ds\, dt \\
&= \int_{I_m} \|\partial_t^2 w(t)\| \int_{t_m^*}^{t} (s - t_m^*)\, ds\, dt \\
&\le C k_m^2 \int_{I_m} \|\partial_t^2 w(t)\|\, dt \le C k_m^{\frac{5}{2}} \|\partial_t^2 w\|_{I_m}.
\end{aligned}$$

Inserting this into (5.1) completes the proof. ∎

LEMMA 5.2. *For the solutions $y \in X$ of* (4.2) *and $y_k \in \tilde{X}_k^0$ of* (4.4), *it holds that*

$$\|y - y_k\|_I \le Ck\|g\|_I.$$

Proof. We split

$$\|y - y_k\|_I \le \|y - P_k y\|_I + \|P_k y - y_k\|_I,$$

where P_k is defined as in (3.10), and for the first term we have by standard estimates for the L^2 projection

$$\|y - P_k y\|_I \le Ck\|\partial_t y\|_I. \tag{5.2}$$

For the second term, we proceed as follows by employing the solution $v_k \in X_k^1$ of (4.3) with $g = y_k - P_k y$, the Galerkin orthogonality of $y - y_k$ to X_k^1, and the fact that $v_{k,0} = 0$:

$$\begin{aligned}\|y_k - P_k y\|^2 &= B(v_k, y_k - P_k y) = B(v_k, y - P_k y)\\ &= (\partial_t v_k, y - P_k y)_I + (\nabla v_k, \nabla(y - P_k y))_I.\end{aligned}$$

Since $\partial_t v_k \in \tilde{X}_k^0$, the term $(\partial_t v_k, y - P_k y)_I$ vanishes, and using the properties of P_k, we get

$$\begin{aligned}\|y_k - P_k y\|^2 = (\nabla(v_k - P_k v_k), \nabla(y - P_k y))_I &= (\nabla(v_k - P_k v_k), \nabla y)_I\\ &= -(v_k - P_k v_k, \Delta y)_I \le Ck\|\partial_t v_k\|_I\|\Delta y\|_I.\end{aligned}$$

By means of Lemma 4.2 this implies

$$\|y_k - P_k y\| \le Ck\|\Delta y\|_I,$$

and together with (5.2) we obtain the assertion by the stability estimate from Proposition 4.1. ∎

We now prove the following supercloseness result between the discrete solution y_k and the interpolation $\Pi_k y$.

LEMMA 5.3. *For the solutions $y \in X$ of (4.2) and $y_k \in \tilde{X}_k^0$ of (4.4), the following holds with the projection Π_k given by (3.11) provided that $y \in H^1(I, H^2(\Omega) \cap V) \cap H^2(I, H)$:*

$$\|y_k - \Pi_k y\|_I \le Ck^2\big\{\|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I\big\}.$$

Proof. We consider the solution $v_k \in X_k^1$ of (4.3) with $g = y_k - \Pi_k y$. Using the Galerkin orthogonality of $y - y_k$ to X_k^1, we get for $\varphi = y_k - \Pi_k y \in \tilde{X}_k^0$ since $(y - \Pi_k y)_0^- = 0$

$$\begin{aligned}\|y_k - \Pi_k y\|_I^2 &= B(v_k, y_k - \Pi_k y) = B(v_k, y - \Pi_k y)\\ &= (\partial_t v_k, y - \Pi_k y)_I + (\nabla v_k, \nabla(y - \Pi_k y))_I.\end{aligned} \tag{5.3}$$

For $(\partial_t v_k, y - \Pi_k y)_I$, Lemma 5.1 with $\psi_k = \partial_t v_k \in \tilde{X}_k^0$ implies, since $H^1(I, H^2(\Omega) \cap V) \cap H^2(I, H) \hookrightarrow C(\bar{I}, V)$, that

$$\begin{aligned}(\partial_t v_k, y - \Pi_k y)_I = \sum_{m=1}^M (\partial_t v_k, y - \Pi_k y)_{I_m} &\le C \sum_{m=1}^M k_m^{\frac{5}{2}} \|(\partial_t v_k)_m\| \|\partial_t^2 y\|_{I_m}\\ &\le C\left(\sum_{m=1}^M k_m^5 \|(\partial_t v_k)_m\|^2\right)^{\frac{1}{2}} \left(\sum_{m=1}^M \|\partial_t^2 y\|_{I_m}^2\right)^{\frac{1}{2}}\\ &\le Ck^2\|\partial_t v_k\|_I \|\partial_t^2 y\|_I.\end{aligned}$$

With the projection P_k defined by (3.10), we obtain for the second term on the right-hand side of (5.3)

$$\begin{aligned}(\nabla v_k, \nabla(y - \Pi_k y))_I &= (\nabla(v_k - P_k v_k), \nabla(y - P_k y))_I + (\nabla P_k v_k, \nabla(y - \Pi_k y))_I \\ &= -(v_k - P_k v_k, \Delta y - P_k \Delta y)_I - (P_k \Delta v_k, y - \Pi_k y)_I.\end{aligned}$$

Here, we have used the fact that the temporal projection P_k commutes with the spacial derivatives by its definition. The term $(P_k \Delta v_k, y - \Pi_k y)_I$ is treated as the first term using Lemma 5.1 with $\psi = P_k \Delta v_k \in \tilde{X}_k^0$, and the error estimate for the L^2 projection P_k yields

$$(\nabla v_k, \nabla(y - \Pi_k y))_I \leq Ck^2 \big\{ \|\partial_t v_k\|_I \|\partial_t \Delta y\|_I + \|P_k \Delta v_k\|_I \|\partial_t^2 y\|_I \big\}.$$

Inserting the estimates in (5.3) and using the assertions of Lemma 4.2, we end up with

$$\|y_k - \Pi_k y\|_I \leq Ck^2 \big\{ \|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I \big\},$$

which completes the proof. □

LEMMA 5.4. *Let $g \in H^1(I, H)$. For the solutions y_k of* (4.4) *and $\tilde{y}_k \in \tilde{X}_k^0$ of*

$$B(\varphi, \tilde{y}_k) = (\varphi, P_k g)_I \quad \forall \varphi \in X_k^1, \tag{5.4}$$

where P_k is given as in (3.10)*, it holds that*

$$\|y_k - \tilde{y}_k\|_I \leq Ck^2 \|\partial_t g\|_I.$$

Proof. To prove the assertion, we subtract (5.4) from (4.4) and choose $\varphi = w_k \in X_k^1$ such that $\partial_t w_k\big|_{I_m} = (y_k - \tilde{y}_k)\big|_{I_m}$ for $m = 1, 2, \ldots, m$ and $w_k(0) = 0$. Then, we obtain

$$\|y_k - \tilde{y}_k\|_I^2 + (\nabla w_k, \partial_t \nabla w_k)_I = (w_k, g - P_k g)_I = (w_k - P_k w_k, g - P_k g)_I.$$

It follows that

$$\|y_k - \tilde{y}_k\|_I^2 + \frac{1}{2}\|\nabla w_{k,M}\|^2 \leq \|w_k - P_k w_k\|_I \|g - P_k g\|_I \leq Ck^2 \|\partial_t w_k\|_I \|\partial_t g\|_I,$$

which yields the desired result. □

COROLLARY 5.5. *Let $g \in H^1(I, H)$. For the solutions y_{kh} of* (4.6) *and $\tilde{y}_{kh} \in \tilde{X}_{k,h}^{0,1}$ of*

$$B(\varphi, \tilde{y}_{hk}) = (\varphi, P_k g)_I \quad \forall \varphi \in X_{k,h}^{1,1}, \tag{5.5}$$

where P_k is given as in (3.10)*, it holds that*

$$\|y_{kh} - \tilde{y}_{kh}\|_I \leq Ck^2 \|\partial_t g\|_I.$$

Proof. The assertion can be proved similarly to how the corresponding assertion for $y_k - \tilde{y}_k$ in Lemma 5.4 was proved. □

LEMMA 5.6. *Let the projection π_k be given by* (3.12).

(i) *For $w \in H^2(I, \mathbb{R}^D)$, it holds that*

$$|w - \pi_k w|_I \leq Ck^2 |\partial_t^2 w|_I.$$

(ii) *For $w_k \in \tilde{Q}_k^0$, it holds that*

$$|\pi_k w_k|_I \leq C|w_k|_I.$$

Proof.

(i) For $w \in H^2(I, \mathbb{R}^D) \hookrightarrow C(\bar{I}, \mathbb{R}^D)$, the interpolation operator π_k coincides on I_m^* for $m = 3, 4, \ldots, M-1$ with the usual nodal interpolation into $\mathcal{P}_1(I_m^*, \mathbb{R}^D)$. Hence, the standard interpolation estimate

$$|w - \pi_k w|_{I_m^*} \leq C|I_m^*|^2|\partial_t^2 w|_{I_m^*}$$

holds. On the "extended" time intervals $I_1^* \cup I_2^*$ and $I_M^* \cup I_{M+1}^*$, one can show by the Bramble–Hilbert lemma and a transformation argument using the assumed regularity of the time grid that

$$|w - \pi_k w|_{I_1^* \cup I_2^*} \leq C(\kappa)|I_1^* \cup I_2^*|^2|\partial_t^2 w|_{I_1^* \cup I_2^*},$$
$$|w - \pi_k w|_{I_M^* \cup I_{M+1}^*} \leq C(\kappa)|I_M^* \cup I_{M+1}^*|^2|\partial_t^2 w|_{I_M^* \cup I_{M+1}^*}.$$

Thus, the relation $\max_{m=1,2,\ldots,M+1}|I_m^*| \leq k$ yields the first assertion.

(ii) For the second assertion, we get by definition

$$\pi_k w_k(t)\big|_{I_1^* \cup I_2^*} := w_{k,1}^- + \frac{t - t_1^*}{t_2^* - t_1^*}(w_{k,2}^- - w_{k,1}^-),$$
$$\pi_k w_k(t)\big|_{I_m^*} := w_{k,m-1}^- + \frac{t - t_{m-1}^*}{t_m^* - t_{m-1}^*}(w_{k,m}^- - w_{k,m-1}^-),$$
$$m = 3, 4, \ldots, M-1,$$
$$\pi_k w_k(t)\big|_{I_M^* \cup I_{M+1}^*} := w_{k,M-1}^- + \frac{t - t_{M-1}^*}{t_M^* - t_{M-1}^*}(w_{k,M}^- - w_{k,M-1}^-).$$

Using these identities, we obtain for $m = 2, 4, \ldots, M$

$$\begin{aligned}
|\pi_k w_k|_{I_m}^2 &\leq C \int_{I_m^*} \left\{|w_{k,m-1}^-|^2 + |w_{k,m}^-|^2\right\} dt \\
&\leq C(k_{m-1} + k_m)\left\{|w_{k,m-1}^-|^2 + |w_{k,m}^-|^2\right\} \\
&\leq C(\kappa)k_{m-1}|w_{k,m-1}^-|^2 + C(\kappa)k_m|w_{k,m}^-|^2 \\
&\leq C(\kappa)\left\{|w_k|_{I_{m-1}}^2 + |w_m|_{I_m}^2\right\}.
\end{aligned}$$

On I_1^*, we obtain

$$\begin{aligned}
|\pi_k w_k|_{I_1^*}^2 &\leq C(\kappa) \int_{I_1^*} \left\{|w_{k,1}^-|^2 + |w_{k,2}^-|^2\right\} dt \leq C(\kappa)k_1\left\{|w_{k,1}^-|^2 + |w_{k,2}^-|^2\right\} \\
&\leq C(\kappa)k_1|w_{k,1}^-|^2 + C(\kappa)k_2|w_{k,2}^-|^2 \leq C(\kappa)\left\{|w_k|_{I_1}^2 + |w_m|_{I_2}^2\right\}.
\end{aligned}$$

A similar calculation on I_{M+1}^* leads to

$$|\pi_k w_k|_{I_{M+1}^*}^2 \leq C(\kappa)\left\{|w_k|_{I_{M-1}}^2 + |w_m|_{I_M}^2\right\}.$$

Collecting all estimates yields the second assertion.

This completes the proof. ∎

5.2. Analysis of the spatial discretization error.

LEMMA 5.7. *Let $g \in H^1(I, H)$. For the solutions $y_k \in \tilde{X}_k^0$ of* (4.4) *and $y_{kh} \in \tilde{X}_{k,h}^{0,1}$ of* (4.6), *it holds that*

$$\|y_k - y_{kh}\|_I \leq C\big\{k^2\|\partial_t g\|_I + h^2\|g\|_I\big\}.$$

Proof. In addition to y_k and y_{kh}, we consider the solutions $\tilde{y}_k \in \tilde{X}_k^0$ and $\tilde{y}_{kh} \in \tilde{X}_{k,h}^{0,1}$ of (5.4) and (5.5). By the triangle inequality, we have

$$\|y_k - y_{kh}\|_I \leq \|y_k - \tilde{y}_k\|_I + \|\tilde{y}_k - \tilde{y}_{kh}\|_I + \|\tilde{y}_{kh} - y_{kh}\|_I,$$

and from Lemma 5.4 and Corollary 5.5, we obtain for the first and last terms

$$\|y_k - \tilde{y}_k\|_I \leq Ck^2\|\partial_t g\|_I \quad \text{and} \quad \|y_{kh} - \tilde{y}_{kh}\|_I \leq Ck^2\|\partial_t g\|_I.$$

It remains to estimate $\|\tilde{y}_k - \tilde{y}_{kh}\|_I$. We split by means of the Ritz projection R_h given by (3.20):

$$\tilde{y}_k - \tilde{y}_{kh} = (\tilde{y}_k - R_h\tilde{y}_k) + (R_h\tilde{y}_k - \tilde{y}_{kh}) = \eta_h + \xi_h.$$

Let $v_{kh} \in X_{k,h}^{1,1}$ be the solution of (4.5) with $g = \xi_h$. First, we observe that by choosing $\varphi \in \tilde{X}_{k,h}^{0,1}$ in (4.5) such that $\varphi\big|_{I_m} = 0$ for $m = 1, 2, \ldots, M$ and $\varphi_0^- = v_{kh,0}$, we obtain that $v_{kh,0} = 0$. By choosing $\varphi = \xi_h \in \tilde{X}_{k,h}^{0,1}$ in (4.5), applying Galerkin orthogonality of $\tilde{y}_k - \tilde{y}_{kh}$ to X_k^1, and using the properties of the Ritz projection, we obtain, since $v_{kh,0} = 0$,

$$\|\xi_h\|_I^2 = B(v_{kh}, \xi_h) = -B(v_{kh}, \eta_h) = -(\partial_t v_{kh}, \eta_h)_I.$$

By Corollary 4.3, we get

$$\|\xi_h\|_I^2 \leq \|\partial_t v_{kh}\|_I\|\eta_h\|_I \leq \|\xi_h\|_I\|\eta_h\|_I.$$

Standard estimates for the Ritz projection error η_h imply

$$\|\tilde{y}_k - \tilde{y}_{kh}\|_I \leq \|\xi_h\|_I + \|\eta_h\|_I \leq 2\|\eta_h\|_I \leq Ch^2\|\Delta\tilde{y}_k\|_I,$$

and by Lemma 4.4, we finally obtain

$$\|\tilde{y}_k - \tilde{y}_{kh}\|_I \leq Ch^2\|P_k g\|_I \leq Ch^2\|g\|_I.$$

This completes the proof. ∎

6. Error analysis for the optimal control problem. In this section we will prove error estimates for the optimal control problem under consideration and derive the main result of this article.

The analysis in this section is based on an assumption on the structure of the active sets. We split the set of indices $\mathcal{I}_k := \{\, m \mid m = 1, 2, \ldots, M \,\}$ depending on the value of $\bar{q}$ on I_m into two disjoint sets $\mathcal{I}_k = \mathcal{I}_k^1 \cup \mathcal{I}_k^2$. The sets are chosen as follows:

$$\mathcal{I}_k^1 = \left\{\, m \in \mathcal{I}_k \;\middle|\; \bar{q}_i(t) = q_i^a \text{ or } \bar{q}_i(t) = q_i^b \text{ or } q_i^a < \bar{q}_i(t) < q_i^b \;\forall t \in I_m, i = 1, 2, \ldots, D \,\right\},$$
$$\mathcal{I}_k^2 = \mathcal{I}_k \setminus \mathcal{I}_k^1.$$

Hence, the set $\mathcal{I}_k^2$ consists of the indices of the time intervals which contain, at least for one component q_i with $i \in \{1, 2, \dots, D\}$, the free boundary between the active and the inactive sets.

ASSUMPTION 2. *We assume that there exists a positive constant C independent of k such that*

$$\sum_{m \in \mathcal{I}_k^2} |I_m| \le Ck.$$

Remark 6.1. An assumption similar to this but on the spatial mesh instead of the time grid is used in [24, 29, 4, 23]. This assumption is fulfilled if the boundary of the active set consists of a finite number of points.

Using Assumption 2, we can formulate the main results of this article.

THEOREM 6.2. *Let Assumption 2 be fulfilled. For the solution $\bar{q} \in Q_{ad}$ of (2.2) and the approximation*

$$\tilde{q}_{kh} = P_{Q_{ad}}(-\alpha^{-1}\pi_k G^* z_{kh}(\bar{q}_{kh})) \tag{6.1}$$

with the solution $\bar{q}_{kh}$ of (3.15), it holds that

$$|\bar{q} - \tilde{q}_{kh}|_I = \mathcal{O}(k^2 + h^2).$$

Proof. The assertion follows directly from Theorems 6.6 and 6.10 proved in the following subsections. ∎

6.1. Estimates for the error due to time discretization of the state. As a first step in proving Theorem 6.2, we derive in this subsection an estimate of the error between the continuous solution $\bar{q} \in Q_{\mathrm{ad}}$ of (2.2) and the approximation

$$\tilde{q}_k = P_{Q_{\mathrm{ad}}}(-\alpha^{-1}\pi_k G^* z_k(\bar{q}_k)) \tag{6.2}$$

with the solution $\bar{q}_k \in Q_{\mathrm{ad}}$ of (3.5) and π_k defined in (3.12). For doing so, we first prove the following sequence of lemmas.

LEMMA 6.3. *For the solutions $u = u(q) \in X$ of the state equation (2.1) and $u_k = u_k(q) \in X_k^1$ of the semidiscrete state equation (3.4) for given control $q \in H^1(I, \mathbb{R}^D) \subset Q$, it holds that*

$$\|u - u_k\|_I \le Ck^2\big\{\|\partial_t^2 u\|_I + \|\partial_t \Delta u\|_I\big\}.$$

Proof. We define $e_k := u - u_k$ and use the splitting $e_k = \eta_k + \xi_k$ with

$$\eta_k := u - i_k u \quad \text{and} \quad \xi_k := i_k u - u_k,$$

where $i_k \colon C(\bar{I}, V) \to X_k^1$ denotes the temporal nodal interpolation defined for $m = 0, 1, \dots, M$ by $(i_k u)(t_m) := u(t_m)$. Employing the solutions $y \in X$ of (4.2) and $y_k \in \tilde{X}_k^0$ of (4.4) with $g = e_k$, we have

$$\|e_k\|_I^2 = (\eta_k, e_k)_I + (\xi_k, e_k)_I = (\eta_k, e_k)_I + B(\xi_k, y_k) = (\eta_k, e_k)_I - B(\eta_k, y_k).$$

For $-B(\eta_k, y_k)$, we obtain by integration by parts in time

$$\begin{aligned} -B(\eta_k, y_k) &= -(\partial_t \eta_k, y_k)_I - (\nabla \eta_k, \nabla y_k)_I - (\eta_{k,0}, y_{k,0}^-) \\ &= -(\nabla \eta_k, \nabla y_k)_I + \sum_{m=0}^{M-1} (\eta_{k,m}, [y_k]_m) + (\eta_{k,M}, y_{k,M}^-). \end{aligned}$$

By the definition of i_k, the estimate for $\|y - y_k\|_I$ from Lemma 5.2, and standard interpolation estimates for $\|\eta_k\|_I$ and $\|\Delta\eta_k\|_I$, we get

$$\begin{aligned}-B(\eta_k, y_k) &= -(\nabla\eta_k, \nabla y_k)_I = (\Delta\eta_k, y_k - y)_I + (\eta_k, \Delta y)_I \\ &\le Ck^2\big\{\|\partial_t\Delta u\|_I + \|\partial_t^2 u\|_I\big\}\big\{\|\partial_t y\|_I + \|\Delta y\|_I\big\},\end{aligned}$$

where we used the regularity $u \in H^1(I, H^2(\Omega) \cap V) \cap H^1(I, H)$ ensured by Proposition 2.1 and $q \in H^1(I, \mathbb{R}^D)$. Then, the stability estimates for y from Proposition 4.1 imply

$$-B(\eta_k, y_k) \le Ck^2\big\{\|\partial_t\Delta u\|_I + \|\partial_t^2 u\|_I\big\}\|e_k\|_I,$$

and the estimate

$$(\eta_k, e_k)_I \le Ck^2\|\partial_t^2 u\|_I\|e_k\|_I$$

for the remaining term implies the assertion. ∎

Lemma 6.4. *Let $\bar q$ be the optimal solution of* (2.2) *and $\psi_k \in \tilde X_k^0$. Then, under Assumption* 2, *it holds for the projection Π_k given by* (3.11) *that*

$$(\psi_k, G\bar q - \Pi_k G\bar q)_I \le \frac{C}{\alpha}k^2 \max_{m=1,2,\dots,M}\|\psi_k\|_{L^\infty(I_m,H)}\|\partial_t^2 z(\bar q)\|_I.$$

Proof. The optimal control $\bar q$ fulfills

$$\bar q = P_{Q_{\mathrm{ad}}}(\alpha^{-1}G^* z(\bar q)).$$

Hence, for $m \in \mathcal{I}_k^1$ we have for $i = 1, 2, \dots, D$ that $\bar q_i\big|_{I_m} = q_i^a$, $\bar q_i\big|_{I_m} = q_i^b$, or $\bar q_i\big|_{I_m} = \alpha^{-1}(G^* z(\bar q))_i$. This implies by the regularity of $z(\bar q)$ from Proposition 2.3 that $\bar q\big|_{I_m} \in H^2(I_m, \mathbb{R}^D)$ for $m \in \mathcal{I}_k^1$. Since $z(\bar q) \in H^2(I, H) \hookrightarrow W^{1,\infty}(I, H)$, we have on I_m for $m \in \mathcal{I}_k^2$ due to the stability of the projection $P_{Q_{\mathrm{ad}}}$ that $\bar q\big|_{I_m} \in W^{1,\infty}(I_m, \mathbb{R}^D)$.

We split

$$(\psi_k, G\bar q - \Pi_k G\bar q)_I = \sum_{m\in\mathcal{I}_k^1}(\psi_k, G\bar q - \Pi_k G\bar q)_{I_m} + \sum_{m\in\mathcal{I}_k^2}(\psi_k, G\bar q - \Pi_k G\bar q)_{I_m}.$$

For the first sum, we have from Lemma 5.1 with $w = G\bar q \in H^2(I, H)$ that

$$\begin{aligned}\sum_{m\in\mathcal{I}_k^1}(\psi_k, G\bar q - \Pi_k G\bar q)_{I_m} &\le C\sum_{m\in\mathcal{I}_k^1} k_m^{\frac{5}{2}}\|\psi_{k,m}\|\|\partial_t^2 G\bar q\|_{I_m} \\ &\le \frac{C}{\alpha}\Bigg(\sum_{m\in\mathcal{I}_k^1} k_m^5\|\psi_{k,m}\|^2\Bigg)^{\frac12}\Bigg(\sum_{m\in\mathcal{I}_k^1}\|\partial_t^2 z(\bar q)\|_I^2\Bigg)^{\frac12} \\ &\le \frac{C}{\alpha}k^2\|\psi_k\|_I\|\partial_t^2 z(\bar q)\|_I \\ &\le \frac{C}{\alpha}k^2 \max_{m=1,2,\dots,M}\|\psi_k\|_{L^\infty(I_m,H)}\|\partial_t^2 z(\bar q)\|_I.\end{aligned} \tag{6.3}$$

For the second sum, we proceed by means of the projection P_k given by (3.10):

$$\begin{aligned}\sum_{m\in\mathcal{I}_k^2}(\psi_k, G\bar q - \Pi_k G\bar q)_{I_m} &= \sum_{m\in\mathcal{I}_k^2}(\psi_k, P_k G\bar q - \Pi_k G\bar q)_{I_m} \\ &\le C\sum_{m\in\mathcal{I}_k^2}|I_m|\|\psi_{k,m}\|\|(P_k G\bar q)_m - (\Pi_k G\bar q)_m\|.\end{aligned}$$

For $\|(P_k G\bar{q})_m - (\Pi_k G\bar{q})_m\|$, we have

$$\begin{aligned}\|(P_k G\bar{q})_m - (\Pi_k G\bar{q})_m\| &= \left\| \frac{1}{k_m} \int_{I_m} \big(G\bar{q}(t) - G\bar{q}(t_m)\big)\, dt \right\| \\ &\le \frac{1}{k_m} \int_{I_m} \int_{t_m}^{t} \|\partial_t G\bar{q}(s)\|\, ds\, dt \\ &\le k_m \|\partial_t G\bar{q}\|_{L^\infty(I_m,H)}.\end{aligned}$$

By Assumption 2, the $W^{1,\infty}$-stability of $P_{Q_{\mathrm{ad}}}$ from (2.9), and the embedding $H^2(I,H) \hookrightarrow W^{1,\infty}(I,H)$, it follows that

$$\begin{aligned}\sum_{m\in\mathcal{I}_k^2} (\psi_k, G\bar{q} - \Pi_k G\bar{q})_{I_m} &\le Ck \sum_{m\in\mathcal{I}_k^2} |I_m| \|\psi_k\|_{L^\infty(I_m,H)} \|\partial_t G\bar{q}\|_{L^\infty(I_m,H)} \\ &\le \frac{C}{\alpha} k^2 \max_{m=1,2,\dots,M} \|\psi_k\|_{L^\infty(I_m,H)} \|\partial_t z(\bar{q})\|_{L^\infty(I,H)} \qquad (6.4)\\ &\le \frac{C}{\alpha} k^2 \max_{m=1,2,\dots,M} \|\psi_k\|_{L^\infty(I_m,H)} \|\partial_t^2 z(\bar{q})\|_I.\end{aligned}$$

Combination of the estimates (6.3) and (6.4) implies the assertion. ∎

LEMMA 6.5. *Let Assumption 2 be fulfilled. For the solutions $\bar{q} \in Q_{ad}$ of (2.2) and $\bar{q}_k \in Q_{ad}$ of (3.5) and the projection Π_k given in (3.11), it holds that*

$$\begin{aligned}&\alpha|\bar{q}_k - \Pi_k\bar{q}|_I^2 + \|u_k(\bar{q}_k) - u_k(\Pi_k\bar{q})\|_I^2 \\ &\quad\le \frac{C}{\alpha} k^4 \Big\{ \|\partial_t^2 u(\bar{q})\|_I^2 + \|\partial_t \Delta u(\bar{q})\|_I^2 + \Big(1 + \frac{1}{\alpha^2}\Big) \|\partial_t^2 z(\bar{q})\|_I^2 + \|\partial_t \Delta z(\bar{q})\|_I^2 \Big\}.\end{aligned}$$

Proof. The optimality conditions (2.4) and (3.6) for $\bar{q}$ and $\bar{q}_k$ yield

$$\langle \alpha\bar{q} + G^* z(\bar{q}), \delta q - \bar{q}\rangle_I \ge 0 \quad \forall \delta q \in Q_{\mathrm{ad}}, \qquad (6.5)$$

$$\langle \alpha\bar{q}_k + G^* z_k(\bar{q}_k), \delta q - \bar{q}_k\rangle_I \ge 0 \quad \forall \delta q \in Q_{\mathrm{ad}}. \qquad (6.6)$$

A pointwise discussion of (6.5) implies

$$\langle \alpha\Pi_k\bar{q} + \Pi_k G^* z(\bar{q}), \Pi_k\delta q - \Pi_k\bar{q}\rangle_I \ge 0 \quad \forall \delta q \in Q_{\mathrm{ad}}. \qquad (6.7)$$

For the choices $\delta q = \bar{q}_k$ in (6.7) and $\delta q = \Pi_k\bar{q}$ in (6.6), we have

$$\langle \alpha\Pi_k\bar{q} + \Pi_k G^* z(\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \ge 0 \quad \text{and} \quad -\langle \alpha\bar{q}_k + G^* z_k(\bar{q}_k), \bar{q}_k - \Pi_k\bar{q}\rangle_I \ge 0$$

since $\Pi_k\bar{q}_k = \bar{q}_k$. By adding these identities, we get by means of the semidiscrete state and adjoint equations (3.4) and (3.8)

$$\begin{aligned}\alpha|\bar{q}_k - \Pi_k\bar{q}|_I^2 &\le \langle \Pi_k G^* z(\bar{q}) - G^* z_k(\bar{q}_k), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &= \langle \Pi_k G^* z(\bar{q}) - G^* z_k(\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &\quad + \langle G^* z_k(\bar{q}) - G^* z_k(\Pi_k\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &\quad + (z_k(\Pi_k\bar{q}) - z_k(\bar{q}_k), G\bar{q}_k - G\Pi_k\bar{q})_I \\ &= \langle \Pi_k G^* z(\bar{q}) - G^* z_k(\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &\quad + \langle G^* z_k(\bar{q}) - G^* z_k(\Pi_k\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &\quad - B(u_k(\bar{q}_k) - u_k(\Pi_k\bar{q}), z_k(\bar{q}_k) - z_k(\Pi_k\bar{q})) \\ &= \langle \Pi_k G^* z(\bar{q}) - G^* z_k(\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &\quad + \langle G^* z_k(\bar{q}) - G^* z_k(\Pi_k\bar{q}), \bar{q}_k - \Pi_k\bar{q}\rangle_I \\ &\quad - \|u_k(\bar{q}_k) - u_k(\Pi_k\bar{q})\|_I^2.\end{aligned}$$

By definition (3.11) of the projection Π_k, we have $\Pi_k G^* z(\bar{q}) = G^* \Pi_k z(\bar{q})$, where Π_k on the left-hand side acts on $C(\bar{I}, \mathbb{R}^D)$ and on the right-hand side acts on $C(\bar{I}, V)$; cf. (3.11). Hence, we obtain by the continuity of G^*

$$\text{(6.8)} \quad \alpha|\bar{q}_k - \Pi_k \bar{q}|_I^2 + \|u_k(\bar{q}_k) - u_k(\Pi_k \bar{q})\|_I^2 \leq \frac{C}{\alpha}\|\Pi_k z(\bar{q}) - z_k(\bar{q})\|_I^2 + \frac{C}{\alpha}\|z_k(\bar{q}) - z_k(\Pi_k \bar{q})\|_I^2.$$

For the first term on the right-hand side of (6.8), we have

$$\|\Pi_k z(\bar{q}) - z_k(\bar{q})\|_I^2 \leq C\big\{\|\Pi_k z(\bar{q}) - \tilde{z}_k\|_I^2 + \|\tilde{z}_k - z_k(\bar{q})\|_I^2\big\}$$

with $\tilde{z}_k \in \tilde{X}_k^0$ solving

$$B(\varphi, \tilde{z}_k) = (u(\bar{q}) - \hat{u}, \varphi)_I \quad \forall \varphi \in X_k^1.$$

From Lemma 5.3, we have

$$\|\Pi_k z(\bar{q}) - \tilde{z}_k\|_I \leq Ck^2\big\{\|\partial_t^2 z(\bar{q})\|_I + \|\partial_t \Delta z(\bar{q})\|_I\big\}.$$

The stability of z_k and $\tilde{z}_k$ from Lemma 4.4 and the estimate from Lemma 6.3 yields

$$\|\tilde{z}_k - z_k(\bar{q})\|_I \leq C\|u(\bar{q}) - u_k(\bar{q})\|_I \leq Ck^2\big\{\|\partial_t^2 u(\bar{q})\|_I + \|\partial_t \Delta u(\bar{q})\|_I\big\},$$

where we used that $\bar{q} \in H^1(I, \mathbb{R}^D)$ which is ensured by Proposition 2.3. Altogether, we have for the first term

$$\|\Pi_k z(\bar{q}) - z_k(\bar{q})\|_I^2 \leq Ck^4\big\{\|\partial_t^2 u(\bar{q})\|_I^2 + \|\partial_t \Delta u(\bar{q})\|_I^2 + \|\partial_t^2 z(\bar{q})\|_I^2 + \|\partial_t \Delta z(\bar{q})\|_I^2\big\}.$$

Due to the stability estimate for z_k from Lemma 4.4, the second term on the right-hand side of (6.8) is bounded by

$$\|z_k(\bar{q}) - z_k(\Pi_k \bar{q})\|_I^2 \leq C\|u_k(\bar{q}) - u_k(\Pi_k \bar{q})\|_I^2.$$

Further, we have by means of the semidiscrete state and adjoint equations (3.4) and (3.8) using the relation $G\Pi_k = \Pi_k G$

$$\begin{aligned}\|u_k(\bar{q}) - u_k(\Pi_k \bar{q})\|_I^2 &= (z_k(\bar{q}) - z_k(\Pi_k \bar{q}), G\bar{q} - G\Pi_k \bar{q})_I \\ &= (z_k(\bar{q}) - z_k(\Pi_k \bar{q}), G\bar{q} - \Pi_k G\bar{q})_I.\end{aligned}$$

With $\psi_k = z_k(\bar{q}) - z_k(\Pi_k \bar{q}) \in \tilde{X}_k^0$ in Lemma 6.4, we have

$$(z_k(\bar{q}) - z_k(\Pi_k \bar{q}), G\bar{q} - \Pi_k G\bar{q})_I \leq \frac{C}{\alpha}k^2 \max_{m=1,2,\dots,M} \|z_k(\bar{q}) - z_k(\Pi_k \bar{q})\|_{L^\infty(I_m,H)} \|\partial_t^2 z(\bar{q})\|_I,$$

and the stability estimate from Lemma 4.6,

$$\max_{m=1,2,\dots,M} \|z_k(\bar{q}) - z_k(\Pi_k \bar{q})\|_{L^\infty(I_m,H)} \leq C\|u_k(\bar{q}) - u_k(\Pi_k \bar{q})\|_I,$$

yields

$$\text{(6.9)} \qquad \|u_k(\bar{q}) - u_k(\Pi_k \bar{q})\|_I \leq \frac{C}{\alpha}k^2\|\partial_t^2 z(\bar{q})\|_I,$$

and consequently,

$$\|z_k(\bar q) - z_k(\Pi_k \bar q)\|_I^2 \le \frac{C}{\alpha^2} k^4 \|\partial_t^2 z(\bar q)\|_I^2,$$

which completes the proof. ∎

After these preparations, we are able to prove the main result of this subsection.

THEOREM 6.6. *Let Assumption 2 be fulfilled. For the solution $\bar q \in Q_{ad}$ of* (2.2) *and $\tilde q_k$ defined by* (6.2) *with the solution $\bar q_k \in Q_{ad}$ of* (3.5)*, it holds that*

$$|\bar q - \tilde q_k|_I \le C(\alpha) k^2 \big\{ \|f\|_{H^1(I,H)} + \|q\|_{H^1(I,\mathbb{R}^D)} + \|\hat u\|_{H^1(I,H)} + \|\nabla f(0)\| + \|\nabla \hat u(T)\| + \|\nabla \Delta u_0\| \big\}.$$

Proof. We have

$$|\bar q - \tilde q_k|_I = |P_{Q_{\mathrm{ad}}}(-\alpha^{-1} G^* z(\bar q)) - P_{Q_{\mathrm{ad}}}(-\alpha^{-1} \pi_k G^* z_k(\bar q_k))|_I \le \frac{C}{\alpha} |G^* z(\bar q) - \pi_k G^* z_k(\bar q_k)|_I.$$

By the identity $\pi_k \Pi_k = \pi_k$ for Π_k given by (3.11) and Lemma 5.6 with $w = G^* z(\bar q) \in H^2(I, \mathbb{R}^D)$ and $w_k = \Pi_k G^* z(\bar q) - G^* z_k(\bar q_k) \in \tilde Q_k^0$, we get

$$\begin{aligned} |G^* z(\bar q) - \pi_k G^* z_k(\bar q_k)|_I &\le |G^* z(\bar q) - \pi_k \Pi_k G^* z(\bar q)|_I + |\pi_k(\Pi_k G^* z(\bar q) - G^* z_k(\bar q_k))|_I \\ &\le |G^* z(\bar q) - \pi_k G^* z(\bar q)|_I + C|\Pi_k G^* z(\bar q) - G^* z_k(\bar q_k)|_I \\ &\le C k^2 \|\partial_t^2 z(\bar q)\|_I + C|\Pi_k G^* z(\bar q) - G^* z_k(\bar q_k)|_I. \end{aligned}$$

For $|\Pi_k G^* z(\bar q) - G^* z_k(\bar q_k)|_I$, we proceed using the the relation $\Pi_k G^* = G^* \Pi_k$ and the continuity of G^* to obtain

$$|\Pi_k G^* z(\bar q) - G^* z_k(\bar q_k)|_I \le C \|\Pi_k z(\bar q) - z_k(\bar q_k)\|_I.$$

Using the solution $\tilde z_k \in \tilde X_k^0$ of

$$B(\varphi, \tilde z_k) = (\varphi, u(\bar q) - \hat u)_I \quad \forall \varphi \in X_k^1,$$

we get by means of Lemma 5.3 and the stability of z_k and $\tilde z_k$

$$\begin{aligned} \|\Pi_k z(\bar q) - z_k(\bar q_k)\|_I &\le \|\Pi_k z(\bar q) - \tilde z_k\|_I + \|\tilde z_k - z_k(\bar q_k)\|_I \\ &\le C k^2 \big\{ \|\partial_t^2 z(\bar q)\|_I + \|\partial_t \Delta z(\bar q)\|_I \big\} + \|u(\bar q) - u_k(\bar q_k)\|_I. \end{aligned}$$

To estimate $\|u(\bar q) - u_k(\bar q_k)\|_I$, we proceed as follows by means of Lemma 6.3, (6.9), and Lemma 6.5:

$$\begin{aligned} \|u(\bar q) - u_k(\bar q_k)\|_I &\le \|u(\bar q) - u_k(\bar q)\|_I + \|u_k(\bar q) - u_k(\Pi_k \bar q)\|_I + \|u_k(\Pi_k \bar q) - u_k(\bar q_k)\|_I \\ &\le C(\alpha) k^2 \big\{ \|\partial_t^2 u(\bar q)\|_I + \|\partial_t \Delta u(\bar q)\|_I + \|\partial_t^2 z(\bar q)\|_I + \|\partial_t \Delta z(\bar q)\|_I \big\}, \end{aligned}$$

where we used that $\bar q \in H^1(I, \mathbb{R}^D)$ which is ensured by Proposition 2.3. Collecting all estimates yields

$$|\bar q - \tilde q_k|_I \le C(\alpha) k^2 \big\{ \|\partial_t^2 u(\bar q)\|_I + \|\partial_t \Delta u(\bar q)\|_I + \|\partial_t^2 z(\bar q)\|_I + \|\partial_t \Delta z(\bar q)\|_I \big\},$$

and the stability estimate from Proposition 2.3 proves the desired result. ∎

6.2. Estimates for the error due to space discretization of the state. In this subsection, we prove the remaining part of the estimate from Theorem 6.2. That is, we derive an estimate for the error between $\tilde{q}_k$ defined by (6.2) and $\tilde{q}_{kh}$ defined by (6.1). Before proving this result, we derive the following sequence of lemmas.

LEMMA 6.7. *For the solutions $u_k = u_k(q) \in X_k^1$ of (3.4) and $u_{kh} = u_{kh}(q) \in X_{k,h}^{1,1}$ of (3.14) for given control $q \in Q$ and the projection P_k given by (3.10), it holds that*

$$\|P_k(u_k - u_{kh})\|_I \le Ch^2\big\{\|f\|_I + |q|_I\big\}.$$

Proof. We consider the solutions $y_k \in \tilde{X}_k^0$ of (4.4) and $y_{kh} \in \tilde{X}_{k,h}^{0,1}$ of (4.6) with $g = P_k(u_k - u_{kh}) \in \tilde{X}_k^0$. Then, we have for $\varphi = u_k$ in (4.4) and $\varphi = u_{kh}$ in (4.6) employing (3.4) and (3.14)

$$\begin{aligned}
\|P_k(u_k - u_{kh})\|_I^2 &= (P_k(u_k - u_{kh}), P_k(u_k - u_{kh}))_I \\
&= (u_k, P_k(u_k - u_{kh}))_I - (u_{kh}, P_k(u_k - u_{kh}))_I \\
&= B(u_k, y_k) - B(u_{kh}, y_{kh}) \\
&= (f + Gq, y_k)_I - (f + Gq, y_{kh})_I \\
&\le \|f + Gq\|_I \|y_k - y_{kh}\|_I.
\end{aligned}$$

As in the proof of Lemma 5.7 for $\|\tilde{y}_k - \tilde{y}_{kh}\|_I$, we have

$$\|y_k - y_{kh}\|_I \le Ch^2 \|P_k(u_k - u_{kh})\|_I,$$

which implies the assertion. ∎

LEMMA 6.8. *For the solutions $z_k = z_k(q) \in \tilde{X}_k^0$ of (3.8) and $z_{kh} = z_{kh}(q) \in \tilde{X}_{k,h}^{0,1}$ of (3.18) for given $q \in Q$, it holds that*

$$\|z_k - z_{kh}\|_I \le C\big\{k^2 + h^2\big\}\big\{\|f\|_I + |q|_I + \|\nabla u_0\| + \|u_0\|\big\} + k^2\|\partial_t \hat{u}\|_I + h^2\|\hat{u}\|_I.$$

Proof. We introduce $\tilde{z}_{kh} \in \tilde{X}_{k,h}^{0,1}$ as the solution of

$$B(\varphi, \tilde{z}_{kh}) = (\varphi, u_k(q) - \hat{u})_I \quad \forall \varphi \in X_{k,h}^{1,1}$$

and split the error as

$$\|z_k(q) - z_{kh}(q)\|_I \le \|z_k(q) - \tilde{z}_{kh}\|_I + \|\tilde{z}_{kh} - z_{kh}(q)\|_I. \tag{6.10}$$

For the first term on the right-hand side of (6.10), we have by Lemma 5.7

$$\|z_k(q) - \tilde{z}_{kh}\|_I \le C\big\{k^2\|\partial_t(u_k(q) - \hat{u})\|_I + h^2\|u_k(q) - \hat{u}\|\big\},$$

and by Lemma 4.7

$$\|z_k(q) - \tilde{z}_{kh}\|_I \le C\big\{k^2 + h^2\big\}\big\{\|f\|_I + |q|_I + \|\nabla u_0\| + \|u_0\|\big\} + k^2\|\partial_t \hat{u}\|_I + h^2\|\hat{u}\|_I.$$

To estimate the second term on the right-hand side of (6.10), we consider the solution $y_{kh} \in \tilde{X}_{k,h}^{0,1}$ of (4.6) with $g = P_k(u_k(q) - u_{kh}(q))$, where the projection P_k is given as in (3.10). We have by Corollary 5.5 and Lemma 4.4

$$\begin{aligned}
\|\tilde{z}_{kh} - z_{kh}(q) - y_{kh}\|_I &\le Ck^2\|\partial_t(u_k(q) - u_{kh}(q))\|_I, \\
\|y_{kh}\|_I &\le C\|P_k(u_k(q) - u_{kh}(q))\|_I.
\end{aligned}$$

Then, Lemmas 6.7 and 4.2 and Corollary 4.3 imply

$$\begin{aligned}\|\tilde{z}_{kh} - z_{kh}(q)\|_I &\leq \|\tilde{z}_{kh} - z_{kh}(q) - y_{kh}\|_I + \|y_{kh}\|_I \\ &\leq Ck^2\|\partial_t(u_k(q) - u_{kh}(q))\|_I + C\|P_k(u_k(q) - u_{kh}(q))\|_I \\ &\leq Ck^2\big\{\|\partial_t u_k(q)\|_I + \|\partial_t u_{kh}(q)\|_I\big\} + Ch^2\big\{\|f\|_I + |q|_I\big\} \\ &\leq C\big\{k^2 + h^2\big\}\big\{\|f\|_I + |q|_I\big\}.\end{aligned}$$

This completes the proof. ∎

LEMMA 6.9. *For the solutions* $\bar{q}_k \in Q_{ad}$ *of* (3.5) *and* $\bar{q}_{kh} \in Q_{ad}$ *of* (3.15), *it holds that*

$$\begin{aligned}&\alpha|\bar{q}_k - \bar{q}_{kh}|_I^2 + \|u_{kh}(\bar{q}_k) - u_{kh}(\bar{q}_{kh})\|_I^2 \\ &\qquad \leq \frac{C}{\alpha}\big\{k^4 + h^4\big\}\big\{\|f\|_I^2 + |\bar{q}_k|_I^2 + \|\nabla u_0\|^2 + \|u_0\|^2\big\} + k^4\|\partial_t \hat{u}\|_I^2 + h^4\|\hat{u}\|_I^2.\end{aligned}$$

Proof. The optimality conditions (3.6) and (3.16) for $\bar{q}_k$ and $\bar{q}_{kh}$ yield

$$\langle \alpha\bar{q}_k + G^* z_k(\bar{q}_k), \delta q - \bar{q}_k\rangle_I \geq 0 \quad \forall \delta q \in Q_{\mathrm{ad}}, \tag{6.11}$$

$$\langle \alpha\bar{q}_{kh} + G^* z_{kh}(\bar{q}_{kh}), \delta q - \bar{q}_{kh}\rangle_I \geq 0 \quad \forall \delta q \in Q_{\mathrm{ad}}. \tag{6.12}$$

For the choices $\delta q = \bar{q}_{kh}$ in (6.11) and $\delta q = \bar{q}_k$ in (6.12), we have

$$\langle \alpha\bar{q}_k + G^* z_k(\bar{q}_k), \bar{q}_{kh} - \bar{q}_k\rangle_I \geq 0 \quad \text{and} \quad -\langle \alpha\bar{q}_{kh} + G^* z_{kh}(\bar{q}_{kh}), \bar{q}_{kh} - \bar{q}_k\rangle_I \geq 0.$$

By adding these identities, we get by means of the discrete state and adjoint equations (3.14) and (3.18)

$$\begin{aligned}\alpha|\bar{q}_k - \bar{q}_{kh}|_I^2 &\leq \langle G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_{kh}), \bar{q}_{kh} - \bar{q}_k\rangle_I \\ &= \langle G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_k), \bar{q}_{kh} - \bar{q}_k\rangle_I \\ &\quad + (z_{kh}(\bar{q}_k) - z_{kh}(\bar{q}_{kh}), G\bar{q}_{kh} - G\bar{q}_k)_I \\ &= \langle G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_k), \bar{q}_{kh} - \bar{q}_k\rangle_I \\ &\quad - B(u_{kh}(\bar{q}_{kh}) - u_{kh}(\bar{q}_k), z_{kh}(\bar{q}_{kh}) - z_{kh}(\bar{q}_k)) \\ &= \langle G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_k), \bar{q}_{kh} - \bar{q}_k\rangle_I - \|u_{kh}(\bar{q}_{kh}) - u_{kh}(\bar{q}_k)\|_I^2.\end{aligned}$$

Hence, we have by the continuity of G^*

$$\alpha|\bar{q}_k - \bar{q}_{kh}|_I^2 + \|u_{kh}(\bar{q}_k) - u_{kh}(\bar{q}_{kh})\|_I^2 \leq C\|z_k(\bar{q}_k) - z_{kh}(\bar{q}_k)\|_I|\bar{q}_k - \bar{q}_{kh}|_I,$$

and Lemma 6.8 implies the assertion. ∎

After these preparations, we are able to prove the main result of this subsection for the difference between $\tilde{q}_k$ and $\tilde{q}_{kh}$.

THEOREM 6.10. *For* $\tilde{q}_k$ *defined by* (6.2) *with the solution* $\bar{q}_k \in Q_{ad}$ *of* (3.5) *and* $\tilde{q}_{kh}$ *defined by* (6.1) *with the solution* $\bar{q}_{kh}$ *of* (3.15), *it holds that*

$$\begin{aligned}|\tilde{q}_k - \tilde{q}_{kh}|_I \leq C(\alpha)\big\{k^2 + h^2\big\}\big\{\|f\|_I + |\bar{q}_k|_I + \|\nabla u_0\| + \|u_0\|\big\}& \\ + k^2\|\partial_t \hat{u}\|_I + h^2\|\hat{u}\|_I&.\end{aligned}$$

Proof. We have

$$\begin{aligned}|\tilde{q}_k - \tilde{q}_{kh}|_I = |P_{Q_{\mathrm{ad}}}(-\alpha^{-1}\pi_k G^* z_k(\bar{q}_k)) - P_{Q_{\mathrm{ad}}}(-\alpha^{-1}\pi_k G^* z_{kh}(\bar{q}_{kh}))|_I& \\ \leq \frac{C}{\alpha}|\pi_k(G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_{kh}))|_I&,\end{aligned}$$

and Lemma 5.6(ii) with $w_k = G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_{kh}) \in \tilde{Q}_k^0$ and the continuity of G^* imply

$$|\tilde{q}_k - \tilde{q}_{kh}|_I \leq \frac{C}{\alpha} |G^* z_k(\bar{q}_k) - G^* z_{kh}(\bar{q}_{kh})|_I \leq \frac{C}{\alpha} \|z_k(\bar{q}_k) - z_{kh}(\bar{q}_{kh})\|_I.$$

Hence, we have to estimate $\|z_k(\bar{q}_k) - z_{kh}(\bar{q}_{kh})\|_I$. By inserting $z_{kh}(\bar{q}_k)$, we get

$$\|z_k(\bar{q}_k) - z_{kh}(\bar{q}_{kh})\|_I \leq \|z_k(\bar{q}_k) - z_{kh}(\bar{q}_k)\|_I + \|z_{kh}(\bar{q}_k) - z_{kh}(\bar{q}_{kh})\|_I. \tag{6.13}$$

For the first term on the right-hand side of (6.13), Lemma 6.8 yields

$$\|z_k(\bar{q}_k) - z_{kh}(\bar{q}_k)\|_I \leq C\{k^2 + h^2\}\{\|f\|_I + |\bar{q}_k|_I + \|\nabla u_0\| + \|u_0\|\} \\ + k^2 \|\partial_t \hat{u}\|_I + h^2 \|\hat{u}\|_I.$$

For the second term on the right-hand side of (6.13), we obtain by means of Corollaries 4.5 and 4.8

$$\|z_{kh}(\bar{q}_k) - z_{kh}(\bar{q}_{kh})\|_I \leq C \|u_{kh}(\bar{q}_k) - u_{kh}(\bar{q}_{kh})\|_I \leq C \|\bar{q}_k - \bar{q}_{kh}\|_I.$$

Then, the stated estimate follows by the assertion of Lemma 6.9. □

The assertion of Theorem 6.2 is then implied by Theorems 6.6 and 6.10.

7. Numerical results. In this section, we are going to validate the a priori error estimates for the error in the control. To this end, we consider the following concretion of the optimal control problem (2.2) with known exact solution on $\Omega \times I = (0,1)^2 \times (0, 0.1)$. The operator $G\colon \mathbb{R}^D \to H$ with $D = 1$ mapping the control to the $L^2(I, V)$ is given by means of the function

$$g_1(x_1, x_2) \coloneqq \sin(\pi x_1) \sin(\pi x_2) \in V$$

by $(Gq)(t, x_1, x_1) \coloneqq q(t) \cdot g_1(x_1, x_2)$. Then, the adjoint operator $G^*\colon H \to \mathbb{R}$ is

$$(G^* z)(t) = \int_\Omega z(t, x_1, x_2) \cdot g(x_1, x_2)\, dx_1\, dx_2.$$

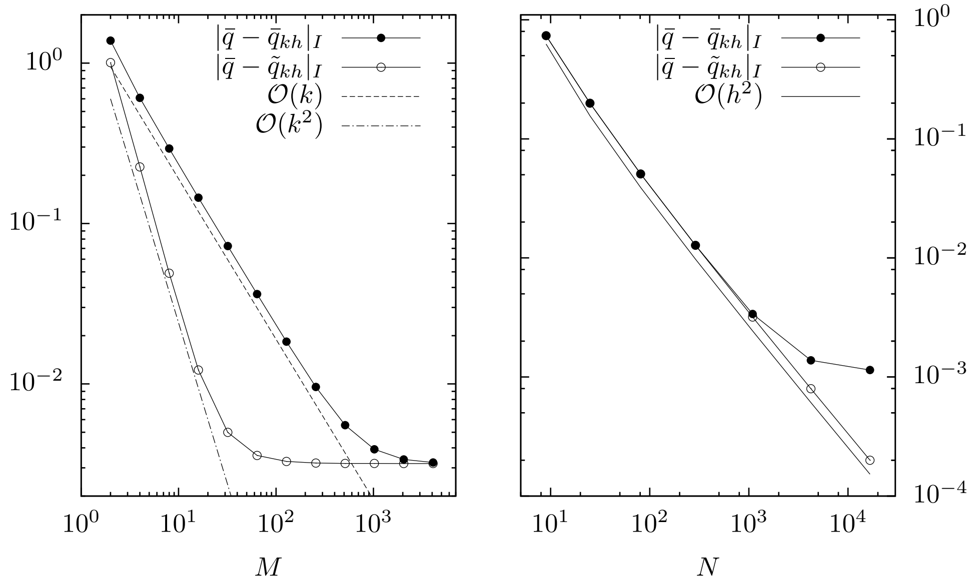
The right-hand side f, the desired state $\hat{u}$, and the initial condition u_0 are given in terms of the eigenfunctions

$$w_a(t, x_1, x_2) \coloneqq \exp(a\pi^2 t) \sin(\pi x_1) \sin(\pi x_2), \quad a \in \mathbb{R},$$

of the operator $\pm\partial_t - \Delta$ as

$$\begin{aligned} f(t, x_1, x_2) &\coloneqq -\pi^4 w_a(t, x_1, x_2) - GP_{Q_{\text{ad}}}\Big(-\frac{\pi^4}{4}\{\exp(a\pi^2 t) - \exp(a\pi^2 T)\}\Big), \\ \hat{u}(t, x_1, x_2) &\coloneqq \frac{a^2 - 5}{2 + a} \pi^2 w_a(t, x_1, x_2) + 2\pi^2 w_a(T, x_1, x_2), \\ u_0(x_1, x_2) &\coloneqq \frac{-1}{2 + a} \pi^2 w_a(0, x_1, x_2), \end{aligned}$$

with $P_{Q_{\text{ad}}}$ given as in (2.7) with $q_a = -70$ and $q_b = -1$. For this choice of data and with the regularization parameter α chosen as $\alpha = \pi^{-4}$, the optimal solution triple



(a) Refinement of the time steps for $N = 1089$ spatial nodes.

(b) Refinement of the spatial triangulation for $M = 2048$ time steps.

FIG. 7.1. *Discretization errors* $|\bar{q}-\bar{q}_{kh}|_I$ *and* $|\bar{q}-\tilde{q}_{kh}|_I$.

$(\bar{q}, \bar{u}, \bar{z})$ of the optimal control problem (2.2) is given by

$$\begin{aligned}\bar{q}(t) &:= P_{Q_{\mathrm{ad}}}\Big(-\frac{\pi^4}{4}\big\{\exp(a\pi^2 t) - \exp(a\pi^2 T)\big\}\Big),\\ \bar{u}(t,x_1,x_2) &:= \frac{-1}{2+a}\pi^2 w_a(t,x_1,x_2),\\ \bar{z}(t,x_1,x_2) &:= w_a(t,x_1,x_2) - w_a(T,x_1,x_2).\end{aligned}$$

We are going to validate the estimates developed in the previous section by separating the discretization errors. That is, we consider first the behavior of the error for a sequence of discretizations with decreasing size of the time steps and a fixed spatial triangulation with $N = 1089$ nodes. Second, we examine the behavior of the error under refinement of the spatial triangulation for $M = 2048$ time steps. For the following computations, we choose the free parameter a to be $-\sqrt{5}$.

The optimal control problems are solved by the finite element toolkit GASCOIGNE [13] and the optimization library RoDoBo [27] using a primal-dual active set strategy (cf. [5, 14]) in combination with a conjugate gradient method applied to the reduced problem (3.15).

Figure 7.1(a) depicts the development of the errors $|\bar{q}-\bar{q}_{kh}|_I$ and $|\bar{q}-\tilde{q}_{kh}|_I$ under refinement of the temporal step size k. Up to the spatial discretization error, we observe the convergence orders $\mathcal{O}(k)$ and $\mathcal{O}(k^2)$, respectively. For $|\bar{q}-\tilde{q}_{kh}|_I$, this discretization error is already reached at 64 time steps, whereas for $|\bar{q}-\bar{q}_{kh}|_I$, the number of time steps could be increased up to $M = 4096$ until reaching the spatial accuracy. The first order convergence of $|\bar{q}-\bar{q}_{kh}|_I$ corresponds to our expectations, since the ansatz functions for the temporal discretization of the adjoint equation are piecewise constant; cf. the discussion in the introduction. The second order convergence for the error $|\bar{q} - \tilde{q}_{kh}|_I$ with respect to k illustrates the convergence results from section 6.1 with respect to the *temporal* discretization.

In Figure 7.1(b) the development of the errors $|\bar{q} - \bar{q}_{kh}|_I$ and $|\bar{q} - \tilde{q}_{kh}|_I$ under spatial refinement is shown. The expected order $\mathcal{O}(h^2)$ is observed. This illustrates the convergence results from subsection 6.2 with respect to the *spatial* discretization.

Remark 7.1. The data of the presented numerical example fulfill the regularity conditions stated in Assumption 1. If this assumption is violated (e.g., through the missing compatibility $u_0, \Delta u_0 \in V$), the Crank–Nicolson scheme may exhibit unstable behavior. This can be circumvented by adding some damping steps as proposed in [17, 26]. An investigation of the resulting damped scheme in the context of optimal control will be a subject of future research.

REFERENCES

[1] T. Apel and T. Flaig, *Crank-Nicolson Schemes for Optimal Control Problems with Evolution Equations*, Preprint SPP1253–113, DFG priority program 1253 "Optimization with PDEs," Friedrich-Alexander-Universität Erlangen-Nürnberg, Erlangen, Germany, 2010.

[2] A. Aziz and P. Monk, *Continuous finite elements in space and time for the heat equation*, Math. Comp., 52 (1989), pp. 255–274.

[3] R. Becker, D. Meidner, and B. Vexler, *Efficient numerical solution of parabolic optimization problems by finite element methods*, Optim. Methods Softw., 22 (2007), pp. 813–833.

[4] R. Becker and B. Vexler, *Optimal control of the convection-diffusion equation using stabilized finite element methods*, Numer. Math., 106 (2007), pp. 349–367.

[5] M. Bergounioux, K. Ito, and K. Kunisch, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37 (1999), pp. 1176–1194.

[6] E. Burman, *Crank-Nicolson finite element methods using symmetric stabilization with an application to optimal control problems subject to transient advection-diffusion equations*, Commun. Math. Sci., 9 (2011), pp. 319–329.

[7] K. Chrysafinos, *Convergence of discontinuous Galerkin approximations of an optimal control problem associated to semilinear parabolic PDE's*, M2AN Math. Model. Numer. Anal., 44 (2010), pp. 189–206.

[8] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Classics Appl. Math. 40, SIAM, Philadelphia, 2002.

[9] R. Dautray and J.-L. Lions, *Evolution Problems* I, Math. Anal. Numer. Methods Sci. Technol. 5, Springer-Verlag, Berlin, 1992.

[10] K. Deckelnick and M. Hinze, *Variational discretization of parabolic control problems in the presence of pointwise state constraints*, J. Comput. Math., 29 (2011), pp. 1–16.

[11] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson, *Computational Differential Equations*, Cambridge University Press, Cambridge, UK, 1996.

[12] L. C. Evans, *Partial Differential Equations*, Grad. Stud. Math. 19, AMS, Providence, RI, 2002.

[13] Gascoigne, *High Performance Finite Element Toolkit*, http://www.gascoigne.uni-hd.de (2002).

[14] K. Kunisch and A. Rösch, *Primal-dual active set strategy for a general class of constrained optimal control problems*, SIAM J. Optim., 13 (2002), pp. 321–334.

[15] I. Lasiecka and K. Malanowski, *On discrete-time Ritz-Galerkin approximation of control constrained optimal control problems for parabolic systems*, Control Cybernet., 7 (1978), pp. 21–36.

[16] J.-L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Grundlehren Math. Wiss. 170, Springer, Berlin, 1971.

[17] M. Luskin and R. Rannacher, *On the smoothing property of the Galerkin method for parabolic equations*, SIAM J. Numer. Anal., 19 (1982), pp. 93–113.

[18] K. Malanowski, *Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems*, Appl. Math. Optim., 8 (1981), pp. 69–95.

[19] R. S. McNight and W. E. Bosarge, Jr., *The Ritz–Galerkin procedure for parabolic control problems*, SIAM J. Control Optim., 11 (1973), pp. 510–524.

[20] D. Meidner, R. Rannacher, and B. Vexler, *A priori error estimates for finite element discretizations of parabolic optimization problems with pointwise state constraints in time*, SIAM J. Control Optim., 49 (2011), pp. 1961–1997.

[21] D. Meidner and B. Vexler, *Adaptive space-time finite element methods for parabolic optimization problems*, SIAM J. Control Optim., 46 (2007), pp. 116–142.

[22] D. Meidner and B. Vexler, *A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part* I: *Problems without control constraints*, SIAM J. Control Optim., 47 (2008), pp. 1150–1177.

[23] D. Meidner and B. Vexler, *A priori error estimates for space-time finite element discretization of parabolic optimal control problems. Part* II: *Problems with control constraints*, SIAM J. Control Optim., 47 (2008), pp. 1301–1329.

[24] C. Meyer and A. Rösch, *Superconvergence properties of optimal control problems*, SIAM J. Control Optim., 43 (2004), pp. 970–985.

[25] I. Neitzel and B. Vexler, *A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems*, Numer. Math., to appear; published online September 15, 2011 (doi: 10.1007/s00211-011-0409-9).

[26] R. Rannacher, *Finite element solution of diffusion problems with irregular data*, Numer. Math., 43 (1984), pp. 309–327.

[27] RoDoBo*, A C++ Library for Optimization with Stationary and Nonstationary PDEs with interface to* Gascoigne, http://www.rodobo.uni-hd.de (2010).

[28] A. Rösch, *Error estimates for parabolic optimal control problems with control constraints*, Z. Anal. Anwendungen, 23 (2004), pp. 353–376.

[29] A. Rösch and B. Vexler, *Optimal control of the Stokes equations: A priori error analysis for finite element discretization with postprocessing*, SIAM J. Numer. Anal., 44 (2006), pp. 1903–1920.

[30] F. Schieweck, *A-stable discontinuous Galerkin-Petrov time discretization of higher order*, J. Numer. Math., 18 (2010), p. 25–57.

[31] V. Thomée, *Galerkin Finite Element Methods for Parabolic Problems*, Springer Ser. Comput. Math. 25, Springer, Berlin, 2006.

[32] R. Winther, *Error estimates for a Galerkin approximation of a parabolic control problem*, Ann. Mat. Pura Appl. (4), 117 (1978), pp. 173–206.

[33] J. Wloka, *Partielle Differentialgleichungen: Sobolevräume und Randwertaufgaben*, B. G. Teubner, Stuttgart, 1982.