Computational Optimization and Applications manuscript No.
(will be inserted by the editor)

Third order convergent time discretization for parabolic optimal control problems with control constraints

Andreas Springer · **Boris Vexler**

Received: date / Accepted: date

Abstract We consider a priori error analysis for a discretization of a linear quadratic parabolic optimal control problem with box constraints on the time-dependent control variable. For such problems one can show that a time-discrete solution with second order convergence can be obtained by a first order discontinuous Galerkin time discretization for the state variable and either the variational discretization approach or a post-processing strategy for the control variable. Here, by combining the two approaches for the control variable, we demonstrate that almost third order convergence with respect to the size of the time steps can be achieved.

Keywords optimal control · heat equation · control constraints · discontinuous Galerkin time stepping · error estimates · post-processing · variational control discretization

Mathematics Subject Classification (2000) 49J20 · 35K20 · 49M05 · 49M25 · 49M29 · 65M12 · 65M60

1 Introduction

Let $\Omega \subset \mathbb{R}^d$ be a spatial domain and $I = (0,T)$ an open time interval. Then our model problem is given by: Minimize the cost functional

$$J(q,u) = \frac{1}{2}\int_0^T \int_\Omega (u(t,x) - u_d(t,x))^2 \,\mathrm{d}x\,\mathrm{d}t + \frac{\alpha}{2}\int_0^T \sum_{i=0}^{D} q_i(t)^2 \,\mathrm{d}t \tag{1a}$$

A. Springer
Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik, Boltzmannstraße 3, 85748 Garching b. München, Germany
E-mail: springer@ma.tum.de

B. Vexler
Lehrstuhl für Mathematische Optimierung, Technische Universität München, Fakultät für Mathematik, Boltzmannstraße 3, 85748 Garching b. München, Germany
E-mail: vexler@ma.tum.de

with the time-dependent control $q\colon (0,T) \to \mathbb{R}^D$ and the state variable $u = u(t,x)$ subject to the state equation

$$\begin{aligned} \partial_t u - \Delta u &= f + Gq && \text{in } I \times \Omega \\ u &= 0 && \text{on } I \times \partial\Omega \\ u(0) &= u_0 && \text{in } \Omega \end{aligned} \tag{1b}$$

and to the point-wise control constraints

$$q^a \le q(t) \le q^b \text{ for almost all } t \in I. \tag{1c}$$

The scalar value α is assumed to be positive, the bounds $q^a, q^b \in \mathbb{R}^D$ are given fixed vectors with $q^a < q^b$ componentwise, and G is a linear operator, see Section 2 for more details.

We are interested in an efficient discretization of this type of problem. Our focus here is on reducing the number of degrees of freedom required for an accurate time discretization by using a higher order convergent scheme. Due to the box constraints on the control, the solution of the model problem (1) has limited regularity which restricts the order of convergence we can achieve for the time discretization. Previous results about time discretization of problems with control constraints include the works [12], [14], [18] that discuss first order convergent schemes and the article [23] that shows convergence with order $\mathcal{O}(k^{\frac{3}{2}})$ for a one-dimensional problem. More recently, the work [19] presented a time discretization based on a first order continuous Petrov-Galerkin scheme for the state and piece-wise constant ansatz functions for the control. After a post-processing step the resulting approximation for the control converges with second order.

Here, we propose a discontinuous Galerkin time discretization with piece-wise linear trial functions for the state variable. The control is treated with the variational discretization concept by Hinze, cf. [10] for application to elliptic problems. We will show that this results in second order accuracy of the time discretization. Based on this variational solution we perform a post processing step and show that the resulting improved optimal control $\tilde{q}_{kh}$ (see Equation (51) for its definition) converges with order $\mathcal{O}(|\log k|^{\frac{1}{2}} k^3)$ with respect to the temporal L^2 norm. The error introduced by the spatial discretization of the state variable can be analyzed independently and decreases for conforming bilinear finite elements with $\mathcal{O}(h^2)$.

The post-processing is based on a higher order continuous interpolation of a discontinuous Galerkin solution at the nodes of the Radau quadrature. This kind of reconstruction has been used previously in improving the order of convergence for discontinuous Galerkin time discretization of various time-dependent differential equations, see, e. g., [15] and [2], and for constructing a posteriori error estimators, see, e. g. [13,3], [27] and [7]. Here, we derive optimal order a priori estimates for this reconstruction applied to the adjoint state.

The remainder of the paper is organized as follows: In Section 2 we discuss the functional analytic setting and the optimality conditions for the model problem (1). The regularity of the optimal solution is discussed in detail. The third section describes the discretization of the problem in time with a discontinuous Galerkin scheme and in space with conforming elements and collects tools for the a priori analysis. In Section 4 we derive estimates for the discretization error when solving the state and adjoint equations for a given fixed control. In particular, we show that, given sufficient regularity, the reconstruction of the semidiscrete adjoint solution converges with order $\mathscr{O}(k^{r+2})$ for a discontinuous Galerkin time discretization of order r. The results for $r = 1$ and for order $s = 1$ of the spatial discretization are used in Section 5 to derive the error estimate for the post-processed optimal control. In the final section, we illustrate our results by a numerical example and provide evidence that the variational treatment of the control variable is in fact necessary to achieve the proposed order of convergence after post-processing.

2 Problem formulation and regularity results

In this section we give a precise formulation of the optimization problem (1), recall optimality conditions, and discuss the regularity of the optimal solution in greater detail.

Let the spatial domain $\Omega \subset \mathbb{R}^d$, $d = 1,2,3$, be polygonal and convex. The Hilbert spaces $V := H_0^1(\Omega)$ and $H := L^2(\Omega)$ together with the dual $V^* = H^{-1}(\Omega)$ of V form a Gelfand triple $V \hookrightarrow H \hookrightarrow V^*$. On a time interval $I := (0,T)$ we define the state space X by

$$X = W(I) = \left\{ v \in L^2(I,V) \,\middle|\, \partial_t v \in L^2(I,V^*) \right\}$$

and the control space Q by

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

where D is a positive integer. For convenience, the following notations for inner products and norms on the spaces H, Q, and $L^2(I,H)$ are used:

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

With these preparations, a weak form of the state equation (1b) can be stated as: Find $u \in X$ such that

$$\begin{aligned} (\partial_t u, \varphi)_I + (\nabla u, \nabla \varphi)_I &= (f + Gq, \varphi)_I \quad && \text{for all } \varphi \in X, \\ u(0) &= u_0 && \text{in } \Omega. \end{aligned} \tag{2}$$

The linear operator G is defined from $\mathbb{R}^D$ to H with image in V. Its operator norm is denoted $\|G\|$. We extend G to time-dependent functions by setting $(Gq)(t) := G(q(t))$.

For the data we assume $u_0 \in V$ and $f \in L^2(I,H)$. Then there exists a unique control-to-state mapping $q \mapsto u(q)$, $Q \to X$, where $u = u(q)$ is the solution of (2) for the given q. After introducing the reduced cost functional

$$j(q) := J(q,u(q)) \tag{3}$$

and the admissible set

$$Q_{\mathrm{ad}} := \{ q \in Q \,|\, q \text{ satisfies (1c)} \}, \tag{4}$$

we can rewrite the optimal control problem in reduced form as

$$\text{Minimize } j(q) \text{ subject to } q \in Q_{\text{ad}}. \tag{5}$$

There exists a unique solution $\bar{q}$ with corresponding optimal state $\bar{u}$ for this problem. Due to the linear quadratic structure of the problem and the convexity of Q_{ad}, the first order necessary optimality condition is also sufficient for optimality. It reads

$$j'(\bar{q})(\delta q - \bar{q}) \geq 0 \quad \forall \delta q \in Q_{\text{ad}}. \tag{6}$$

The first derivative of the reduced cost functional j can be computed by

$$j'(q)(\delta q - q) = \langle \alpha q + G^* z, \delta q - q \rangle_I$$

where z is the solution of the adjoint problem given by: Find $z \in X$ such that

$$\begin{aligned} -(\varphi, \partial_t z)_I + (\nabla \varphi, \nabla z)_I &= (u - u_d, \varphi)_I \quad \forall \varphi \in X, \\ z(T) &= 0. \end{aligned} \tag{7}$$

Using the solution $\bar{z}$ of (7) with $\bar{u}$ entering the right hand side, the first order optimality condition (6) can be expressed equivalently as

$$\bar{q} = P_{Q_{\text{ad}}}\left(-\alpha^{-1} G^* \bar{z}\right), \tag{8}$$

where the pointwise projection $P_{Q_{\text{ad}}}$ onto the admissible set is given by

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

Lemma 1 *The solution of the state equation* (2) *has the improved regularity*

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

and satisfies the stability estimate

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

Proof In [6, Chapter 7, Theorem 2] the statement is shown for a domain Ω with C^2 boundary. Looking at the proof, we note that the smoothness requirement on the boundary of Ω is only needed in step 3 to apply the corresponding elliptic regularity result Theorem 4 in Chapter 6 on the spatial part of the differential operator. However, this result can also be shown for convex polygonal domains, see, e. g., [9, Theorem 4.4.3.7] for a two-dimensional domain and [16, Theorem 4.3.2] for the three-dimensional case. □

In what follows, we need some additional regularity assumptions on the data of the optimal control problem:

Assumption 1 We assume the data u_0, f and u_d satisfy the following conditions:

- $u_0 \in V$ with $\Delta u_0 \in V$,
- $f \in H^1(I, H) \cap C(\bar{I}, V)$,
- $u_d \in H^2(I, H) \cap H^1(I, H^2(\Omega) \cap V)$, and $\Delta u_d(T) \in V$.

Lemma 2 *Let $(\bar{q},\bar{u})$ be the solution of the optimal control problem (1) and $\bar{z}$ the corresponding adjoint state. If Assumption 1 is satisfied, we obtain the improved regularities*

$$\begin{aligned}&\bar{u}\in H^2(I,H)\cap H^1(I,H^2(\Omega)\cap V)\hookrightarrow C^1(\bar{I},V),\\&\bar{z}\in H^3(I,H)\cap H^2(I,H^2(\Omega)\cap V)\hookrightarrow C^2(\bar{I},V),\ \textit{and}\\&\bar{q}\in W^{1,\infty}(I,\mathbb{R}^D).\end{aligned}$$

Moreover we have the stability estimates

$$\begin{aligned}\|\partial_t\Delta\bar{u}\|_I+\left\|\partial_t^2\bar{u}\right\|_I &\le C\Big\{\|f\|_{H^1(I,H)}+\|\bar{q}\|_{H^1(I,\mathbb{R}^D)}+\|\nabla f(0)\|\\&\qquad+\|\nabla\Delta u_0\|\Big\}\ \textit{and}\\ \left\|\partial_t^3\bar{z}\right\|_I+\left\|\partial_t^2\Delta\bar{z}\right\|_I+\left\|\nabla\partial_t^2\bar{z}(T)\right\| &\le C\Big\{\left\|\partial_t^2u_d\right\|_I+\|\nabla\partial_t u_d(T)\|+\|\nabla\Delta u_d(T)\|\\&\qquad+\|f\|_{H^1(I,H)}+\|\nabla f(0)\|+\|\nabla f(T)\|\\&\qquad+\|\bar{q}\|_{H^1(I,\mathbb{R}^D)}+|\bar{q}(T)|+\|\nabla\Delta u_0\|\Big\}.\end{aligned}$$

Proof The stated regularity results for $\bar{u}$ and $\bar{q}$ and the stability estimate for $\bar{u}$ are shown in [19, Proposition 2.3]. Furthermore, the authors prove that the adjoint solution satisfies $\bar{z}\in H^2(I,H)\cap H^1(I,H^2(\Omega)\cap V)\hookrightarrow C^1(\bar{I},V)$ with the stability estimate

$$\|\partial_t\Delta\bar{z}\|_I+\left\|\partial_t^2\bar{z}\right\|_I\le C\left\{\|u_d\|_{H^1(I,H)}+\|\nabla u_d(T)\|+\|f\|_I+|\bar{q}|_I+\|\nabla u_0\|\right\}.$$

Using the regularity already shown, we verify that $\hat{z}:=\partial_t\bar{z}$ satisfies the equation

$$-\partial_t\hat{z}-\Delta\hat{z}=\partial_t(\bar{u}-u_d) \tag{9}$$

with the terminal condition

$$\hat{z}(T)=\partial_t\bar{z}(T)=-\Delta\bar{z}(T)-(\bar{u}-u_d)(T)=-(\bar{u}-u_d)(T) \tag{10}$$

since $\bar{z}(T)=0$ and hence $-\Delta\bar{z}(T)=0$. We differentiate equation (9) formally another time with respect to the time variable resulting in

$$-\partial_t\tilde{z}-\Delta\tilde{z}=\partial_t^2(\bar{u}-u_d). \tag{11}$$

for a new variable $\tilde{z}$. For the terminal condition of this equation we set

$$\begin{aligned}\tilde{z}(T)&=\partial_t\hat{z}(T)=-\Delta\hat{z}(T)-\partial_t(\bar{u}-u_d)(T)\\&=-\partial_t(\bar{u}-u_d)(T)+\Delta(\bar{u}-u_d)(T)=(\partial_t u_d-\Delta u_d)(T)-(f+G\bar{q})(T).\end{aligned}$$

In the second line, the terminal condition (10) for $\hat{z}$ and the state equation (2) were plugged in.

We note that with Assumption 1 and the shown regularity for $\bar{q}$, the terminal value for $\tilde{z}$ is in V. Hence we can apply the regularity result from Lemma 1 to Equation (11). This gives $\tilde{z} \in H^1(I,H) \cap L^2(I,H^2(\Omega) \cap V)$. The corresponding stability estimate reads

$$\begin{aligned}\|\partial_t \tilde{z}\|_I + \|\Delta \tilde{z}\|_I + \|\nabla \tilde{z}(T)\| \\ \leq C \left(\left\| \partial_t^2 (\bar{u} - u_d) \right\|_I + \|\nabla (\partial_t u_d(T) - \Delta u_d(T))\| + \|\nabla f(T)\| + |\bar{q}(T)| \right).\end{aligned}$$

The first term on the right hand side can be estimated by the stability estimate for the state equation. We verify in the same way as in the proof of Theorem 27.2 in [29] that in fact $\tilde{z} = \partial_t^2 \bar{z}$. This completes the proof. □

3 Discretization of the optimal control problem

3.1 Semidiscretization in time

For the temporal discretization of the state equation, we use a discontinuous Galerkin discretization of order r, which we subsequently refer to as a dG(r) scheme. Therefore we consider a partitioning of the half-open time interval $\tilde{I} = (0,T]$ of the form

$$\tilde{I} = I_1 \cup I_2 \cup \cdots \cup I_M$$

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

$$0 = t_0 < t_1 < \cdots < t_M = T.$$

The discretization parameter k denotes the maximal size of the time steps, that is $k := \max\{k_1, k_2, \ldots, k_M\}$.

Assumption 2 We impose a regularity condition on the temporal mesh and require that there is a constant $\kappa \geq 1$ independent of the mesh width k such that

$$\kappa^{-1} \leq \frac{k_m}{k_{m-1}} \leq \kappa \quad \forall m = 2,3,\ldots,M.$$

The semidiscrete test and trial space for the discontinuous Galerkin method of order r is given by

$$X_k^r := \left\{ v \in L^2(I,V) \,\middle|\, v|_{I_m} \in \mathscr{P}_r(I_m,V),\ m = 1,\ldots,M \right\},$$

where $\mathscr{P}_r(I,V)$ denotes the space of polynomials of degree less or equal r on I with values in V. To simplify notation we introduce the abbreviations

$$\begin{aligned} v_k(t)^- &:= \lim_{\tau \to 0^+} v_k(t - \tau), & v_k(t)^+ &:= \lim_{\tau \to 0^+} v_k(t + \tau), \\ v_{k,m}^{-/+} &:= v_k(t_m)^{-/+}, & \text{and} \quad [v_k]_m &:= v_{k,m}^+ - v_{k,m}^- \end{aligned}$$

for functions $v_k \in X_k^r$. Introducing the bilinear form $B : X_k^r \times X_k^r \to \mathbb{R}$ given by

$$B(u_k,\varphi) := \sum_{m=1}^{M} (\partial_t u_k,\varphi)_{I_m} + (\nabla u_k,\nabla\varphi)_I + \sum_{m=1}^{M-1} ([u_k]_m,\varphi_m^+) + (u_0^+,\varphi_0^+) \tag{12}$$

the dG(r) semidiscrete state equation reads: For given control q, find $u_k \in X_k^r$ such that

$$B(u_k,\varphi) = (f+Gq,\varphi)_I + (u_0,\varphi_0^+) \quad \forall\varphi \in X_k^r. \tag{13}$$

A semidiscrete adjoint type equation with some given right hand side g and terminal condition $z_k(T) = 0$ takes the form

$$B(\varphi,z_k) = (\varphi,g)_I \quad \forall\varphi \in X_k^r. \tag{14}$$

Remark 1 As noted for example in [17], the continuous solution u of (2) for given control $q \in Q$ fulfills the semidiscrete state equation as well. Hence, although the dG(r) semidiscretization is non-conforming, we get *Galerkin orthogonality*, i. e.,

$$B(u-u_k,\varphi) = 0 \quad \forall\varphi \in X_k^r \tag{15}$$

holds.

The semidiscrete optimization problem reads:

$$\text{Minimize } J(q_k,u_k) \text{ subject to (13) and } (q_k,u_k) \in Q_{\text{ad}} \times X_k^r. \tag{16}$$

As in the continuous case, the problem has a unique solution $(\bar{q}_k,\bar{u}_k)$ and the first order optimality condition can be expressed in terms of the corresponding semidiscrete adjoint state $\bar{z}_k$.

The adjoint equation has the form (14) with right hand side $g := \bar{u}_k - u_d$. Note that, since we use a Galerkin discretization, we end up with the same equation if we discretize the continuous adjoint equation (7) with a dG(r) scheme in time, that is, the two approaches *"optimize-then-discretize"* and *"discretize-then-optimize"* are equivalent. To see this, we transform the bilinear form B. After interval-wise integration by parts with respect to time we obtain the equivalent representation

$$B(\varphi,\psi) = -\sum_{m=1}^{M} (\varphi,\partial_t\psi)_{I_m} + (\nabla\varphi,\nabla\psi)_I - \sum_{m=1}^{M-1} (\varphi_m^-,[\psi]_m) + (\varphi_M^-,\psi_M^-). \tag{17}$$

The first order optimality condition for the semidiscrete problem is given as

$$\langle \alpha\bar{q}_k + G^*\bar{z}_k, \delta q - \bar{q}_k\rangle_I \ge 0 \quad \forall\delta q \in Q_{\text{ad}}, \tag{18}$$

or, equivalently,

$$\bar{q}_k = P_{Q_{\text{ad}}}\left(-\alpha^{-1}G^*\bar{z}_k\right). \tag{19}$$

Later on, we need the following stability estimates for the semidiscrete equations:

Theorem 1 *For the solution $u_k \in X_k^r$ of the semidiscrete state equation* (13) *with right-hand side $f \in L^2(I,H)$, control $q \in Q_{ad}$ and initial condition $u_0 \in V$ the stability estimate*

$$\|u_k\|_I^2 + \sum_{m=1}^{M} \|\partial_t u_k\|_{I_m}^2 + \|\Delta u_k\|_I^2 + \sum_{m=1}^{M} k_m^{-1} \|[u_k]_{m-1}\|^2 \\ \leq C\left(\|f+Gq\|_I^2 + \|u_0\|^2 + \|\nabla u_0\|^2\right)$$

holds when defining the jump $[u_k]_0$ as $u_{k,0}^+ - u_0$. The constant C depends only on the domain Ω, the final time T and the order r of the semidiscretization.

Proof See [17, Theorems 4.1 and 4.3]. □

Corollary 1 *The solution $z_k \in X_k^r$ of the semidiscrete adjoint equation* (14) *for any right hand side $g \in L^2(I,H)$ satisfies the stability estimate*

$$\|z_k\|_I^2 + \sum_{m=1}^{M} \|\partial_t z_k\|_{I_m}^2 + \|\Delta z_k\|_I^2 + \sum_{m=1}^{M} k_m^{-1} \|[z_k]_m\|^2 \leq C\|g\|_I^2 .$$

Here, the jump term $[z_k]_M$ at final time is set to be $-z_{k,M}^-$.

Lemma 3 *For the solution z_k of the semidiscrete adjoint equation* (14) *we have additionally the stability estimate*

$$\|z_k\|_{L^\infty(I,H)} \leq C\|g\|_I$$

with the constant C only depending on the domain Ω, the final time T, and the order r of the discretization.

Remark 2 An analogous estimate can be shown for the semidiscrete state solution. It reads

$$\|u_k\|_{L^\infty(I,H)} \leq C\left(\|f+Gq\|_I + \|u_0\| + \|\nabla u_0\|\right).$$

Proof We estimate the spatial L^2 norm of z_k at a fixed time $t^* \in I$ with $t^* \neq t_m$ for any m. Therefore let m^* denote the smallest index such that $t_{m^*} \geq t^*$. Then, the norm of $z_k(t^*)$ can be written as

$$\|z_k(t^*)\| = \left\| -\int_{t^*}^{t_{m^*}} \partial_t z_k \,\mathrm{d}t - \sum_{m=m^*+1}^{M} \int_{I_m} \partial_t z_k \,\mathrm{d}t - \sum_{m=m^*}^{M} [z_k]_m \right\| .$$

We define the function v_k interval-wise by $v_k|_{I_m} = \partial_t z_k$. Together with the triangle inequality we obtain

$$\|z_k(t^*)\| \leq \left\| -\int_{t^*}^{T} v_k \,\mathrm{d}t \right\| + \sum_{m=m^*}^{M} \|[z_k]_m\| .$$

The sum on the right hand side can be estimated by means of Corollary 1, giving

$$\sum_{m=m^*}^{M} \|[z_k]_m\| \le \sum_{m=1}^{M} k_m^{\frac{1}{2}} \cdot k_m^{-\frac{1}{2}} \|[z_k]_m\| \le \sqrt{T} \left(\sum_{m=1}^{M} \frac{1}{k_m} \|[z_k]_m\|^2 \right)^{\frac{1}{2}} \le C \|g\|_I .$$

For the integral term, we get with Hölder's inequality and the stability estimate from Corollary 1

$$\left\| -\int_{t^*}^{T} v_k \, \mathrm{d}t \right\| \le \sqrt{T} \|v_k\|_I \le C \|g\|_I .$$

This shows the claim. □

Additionally we need the following stability estimate for a semidiscrete auxiliary adjoint equation:

Theorem 2 *Let $w \in L^2(I, H^2(\Omega) \cap V)$ be given. Then the solution $y_k \in X_k^r$ of the equation*

$$B(\varphi, y_k) = (\varphi, w)_I \quad \forall \varphi \in X_k^r \tag{20}$$

satisfies the estimate

$$\left\|\Delta^2 y_k\right\|_I + \left(\sum_{m=1}^{M} \|\partial_t \Delta y_k\|_{I_m} \right)^{\frac{1}{2}} \le C \|\Delta w\|_I .$$

The proof uses a Galerkin approximation in the spatial variable and is given in detail in the appendix.

As key tools for obtaining and proving almost third order convergence of our time discretization we need several projection and interpolation operators. Some of them are defined in terms of the nodes of the $r+1$ point Radau quadrature rule on the intervals of the time discretization with a fixed node at the left interval boundary (see, e. g., [1, p. 888]). To give a characterization of those points, let the $(r+1)^{\text{th}}$ Radau polynomial on the interval $[-1,1]$ be given by

$$\hat{R}_{r+1}(t) = \hat{L}_{r+1}(t) + \hat{L}_r(t),$$

where $\hat{L}_j$ is the Legendre polynomial of degree j. We denote the $r+1$ roots of this polynomial by $-1 = b_0^{r+1}, \dots, b_r^{r+1}$. Transforming them onto the interval I_m gives the points $\theta_{m,j} := t_{m-1} + \frac{k_m}{2}\left(b_j^{r+1} + 1\right)$ for j in $0, \dots, r$.

With these preparations, we define the following interpolation and projection operators:

1. the L^2 *projection* Π_k^0 onto the space of piece-wise constant functions in time given by

$$\Pi_k^0 \colon L^2(I, V) \to X_k^0, \; \Pi_k^0 v\big|_{I_m} = \frac{1}{k_m} \int_{I_m} v(t) \, \mathrm{d}t .$$

2. a projection $P_k\colon C(\bar I,V)\to X_k^r$ that is defined interval-wise by the two conditions

$$(P_k v - v, \varphi)_{I_m} = 0 \quad \forall \varphi \in \mathscr{P}_{r-1}(I_m,V), \tag{21a}$$

$$P_k v(t_m)^- = v(t_m)^- \tag{21b}$$

   for each $m=1,\dots,M$. This operator is commonly employed in the error analysis of discontinuous Galerkin methods, see, e. g., [28, Chapter 12].

3. an *interpolation operator* $\pi_k : C(\bar I,V)\to X_k^r$ which is defined interval-wise by

$$\pi_k v(\theta_{m,j})^+ = v(\theta_{m,j}) \quad \forall m=1,\dots,M,\ j=0,\dots,r.$$

4. an auxiliary operator $\hat\pi_k : X_k^r \cup C(\bar I,V)\to X_k^{r+1}\cap C(\bar I,V)$ with

$$\hat\pi_k v(\theta_{m,j}) = v(\theta_{m,j})^+ \quad \forall m=1,\dots,M,\ j=0,\dots,r \quad \text{and}$$
$$\hat\pi_k v(T) = v(T)^-.$$

   A time-reversed version of this kind of operator has been used previously, for example in [3,27,7] to construct a posteriori error estimates or in [2,15] to recover improved solutions from the computed ones.

5. the *reconstruction operator* $\tilde\pi_k : X_k^r \cup C(\bar I,V)\to X_k^{r+1}\cap C(\bar I,V)$, which is the same as the previous operator except for on the last interval. It is determined by the following conditions:

$$\tilde\pi_k v\big|_{I_{M-1}\cup I_M} \in \mathscr{P}_{r+1}(I_{M-1}\cup I_M,V), \tag{22a}$$

$$\tilde\pi_k v\big|_{I_m} = \hat\pi_k v \quad \forall m=1,\dots,M-1. \tag{22b}$$

   That means we extend the reconstruction polynomial on the second-last interval onto the last interval.

The effect of the operators π_k, $\hat\pi_k$ and $\tilde\pi_k$ for the case $r=1$ is visualized in Figure 1. Since all those operators act only on the time variable, we can extend the definitions in the obvious way to control-type variables by replacing the spatial spaces V and H by $\mathbb{R}^D$ while requiring the same temporal regularities. We will use the same notations for those operators acting on time-dependent functions with values in $\mathbb{R}^D$.

Lemma 4 *Let $v\in H^1(I,H)$. For the operators defined above we have the estimates*

$$\left\|v - \Pi_k^0 v\right\|_I \le Ck\left\|\partial_t v\right\|_I, \tag{23}$$

and if we require additionally $v\in C(\bar I,V)$

$$\left\|v - P_k v\right\|_I \le Ck^{r+1}\left\|\partial_t^{r+1} v\right\|_I \qquad \textit{for } v\in H^{r+1}(I,H), \tag{24}$$

$$\left\|v - \pi_k v\right\|_I \le Ck^{r+1}\left\|\partial_t^{r+1} v\right\|_I \qquad \textit{for } v\in H^{r+1}(I,H), \tag{25}$$

$$\left\|v - \hat\pi_k v\right\|_I \le Ck^{r+2}\left\|\partial_t^{r+2} v\right\|_I \qquad \textit{for } v\in H^{r+2}(I,H), \tag{26}$$

$$\left\|v - \tilde\pi_k v\right\|_I \le Ck^{r+2}\left\|\partial_t^{r+2} v\right\|_I \qquad \textit{for } v\in H^{r+2}(I,H). \tag{27}$$

Proof All of the above estimates can be shown in the standard way by transforming each interval to the unit interval, applying the Bramble-Hilbert Lemma and transforming back, for the estimate (24) this is done in the proof of Theorem 12.1 in [28]. For the reconstruction operator $\tilde\pi_k$, the last two discretization intervals are treated as a single interval. □

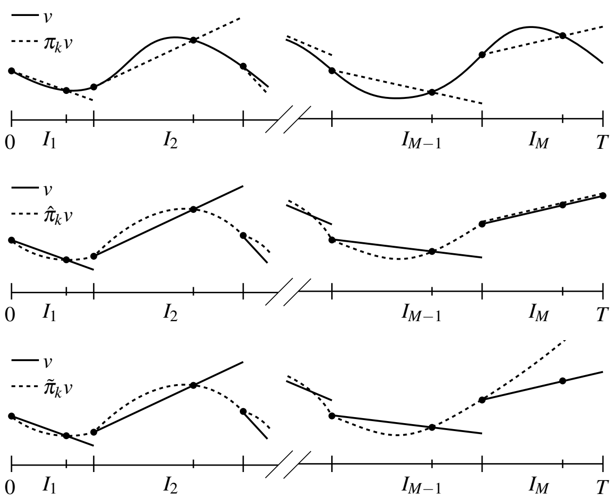


Fig. 1 Visualization of the operators π_k, $\hat{\pi}_k$ and $\tilde{\pi}_k$ for $r = 1$.

3.2 Spatial discretization of the state

For the finite element discretization in space, we consider shape regular meshes consisting of intervals, quadrilaterals or hexahedrons respectively, depending on the dimension d of the domain Ω. The individual cells are denoted by K, the mesh with discretization parameter h representing the maximum cell diameter is written as $\mathcal{T}_h = \{K\}$. On the mesh $\mathcal{T}_h$ we define the conforming finite element space $V_h^s \subset V$ in the standard way by

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

The space $\mathcal{Q}_s(K)$ is obtained by bi- or tri-linear transformations of the polynomial space $\hat{\mathcal{Q}}_s(\hat{K})$ given on the reference cell $\hat{K} = [0,1]^d$ as

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

Using the space V_h^s we can define the fully discrete state space as

$$X_{k,h}^{r,s} = \left\{ \varphi \in L^2(I,V) \,\middle|\, \varphi|_{I_m} \in \mathcal{P}_r(I_m, V_h^s),\ m = 1,\dots,M \right\} \subseteq X_k^r.$$

Then the fully discretized state equation reads: for given control $q \in Q$ find $u_{kh} \in X_{k,h}^{r,s}$ such that

$$B(u_{kh}, \varphi) = (f + Gq, \varphi)_I \quad \forall \varphi \in X_{k,h}^{r,s}. \tag{28}$$

For the state-discrete optimal control problem we have

$$\text{Minimize } J(q_{kh}, u_{kh}) \text{ subject to (28) and } (q_{kh}, u_{kh}) \in Q_{\text{ad}} \times X_{k,h}^{r,s}, \tag{29}$$

and the corresponding adjoint equation is: find $z_{kh} \in X_{k,h}^{r,s}$ such that

$$B(\varphi, z_{kh}) = (\varphi, g)_I \quad \forall \varphi \in X_{k,h}^{r,s} \tag{30}$$

with $g := u_{kh} - u_d$. The first order optimality condition for the optimal solution $(\bar{q}_{kh}, \bar{u}_{kh})$ in terms of the corresponding adjoint state $\bar{z}_{kh}$ can be stated as

$$(\alpha \bar{q}_{kh} + G^* \bar{z}_{kh}, \delta q - \bar{q}_{kh})_I \geq 0 \quad \forall \delta q \in Q_{\text{ad}}, \tag{31}$$

or, equivalently

$$\bar{q}_{kh} = P_{Q_{\text{ad}}} \left(-\frac{1}{\alpha} G^* \bar{z}_{kh} \right). \tag{32}$$

We quote the following stability estimate from Theorem 4.6 and Corollary 4.7 in [17]:

Theorem 3 *Let $\Pi_h : V \to V_h^s$ denote the L^2 projection onto the space V_h^s. For the solution $u_{kh} \in X_{k,h}^{r,s}$ of the discrete state equation* (28) *with right hand side $f \in L^2(I,H)$, control $q \in Q_{ad}$ and initial condition $u_0 \in V$ the stability estimate*

$$\|u_{kh}\|_I^2 + \|\nabla u_{kh}\|_I^2 \leq C \left\{ \|f + Gq\|_I^2 + \|\nabla \Pi_h u_0\|_I^2 + \|\Pi_h u_0\|_I^2 \right\}$$

holds. The solution z_{kh} of the discrete adjoint equation (30) *with some right hand side $g \in L^2(I,H)$ satisfies*

$$\|z_{kh}\|_I^2 + \|\nabla z_{kh}\|_I^2 \leq C \|g\|_I^2.$$

3.3 Variational discretization of the control

We avoid discretizing the control space (which is infinite-dimensional with respect to time) by using the so called variational approach first proposed in [10]. Looking at the optimality condition (32), we note that given the discrete adjoint $\bar{z}_{kh}$, the corresponding optimal control is numerically computable by evaluating the point-wise projection operator $P_{Q_{\text{ad}}}$. The resulting control is piece-wise polynomial but in general not with respect to the partitioning $I_1, \ldots, I_M$.

4 Error estimates for state and adjoint state for a fixed control

In this section we derive error estimates for the semidiscrete and discrete state and adjoint state computed from a given fixed control q. For the state equation we quote results for the temporal discretization error with respect to the $L^2(I,H)$ and the $L^\infty(I,H)$ norms. Subsequently we derive an a priori estimate for the error of the reconstructed adjoint solution which is obtained by applying the reconstruction operator $\tilde{\pi}_k$ to the computed semidiscrete adjoint solution. Given sufficient regularity of the solutions, we show that for a dG(r) semidiscretization, the $L^2(I,H)$ error of the reconstruction converges with order $r+2$ with respect to the step size k to the exact adjoint solution, that is, we gain one power of k compared to the plain dG(r) solution. When setting

$r = 1$, the regularities shown in Lemma 2 are sufficient to apply the estimate to the solution of the optimal control problem.

Throughout this section we assume to be given a fixed control $q \in Q$ and denote the corresponding continuous state solution by $u(q)$ and the solution of the semidiscrete problem (13) with control q by $u_k(q)$. The solution of the adjoint equation (7) with $u(q)$ entering the right hand side is represented by $z(q)$ and the solution of the semidiscrete adjoint equation (14) with $u_k(q)$ on the right hand side by $z_k(q)$. The notations $u_{kh}(q)$ and $z_{kh}(q)$ are used analogously.

4.1 Estimates for the semidiscrete state solution

The following result for the error with respect to the $L^2(I,H)$ norm can be found for example as Theorem 5.1 in [17].

Theorem 4 *The error with respect to the $L^2(I,H)$ norm between the solution $u = u(q)$ of Equation* (2) *and the solution $u_k = u_k(q)$ of its semidiscretization* (13) *can be estimated by*

$$\|u - u_k\|_I \leq Ck^{r+1} \left\|\partial_t^{r+1} u\right\|_I$$

provided that the exact solution u is in $H^{r+1}(I,H)$.

In [28, Theorem 12.4], an estimate for the $L^\infty(I,H)$ norm error is given:

Theorem 5 *For the error between the solution $u = u(q)$ of Equation* (2) *and the solution $u_k = u_k(q)$ of its semidiscretization* (13) *we have the estimate*

$$\|u - u_k\|_{L^\infty(I,H)} \leq C\gamma(k)k^{r+1} \left\|\partial_t^{r+1} u\right\|_{L^\infty(I,H)}$$

with the logarithmic factor $\gamma(k) := |\log k|^{\frac{1}{2}} + 1$.

Note that this estimate requires Assumption 2 on the regularity of the temporal mesh.

4.2 Superconvergence of the reconstructed semidiscrete adjoint solution

Next we show that the reconstruction $\tilde{\pi}_k z_k(q)$ converges with one order more to the exact adjoint solution than $z_k(q)$ w. r. t. the $L^2(I,H)$ norm. We assume $r \geq 1$ since the dG(0) method does not have this superconvergence property.

Theorem 6 *For the reconstruction $\tilde{\pi}_k z_k$ computed from the solution $z_k = z_k(q)$ of the semidiscrete adjoint equation* (14) *with right hand side $u_k(q) - u_d$ the estimate*

$$\|z - \tilde{\pi}_k z_k\|_I \leq Ck^{r+2} \left(\left\|\partial_t^{r+1} \Delta z\right\|_I + \left\|\partial_t^{r+2} z\right\|_I + \left\|\partial_t^{r+1} u\right\|_I \right)$$

holds true where $z = z(q)$ is the solution of the adjoint equation (7) *with $u = u(q)$ entering the right hand side.*

As a preparation for the proof of Theorem 6, we need the following Lemma.

Lemma 5 *The reconstruction operator $\tilde{\pi}_k : X_k^r \to X_k^{r+1}$ is stable with respect to the $L^2(I,H)$ norm, that is, there is a constant C independent of k such that for any $v_k \in X_k^r$*

$$\|\tilde{\pi}_k v_k\|_I \le C \|v_k\|_I .$$

Proof Consider an interval I_m with $m \neq M$. Then

$$\|\tilde{\pi}_k v_k\|_{I_m} \le \|\tilde{\pi}_k v_k - v_k\|_{I_m} + \|v_k\|_{I_m}$$

According to [13, Lemma 2.2] we have for the first term

$$\|\tilde{\pi}_k v_k - v_k\|_{I_m}^2 = k_m \alpha_2^2 \|[v_k]_m\|^2 \le C k_m \left(\left\| v_{k,m}^+ \right\|^2 + \left\| v_{k,m}^- \right\|^2 \right),$$

where the constant α_2 is determined by the order r. The Lobatto quadrature rule with $r+2$ nodes is exact for polynomials of up to degree $2r+1$ and has positive weights (see, e. g., [21]). Let ω_j with $j = 0, \dots, r+1$ denote the weights of the Lobatto quadrature rule on the unit interval and $c_{m,j}$ the corresponding nodes transformed onto the interval I_m. Note that $c_{m,0} = t_{m-1}$ and $c_{m,r+1} = t_m$. Then we obtain for the left-sided limit $v_{k,m}^-$ at t_m the estimate

$$\begin{aligned} \frac{1}{\omega_{r+1}} \int_{I_m} \|v_k\|^2 \, \mathrm{d}t &= \frac{k_m}{\omega_{r+1}} \left(\omega_0 \left\| v_{k,m-1}^+ \right\|^2 + \sum_{j=1}^{r} \omega_j \left\| v_k(c_{m,j}) \right\|^2 + \omega_{r+1} \left\| v_{k,m}^- \right\|^2 \right) \\ &\ge k_m \left\| v_{k,m}^- \right\|^2 . \end{aligned}$$

For the right-sided limit $v_{k,m}^+$ we proceed similarly and get

$$k_m \left\| v_{k,m}^+ \right\|^2 \le \frac{k_m}{k_{m+1}} \frac{1}{\omega_0} \|v_k\|_{I_{m+1}}^2 .$$

Due to the assumption we made for the temporal mesh, the ratio $\frac{k_m}{k_{m+1}}$ is bounded by κ. So apart from the last subinterval we get

$$\sum_{m=1}^{M-1} \|\tilde{\pi}_k v_k - v_k\|_{I_m}^2 \le C \sum_{m=1}^{M-1} k_m \left(\|v_k\|_{I_m}^2 + \|v_k\|_{I_{m+1}}^2 \right) \le C \|v_k\|_I^2 . \tag{33}$$

The last interval I_M requires a separate treatment. Therefore we will show that there exists a constant $C > 0$ depending only on the mesh regularity parameter κ and the order of discretization r such that

$$\int_{I_M} \|\tilde{\pi}_k v_k\|^2 \, \mathrm{d}t \le C \int_{I_{M-1}} \|\tilde{\pi}_k v_k\|^2 \, \mathrm{d}t \tag{34}$$

holds. To see this, we transform the temporal integrals such that the integral over I_{M-1} is transformed to the negative unit interval $(-1,0)$, which gives

$$\int_{I_{M-1}} \|\tilde{\pi}_k v_k\|^2 \, \mathrm{d}t = \int_{-1}^{0} \frac{1}{k_{M-1}} \|\tilde{\pi}_k v_k(t_{M-1} + k_{M-1}\tau)\|^2 \, \mathrm{d}\tau$$

and

$$\int_{I_M} \|\tilde{\pi}_k v_k\|^2 \,\mathrm{d}t = \int_0^{\frac{k_M}{k_{M-1}}} \frac{1}{k_{M-1}} \|\tilde{\pi}_k v_k(t_{M-1} + k_{M-1}\tau)\|^2 \,\mathrm{d}\tau.$$

We define the polynomial $p\colon \mathbb{R} \to \mathbb{R}$ with maximum degree $2r+2$ by requiring $p(\tau) := \frac{1}{k_{M-1}} \|\tilde{\pi}_k v_k(t_{M-1} + k_{M-1}\tau)\|^2$ for $\tau \in \left(-1, \frac{k_M}{k_{M-1}}\right)$. Since the values of p are non-negative, the second integral can be estimated by

$$\int_0^{\frac{k_M}{k_{M-1}}} \frac{1}{k_{M-1}} \|\tilde{\pi}_k v_k(t_{M-1} + k_{M-1}\tau)\|^2 \,\mathrm{d}\tau = \int_0^{\frac{k_M}{k_{M-1}}} p(\tau)\,\mathrm{d}\tau \le \int_0^{\kappa} p(\tau)\,\mathrm{d}\tau$$

When considering both integrals as L^1 norms on the finite dimensional polynomial space $\mathscr{P}_{2r+2}(\mathbb{R})$ we see that those two norms are equivalent and in particular there exists a constant C such that

$$\int_{I_M} \|\tilde{\pi}_k v_k\|^2 \,\mathrm{d}t \le \int_0^{\kappa} p(\tau)\,\mathrm{d}\tau \le C \int_{-1}^{0} p(\tau)\,\mathrm{d}\tau = C \int_{I_{M-1}} \|\tilde{\pi}_k v_k\|^2 \,\mathrm{d}t.$$

The constant C depends only on κ and r. Hence the estimate (34) is shown and together with (33) we obtain the assertion. □

Proof of Theorem 6 The error $\|z - \tilde{\pi}_k z_k\|_I$ is split into two parts. We note that the identity $\tilde{\pi}_k \circ \pi_k = \tilde{\pi}_k$ holds true for arguments in $C(\bar{I}, V)$. Together with Lemma 5, we have

$$\|z - \tilde{\pi}_k z_k\|_I \le \|z - \tilde{\pi}_k z\|_I + \|\tilde{\pi}_k (z - z_k)\|_I \le \|z - \tilde{\pi}_k z\|_I + C\|\pi_k z - z_k\|_I. \tag{35}$$

The first term is bounded by the projection estimate (27), which results in

$$\|z - \tilde{\pi}_k z\|_I \le C k^{r+2} \left\|\partial_t^{r+2} z\right\|_I.$$

For the second term we pose a discrete dual equation (which is a forward equation again): find $w_k \in X_k^r$ satisfying

$$B(w_k, \varphi) = (\pi_k z - z_k, \varphi)_I \quad \forall \varphi \in X_k^r.$$

We choose $\varphi = \pi_k z - z_k \in X_k^r$ which gives us

$$\|\pi_k z - z_k\|_I^2 = B(w_k, \pi_k z - z_k) = B(w_k, \pi_k z - z) + B(w_k, z - z_k). \tag{36}$$

To estimate the first term on the right hand side we note that $(\pi_k z)_m^+ = z_m^+$ and hence the jump terms in representation (12) of the bilinear form B vanish. We get

$$B(w_k, \pi_k z - z) = \sum_{m=1}^{M} (\partial_t w_k, \pi_k z - z)_{I_m} + (\nabla w_k, \nabla(\pi_k z - z))_I. \tag{37}$$

For the first term we make use of the fact that the Radau quadrature formula with $r+1$ nodes is exact for polynomials up to degree $2r$ (see [1]). Using the auxiliary operator $\hat{\pi}_k$ we get the identity

$$(\partial_t w_k, \pi_k z)_{I_m} = (\partial_t w_k, \hat{\pi}_k z)_{I_m},$$

since on both sides we have a temporal integral over a polynomial with respect to time with the polynomial on the left having degree $2r-1$ and the polynomial on the right having degree $2r$. Hence, both can be evaluated exactly with Radau's integration formula, which gives the same result in both cases. With the above identity we obtain for the first term of (37)

$$\sum_{m=1}^{M}(\partial_t w_k, \pi_k z - z)_{I_m} = \sum_{m=1}^{M}(\partial_t w_k, \hat{\pi}_k z - z)_{I_m} \leq \left(\sum_{m=1}^{M}\|\partial_t w_k\|_{I_m}^2\right)^{\frac{1}{2}} \|\hat{\pi}_k z - z\|_I .$$

With the interpolation estimate (26) for $\hat{\pi}_k$ and the stability estimate for $\partial_t w_k$ from Theorem 1, this gives the estimate

$$\sum_{m=1}^{M}(\partial_t w_k, \pi_k z - z)_{I_m} \leq Ck^{r+2}\left\|\partial_t^{r+2} z\right\|_I \|\pi_k z - z_k\|_I \tag{38}$$

for the time derivative term of (37).

To estimate the second term on the right hand side of (37), we split it into two parts using the L^2 projection Π_k^0 into the space X_k^0,

$$(\nabla w_k, \nabla(\pi_k z - z))_I = (\nabla(w_k - \Pi_k^0 w_k), \nabla(\pi_k z - z))_I + (\nabla \Pi_k^0 w_k, \nabla(\pi_k z - z))_I. \tag{39}$$

To estimate the first term, we integrate by parts with respect to the spatial domain and obtain

$$(\nabla(w_k - \Pi_k^0 w_k), \nabla(\pi_k z - z))_I = (w_k - \Pi_k^0 w_k, -\Delta(\pi_k z - z))_I$$

The temporal interpolation operator π_k commutes with the Laplacian and together with the error estimates (23) and (25) for Π_k^0 and π_k, we get

$$\begin{aligned}(\nabla(w_k - \Pi_k^0 w_k), \nabla(\pi_k z - z))_I &\leq C\left(\sum_{m=1}^{M} k_m^2 \|\partial_t w_k\|_{I_m}^2\right)^{\frac{1}{2}} \|\pi_k \Delta z - \Delta z\|_I \\ &\leq C\left(\sum_{m=1}^{M} k_m^2 \|\partial_t w_k\|_{I_m}^2\right)^{\frac{1}{2}} k^{r+1}\left\|\partial_t^{r+1}\Delta z\right\|_I .\end{aligned}$$

The stability estimate from Theorem 1 gives the desired estimate for the first term of (39),

$$(\nabla(w_k - \Pi_k^0 w_k), \nabla(\pi_k z - z))_I \leq Ck^{r+2}\|\pi_k z - z_k\|_I \left\|\partial_t^{r+1}\Delta z\right\|_I . \tag{40}$$

For the second term we have, since the temporal L^2 projection commutes with spatial derivatives,

$$(\nabla \Pi_k^0 w_k, \nabla(\pi_k z - z))_I = -\sum_{m=1}^{M}(\Pi_k^0 \Delta w_k, \pi_k z - z)_{I_m}.$$

The product $(\Pi_k^0 \Delta w_k, \pi_k z)$ is a polynomial of degree $r < 2r$ with respect to time. Thus with the same reasoning as we used when estimating the temporal derivative term, $\pi_k z$ can be replaced by $\hat{\pi}_k z$ without changing the value of the above expression.

Subsequently applying the interpolation estimate (26) and taking the continuity of Π_k^0 into consideration gives

$$\begin{aligned}(\nabla \Pi_k^0 w_k, \nabla(\pi_k z - z))_I &= -\sum_{m=1}^{M} (\Pi_k^0 \Delta w_k, \hat{\pi}_k z - z)_{I_m} \\ &\le C \sum_{m=1}^{M} k_m^{r+2} \|\Delta w_k\|_{I_m} \left\|\partial_t^{r+2} z\right\|_{I_m} \\ &\le C k^{r+2} \|\pi_k z - z_k\|_I \left\|\partial_t^{r+2} z\right\|_I .\end{aligned} \tag{41}$$

In the last step we used again Theorem 1. Plugging equations (38), (39), (40), and (41) into (37) we get for the first term of (36)

$$B(w_k, \pi_k z - z) \le C k^{r+2} \|\pi_k z - z_k\|_I \left(\left\|\partial_t^{r+2} z\right\|_I + \left\|\partial_t^{r+1} \Delta z\right\|_I\right). \tag{42}$$

We use the projection operator P_k into the semidiscrete space to split the second term on the right hand side of (36) into two parts

$$B(w_k, z - z_k) = (w_k, u - u_k)_I = (w_k, u - P_k u)_I + (w_k, P_k u - u_k)_I. \tag{43}$$

Due to Condition (21a), we have $(\Pi_k^0 w_k, u - P_k u)_I = 0$ and hence the first term can be bounded using the interpolation estimates (23) and (24) for Π_k^0 and P_k respectively, which gives

$$(w_k, u - P_k u)_I = (w_k - \Pi_k^0 w_k, u - P_k u)_I \le C k^{r+2} \left(\sum_{m=1}^{M} \|\partial_t w_k\|_{I_m}^2\right)^{\frac{1}{2}} \left\|\partial_t^{r+1} u\right\|_I .$$

For the second term we use another duality argument. Let y_k be the solution of the semidiscrete dual equation

$$B(\varphi, y_k) = (\varphi, w_k)_I \quad \forall \varphi \in X_k^r. \tag{44}$$

Then testing with $P_k u - u_k$ results in

$$(w_k, P_k u - u_k)_I = B(P_k u - u_k, y_k) = B(P_k u - u, y_k) + B(u - u_k, y_k), \tag{45}$$

where the last term vanishes due to Galerkin orthogonality. For the first term we expand the bilinear form in its dual formulation and note that the jump terms vanish due to Condition (21b). Since the time derivative $\partial_t y_k|_{I_m}$ is in $\mathscr{P}_{r-1}(I_m, V)$, the time derivative terms vanish as well with Condtion (21a). This leaves only the spatial operator, that is,

$$B(P_k u - u, y_k) = (\nabla(P_k u - u), \nabla y_k)_I = (P_k u - u, -\Delta y_k + \Pi_k^0 \Delta y_k)_I.$$

In the last step we performed integration by parts and used the orthogonality condition (21a). The interpolation estimates (23) and (24) together with the stability estimates Theorem 2 for Equation (44) and Theorem 1 for the equation for w_k finally result in

$$\begin{aligned}B(P_k u - u, y_k) &\le C k^{r+2} \left\|\partial_t^{r+1} u\right\|_I \left(\sum_{m=1}^{M} \|\partial_t \Delta y_k\|_{I_m}\right)^{\frac{1}{2}} \le C k^{r+2} \left\|\partial_t^{r+1} u\right\|_I \|\Delta w_k\|_I \\ &\le C k^{r+2} \left\|\partial_t^{r+1} u\right\|_I \|\pi_k z - z_k\|_I .\end{aligned}$$

Plugging this estimate into Equation (45) gives an estimate for the second term on the right hand side of (43). Collecting all estimates shows the claim. □

Remark 3 Analogously to $\tilde{\pi}_k$, but backwards in time, a reconstruction operator $\tilde{\pi}_k^u$ for the state solution can be defined. In the same way as in the proof of Theorem 6, it can be shown that for the reconstruction $\tilde{\pi}_k^u u_k$ of the semidiscrete state, the estimate

$$\|u - \tilde{\pi}_k^u u_k\|_I \leq Ck^{r+2} \left(\left\|\partial_t^{r+1} \Delta u\right\|_I + \left\|\partial_t^{r+2} u\right\|_I \right) \tag{46}$$

holds if the exact solution u is in $H^{r+2}(I,H) \cap H^{r+1}(I,H^2(\Omega) \cap V)$. For our optimization problem, we cannot apply this estimate, even in the case $r = 1$ since the optimal state $\bar{u}$ will be in $H^2(I,H)$ but in general not in $H^3(I,H)$. However we mention it due to its potential use for improving numerical solutions of PDEs outside the optimization context and for giving a rigorous justification for the use of the reconstruction operator, e. g., in evaluating dual-weighted-residual based a posteriori error estimators.

4.3 Error analysis for the spatial discretization

Theorem 7 *For the solution u_{kh} of the discrete state equation* (28) *and the semidiscrete solution u_k satisfying* (13)*, the a priori error estimate*

$$\|u_k - u_{kh}\|_I \leq Ch^{s+1} \left\|\nabla^{s+1} u_k\right\|_I$$

holds true with the constant C independent of k and h if $u_k \in L^2(I,H^{s+1})$.

If additionally $z_k \in L^2(I,H^{s+1})$, the solutions z_k of the semidiscrete adjoint equation (14) *with right hand side $u_k - u_d$ and z_{kh} of the discrete adjoint equation* (30) *with right hand side $u_{kh} - u_d$ fulfill the estimate*

$$\|z_k - z_{kh}\|_I \leq Ch^{s+1} \left(\left\|\nabla^{s+1} u_k\right\|_I + \left\|\nabla^{s+1} z_k\right\|_I \right).$$

Proof Both parts of the claim are shown as Theorem 5.5 and as step in the proof of Lemma 6.2 in [17] respectively. □

Corollary 2 *The error $\tilde{\pi}_k z_k - \tilde{\pi}_k z_{kh}$ between the reconstructed semidiscrete and discrete adjoint solution satisfies the a priori bound*

$$\|\tilde{\pi}_k z_k - \tilde{\pi}_k z_{kh}\|_I \leq Ch^{s+1} \left(\left\|\nabla^{s+1} u_k\right\|_I + \left\|\nabla^{s+1} z_k\right\|_I \right).$$

Proof Applying Lemma 5 to the left hand side of the estimate reduces it to the statement of Theorem 7. □

5 Error analysis for the constrained optimal control problem

Now we turn to the analysis of the control-constrained optimal control problem (1). Looking at the temporal and spatial regularity of the optimal state and adjoint state as discussed in Lemma 2, we observe that the error estimates in Theorem 6 and Corrollary 2 can only be applied for the case $r = s = 1$. Therefore, in this section we restrict our considerations to first order elements in both, time and space.

Remark 4 Choosing an order $s > 1$ for the spatial discretization can lead to improved convergence with respect to h if we require a domain Ω with smooth boundary and enforce additional compatibility conditions on the data.

5.1 Time discretization

As a preparation for our main result we show almost second order convergence of the control with respect to the $L^\infty(I,\mathbb{R}^D)$ norm. Therefore, we proceed as in [10] by first proving convergence with order $\mathscr{O}(k^2)$ with respect to the $L^2(I,H)$ norm.

Lemma 6 *For the error between the solution $\bar{q}$ of the continuous optimal control problem* (1) *and the solution $\bar{q}_k$ of the semidiscrete problem* (16) *we have the estimate*

$$|\bar{q}-\bar{q}_k|_I \le Ck^2\left(\alpha^{-\frac{1}{2}}\left\|\partial_t^2\bar{u}\right\|_I + \alpha^{-1}\left\|\partial_t^2\bar{z}\right\|_I\right)$$

and for the corresponding state error we obtain

$$\|\bar{u}-\bar{u}_k\|_I \le Ck^2\left(\left\|\partial_t^2\bar{u}\right\|_I + \alpha^{-\frac{1}{2}}\left\|\partial_t^2\bar{z}\right\|_I\right).$$

Proof Testing the optimality condition (6) with $\delta q = \bar{q}_k$, its semidiscrete counterpart (18) with $\bar{q}$, and adding up the results gives

$$\begin{aligned}\alpha|\bar{q}-\bar{q}_k|_I^2 &\le \langle G^*(\bar{z}-\bar{z}_k), \bar{q}_k-\bar{q}\rangle_I \\ &= (\bar{z}-z_k(\bar{u}), G(\bar{q}_k-\bar{q}))_I + (z_k(\bar{u})-\bar{z}_k, G(\bar{q}_k-\bar{q}))_I.\end{aligned} \tag{47}$$

Here, $z_k(\bar{u})$ denotes the solution of the semidiscrete adjoint with $\bar{u}$ entering the right hand side. For the second term on the right, we apply the semidiscrete state equation followed by Galerkin orthogonality and the semidiscrete adjoint and obtain

$$\begin{aligned}(z_k(\bar{u})-\bar{z}_k, G(\bar{q}_k-\bar{q}))_I &= B(\bar{u}_k-\bar{u}, z_k(\bar{u})-\bar{z}_k) \\ = B(\bar{u}_k-u_k(\bar{q}), z_k(\bar{u})-\bar{z}_k) &= (\bar{u}_k-u_k(\bar{q}), \bar{u}-\bar{u}_k)_I.\end{aligned}$$

We plug this result into (47) and add on both sides $\|\bar{u}-\bar{u}_k\|_I^2$. With the scaled version of Young's inequality, this yields

$$\begin{aligned}\alpha|\bar{q}-\bar{q}_k|_I^2 + \|\bar{u}-\bar{u}_k\|_I^2 &\le (\bar{z}-z_k(\bar{u}), G(\bar{q}_k-\bar{q}))_I \\ &\quad + (\bar{u}_k-u_k(\bar{q}), \bar{u}-\bar{u}_k)_I + (\bar{u}-\bar{u}_k, \bar{u}-\bar{u}_k)_I \\ &\le \|G\|\,\|\bar{z}-z_k(\bar{u})\|_I\,|\bar{q}-\bar{q}_k|_I + \|\bar{u}-u_k(\bar{q})\|_I\,\|\bar{u}-\bar{u}_k\|_I \\ &\le \frac{1}{2\alpha}\|G\|^2\,\|\bar{z}-z_k(\bar{u})\|_I^2 + \frac{\alpha}{2}|\bar{q}-\bar{q}_k|_I^2 \\ &\quad + \frac{1}{2}\|\bar{u}-u_k(\bar{q})\|_I^2 + \frac{1}{2}\|\bar{u}-\bar{u}_k\|_I^2.\end{aligned}$$

Therefore we have with the a priori estimate Theorem 4, applied once to the state equation and once to the adjoint equation

$$\frac{\alpha}{2}|\bar{q}-\bar{q}_k|_I^2+\frac{1}{2}\|\bar{u}-\bar{u}_k\|_I^2 \le \frac{1}{2\alpha}\|G\|^2\|\bar{z}-z_k(\bar{u})\|_I^2+\frac{1}{2}\|\bar{u}-u_k(\bar{q})\|_I^2$$
$$\le Ck^4\left(\left\|\partial_t^2\bar{u}\right\|_I^2+\alpha^{-1}\left\|\partial_t^2\bar{z}\right\|_I^2\right)$$

which results in the desired estimates

$$|\bar{q}-\bar{q}_k|_I \le Ck^2\left(\alpha^{-\frac{1}{2}}\left\|\partial_t^2\bar{u}\right\|_I+\alpha^{-1}\left\|pt^2\bar{z}\right\|_I\right)$$

and

$$\|\bar{u}-\bar{u}_k\|_I^2 \le Ck^2\left(\left\|\partial_t^2\bar{u}\right\|_I+\alpha^{-\frac{1}{2}}\left\|pt^2\bar{z}\right\|_I\right).$$

□

Lemma 7 *For the error between the solution $\bar{q}$ of the continuous optimal control problem* (1) *and the solution $\bar{q}_k$ of the semidiscrete problem* (16) *with respect to the $L^\infty(I,\mathbb{R}^D)$ norm, the estimate*

$$\|\bar{q}-\bar{q}_k\|_{L^\infty(I,\mathbb{R}^D)} \le \alpha^{-1}k^2\left\{\gamma(k)\left\|\partial_t^2\bar{z}\right\|_{L^\infty(I,H)}+\left\|\partial_t^2\bar{u}\right\|_I+\alpha^{-\frac{1}{2}}\left\|\partial_t^2\bar{z}\right\|_I\right\}$$

holds true with the logarithmic factor $\gamma(k)=|\log k|^{\frac{1}{2}}+1$ as introduced in Theorem 5.

Proof Taking into account that $\left\|P_{Q_{\text{ad}}}(f)-P_{Q_{\text{ad}}}(g)\right\|_{L^\infty(I,\mathbb{R}^D)} \le \|f-g\|_{L^\infty(I,\mathbb{R}^D)}$, the optimality conditions (8) and (19) yield

$$\|\bar{q}-\bar{q}_k\|_{L^\infty(I,\mathbb{R}^D)} \le \alpha^{-1}\|G\|\,\|\bar{z}-\bar{z}_k\|_{L^\infty(I,H)}. \tag{48}$$

We introduce an auxiliary adjoint solution $\tilde{z}_k \in X_k^r$ satisfying

$$B(\varphi,\tilde{z}_k)=(\varphi,\bar{u}-u_d)_I \quad \forall\varphi\in X_k^r$$

and split the adjoint error into

$$\|\bar{z}-\bar{z}_k\|_{L^\infty(I,H)} \le \|\bar{z}-\tilde{z}_k\|_{L^\infty(I,H)}+\|\tilde{z}_k-\bar{z}_k\|_{L^\infty(I,H)}.$$

The first term can be estimated by the supremum norm estimate from Theorem 5 applied backward in time, which gives

$$\|\bar{z}-\tilde{z}_k\|_{L^\infty(I,H)} \le C\gamma(k)k^2\left\|\partial_t^2\bar{z}\right\|_{L^\infty(I,H)}. \tag{49}$$

For the second term we apply the stability estimate from Lemma 3 and the state error bound from Lemma 6 to obtain

$$\|\tilde{z}_k-\bar{z}_k\|_{L^\infty(I,H)} \le C\|\bar{u}-\bar{u}_k\|_I \le Ck^2\left(\left\|\partial_t^2\bar{u}\right\|_I+\alpha^{-\frac{1}{2}}k^2\left\|\partial_t^2\bar{z}\right\|_I\right). \tag{50}$$

Plugging the inequalities (49) and (50) into (48) shows the claim. □

Definition 1 Let $\mathscr{M} := \{1,2,\dots,M\}$ denote the set of all time indices. For a given $z \in X \cup X_k^r$ we define the sets of active and inactive indices for each of the D components of the resulting control by

$$\mathscr{A}_i(z) := \left\{ m \in \mathscr{M} \,\middle|\, \exists t \in I_m : \left(-\alpha^{-1} G^* z(t)\right)_i > q_i^b \vee \left(-\alpha^{-1} G^* z(t)\right)_i < q_i^a \right\} \text{ and}$$
$$\mathscr{I}_i(z) := \left\{ m \in \mathscr{M} \,\middle|\, \exists t \in I_m : \left(-\alpha^{-1} G^* z(t)\right)_i \in (q_i^a, q_i^b) \right\}$$

respectively where $i \in \{1,\dots,D\}$. For convenience we also define the sets

$$\mathscr{E}_i(z) := \left\{ m \in \mathscr{M} \,\middle|\, \forall t \in I_m : \left(-\alpha^{-1} G^* z(t)\right)_i = q_i^b \vee \left(-\alpha^{-1} G^* z(t)\right)_i = q_i^a \right\}.$$

Note that $\mathscr{A}_i(z) \cup \mathscr{I}_i(z) \cup \mathscr{E}_i(z) = \mathscr{M}$ for any $i \in \{1,\dots,D\}$ and any $z \in X \cup X_k^r$.

With this notation we can introduce the set $\mathscr{K}$ of critical indices collecting all intervals where at least one component is both active and inactive for either of the functions $\bar{z}$ and $\pi_k \bar{z}$. The remaining indices are collected in the set $\mathscr{R}$.

$$\mathscr{K} := \bigcup_{i=1}^{D} \left\{ [\mathscr{A}_i(\bar{z}) \cap \mathscr{I}_i(\bar{z})] \cup [\mathscr{A}_i(\pi_k \bar{z}) \cap \mathscr{I}_i(\pi_k \bar{z})] \right\},$$
$$\mathscr{R} := \mathscr{M} \setminus \mathscr{K}.$$

Assumption 3 We assume that the set $\mathscr{K}$ satisfies

$$\sum_{m \in \mathscr{K}} |I_m| \le Ck$$

for a constant C independent from k.

Remark 5 Similar assumptions are used in [20,25,4,18,19]. The assumption is satisfied if the boundary of the active set for the continuous problem consists of finitely many points and additionally the time derivative of $\left(-\alpha^{-1} G^* \bar{z}\right)_i$ in those points has non-zero value.

With this assumption, we get the following estimate for the reconstructed semidiscrete adjoint solution.

Theorem 8 *Let the Assumptions 1, 2, and 3 be fulfilled. Then the error between the optimal adjoint $\bar{z}$ of the continuous problem* (1) *and the piecewise quadratic reconstruction $\tilde{\pi}_k \bar{z}_k$ of the adjoint for the semidiscrete problem* (16) *satisfies*

$$\|\bar{z} - \tilde{\pi}_k \bar{z}_k\|_I \le C(\alpha) k^3 \left\{ \gamma(k) \left\|\partial_t^2 \bar{z}\right\|_{L^\infty(I,H)} + \left\|\partial_t^2 \bar{u}\right\|_I + \left\|\partial_t^2 \bar{z}\right\|_I + \left\|\partial_t^3 \bar{z}\right\|_I + \left\|\partial_t^2 \Delta \bar{z}\right\|_I \right\}$$

where the constant $C(\alpha)$ can be estimated by $C(\alpha) \le C\left(1 + \alpha^{-\frac{5}{2}}\right)$ and $\gamma(k)$ is given by $\gamma(k) = |\log k|^{\frac{1}{2}} + 1$.

With Lipschitz continuity of $P_{Q_{\mathrm{ad}}}$ we immediately get the main result of this subsection.

Corollary 3 *Let $(\bar{q}_k, \bar{u}_k, \bar{z}_k)$ be the solutions of the semidiscrete optimization problem. Then for the control solution $\tilde{q}_k$ obtained by the post-processing step*

$$\tilde{q}_k = P_{Q_{ad}}\left(-\alpha^{-1} G^* \tilde{\pi}_k \bar{z}_k\right) \tag{51}$$

the estimate

$$|\bar{q} - \tilde{q}_k|_I \leq C(\alpha) k^3 \left\{ \gamma(k) \left\|\partial_t^2 \bar{z}\right\|_{L^\infty(I,H)} + \left\|\partial_t^2 \bar{u}\right\|_I + \left\|\partial_t^2 \bar{z}\right\|_I + \left\|\partial_t^3 \bar{z}\right\|_I + \left\|\partial_t^2 \Delta \bar{z}\right\|_I \right\}$$

holds true with $C(\alpha) \leq C\left(\alpha^{-1} + \alpha^{-\frac{7}{2}}\right)$.

A key ingredient for the proof of Theorem 8 are the following two auxiliary controls which are constructed in a similar fashion as in [24].

Definition 2 The function $p_k \in L^2(I, \mathbb{R}^D)$ is given piecewise as

$$p_k|_{I_m} = \begin{cases} \bar{q}_k, & \text{if } m \in \mathscr{K}, \\ \pi_k \bar{q}, & \text{if } m \in \mathscr{R}, \end{cases}$$

that is, it is identical to the semidiscrete solution on the critical set and interpolates the exact solution on the remaining intervals. In particular, this function is a linear polynomial on each subinterval. Additionally, we will use the function $\hat{p}_k \in L^2(I, \mathbb{R}^D)$ given by

$$\hat{p}_k|_{I_m} = \begin{cases} \bar{q}_k, & \text{if } m \in \mathscr{K}, \\ \bar{q}, & \text{if } m \in \mathscr{R}. \end{cases}$$

Lemma 8 *For the difference between the semidiscrete adjoint states computed from the exact control $\bar{q}$ and the auxiliary control p_k, the estimate*

$$\|z_k(\bar{q}) - z_k(p_k)\|_I \leq C(\alpha) k^3 \left(\gamma(k) \left\|\partial_t^2 \bar{z}\right\|_{L^\infty(I,H)} + \left\|\partial_t^2 \bar{u}\right\|_I + \left\|\partial_t^2 \bar{z}\right\|_I + \left\|\partial_t^3 \bar{z}\right\|_I \right)$$

holds true with $C(\alpha) \leq C\alpha^{-1}\left(1 + \alpha^{-\frac{1}{2}}\right)$ and $\gamma(k) = |\log k|^{\frac{1}{2}} + 1$.

Proof We start with an estimate for the difference between the corresponding semidiscrete states $u_k(\bar{q})$ and $u_k(p_k)$. With the semidiscrete adjoint (14) and the semidiscrete state equation (13) we get

$$\begin{aligned} \|u_k(\bar{q}) - u_k(p_k)\|_I^2 &= B(u_k(\bar{q}) - u_k(p_k), z_k(\bar{q}) - z_k(p_k)) \\ &= (G(\bar{q} - p_k), z_k(\bar{q}) - z_k(p_k))_I. \end{aligned} \tag{52}$$

For convenience, we introduce the abbreviation $v_k := z_k(\bar{q}) - z_k(p_k)$. Then, using the auxiliary control $\hat{p}_k$ and the L^2 projection Π_k^0 onto the space of piecewise constant functions we split the right hand side into

$$\begin{aligned} (G(\bar{q} - p_k), v_k)_I = (G(\bar{q} - \hat{p}_k), v_k)_I &+ (G(\hat{p}_k - p_k), v_k - \Pi_k^0 v_k)_I \\ &+ (G(\hat{p}_k - p_k), \Pi_k^0 v_k)_I. \end{aligned} \tag{53}$$

Each of the three terms on the right hand side can be estimated separately. For the first term, we obtain since $\bar{q}$ and $\hat{p}_k$ agree on I_m for $m \in \mathscr{R}$

$$
\begin{aligned}
(G\bar{q} - G\hat{p}_k, v_k)_I &= \sum_{m\in\mathscr{K}} \int_{I_m} (G\bar{q} - G\hat{p}_k, v_k)\,\mathrm{d}t \\
&\leq \sum_{m\in\mathscr{K}} |I_m|\,\|G\|\,\|\bar{q} - \hat{p}_k\|_{L^\infty(I_m,\mathbb{R}^D)}\,\|v_k\|_{L^\infty(I_m,H)} \\
&\leq C \sum_{m\in\mathscr{K}} |I_m|\,\|\bar{q} - \bar{q}_k\|_{L^\infty(I,\mathbb{R}^D)}\,\|v_k\|_{L^\infty(I,H)}\,.
\end{aligned}
$$

Plugging the estimate from Lemma 7, the maximum norm stability estimate from Lemma 3 and Assumption 3 into this inequality yields

$$
\begin{aligned}
&(G(\bar{q} - \hat{p}_k), v_k)_I \\
&\leq C\alpha^{-1}k^3 \left(\gamma(k)\left\|\partial_t^2 \bar{z}\right\|_{L^\infty(I,H)} + \left\|\partial_t^2 \bar{u}\right\|_I + \alpha^{-\frac{1}{2}}\left\|\partial_t^2 \bar{z}\right\|_I\right)\|u_k(\bar{q}) - u_k(p_k)\|_I\,. \qquad (54)
\end{aligned}
$$

The second term of (53) vanishes on all intervals in the critical set $\mathscr{K}$ since $p_k = \hat{p}_k$ there. Hence we have

$$
\begin{aligned}
(G(\hat{p}_k - p_k), v_k - \Pi_k^0 v_k)_I &\leq \|G\| \sum_{m\in\mathscr{R}} |\bar{q} - \pi_k\bar{q}|_{I_m}\left\|v_k - \Pi_k^0 v_k\right\|_{I_m} \\
&\leq Ck\left(\sum_{m\in\mathscr{R}}\sum_{i=1}^{D}\|\bar{q}_i - \pi_k\bar{q}_i\|^2_{L^2(I_m)}\right)^{\frac{1}{2}}\left(\sum_{m=1}^{M}\|\partial_t v_k\|^2_{I_m}\right)^{\frac{1}{2}}. \qquad (55)
\end{aligned}
$$

To estimate the factor involving the control, we note that for each component $\bar{q}_i$ of the optimal control, all interval indices in $\mathscr{R}$ are contained either in the active set $\mathscr{A}_i(\bar{z}) \setminus \mathscr{I}_i(\bar{z})$, the inactive set $\mathscr{I}_i(\bar{z}) \setminus \mathscr{A}_i(\bar{z})$, or the rest set $\mathscr{E}_i(\bar{z})$ belonging to the index i. For the indices in the active set and the rest set, component $\bar{q}_i$ is constant on the corresponding intervals. Hence the difference $(\bar{q} - \pi_k\bar{q})_i$ vanishes on those intervals. On the inactive intervals, the component is in $H^3(I,\mathbb{R})$ since the relationship $\bar{q}_i = -\alpha^{-1}(G^*\bar{z})_i$ holds and $\bar{z}$ is in $H^3(I,H)$. Hence we can apply an interpolation estimate for π_k and obtain

$$
\begin{aligned}
\Big(\sum_{m\in\mathscr{R}}\sum_{i=1}^{D}\|\bar{q}_i - \pi_k\bar{q}_i\|^2_{L^2(I_m)}\Big)^{\frac{1}{2}} &= \Big(\sum_{i=1}^{D}\sum_{m\in\mathscr{R}\cap\mathscr{I}_i(\bar{z})}\|\bar{q}_i - \pi_k\bar{q}_i\|^2_{L^2(I_m)}\Big)^{\frac{1}{2}} \\
&\leq Ck^2\Big(\sum_{i=1}^{D}\sum_{m\in\mathscr{R}\cap\mathscr{I}_i(\bar{z})}\left\|\partial_t^2\bar{q}_i\right\|^2_{L^2(I_m)}\Big)^{\frac{1}{2}} \\
&\leq C\alpha^{-1}k^2\Big(\sum_{m=1}^{M}\left\|\partial_t^2\bar{z}\right\|^2_{I_m}\Big)^{\frac{1}{2}}.
\end{aligned}
$$

Plugging this estimate into Equation (55) yields for the second term of (53) the bound

$$
\begin{aligned}
(G(\hat{p}_k - p_k), v_k - \Pi_k^0 v_k)_I &\leq C\alpha^{-1}k^3\left\|\partial_t^2 z\right\|_I\left(\sum_{m=1}^{M}\|\partial_t v_k\|^2_{I_m}\right)^{\frac{1}{2}} \\
&\leq C\alpha^{-1}k^3\left\|\partial_t^2 z\right\|_I\|u_k(\bar{q}) - u_k(p_k)\|_I\,.
\end{aligned} \qquad (56)
$$

In the last step the stability estimate from Corollary 1 was used.

For estimating the third term on the right hand side of (53), we introduce a third auxiliary control $\tilde{p}_k$ given by

$$\tilde{p}_k|_{I_m} = \begin{cases} \bar{q}_k, & \text{if } m \in \mathscr{K}, \\ \tilde{\pi}_k \bar{q}, & \text{if } m \in \mathscr{R}. \end{cases}$$

We observe that $(Gp_k, \Pi_k^0 v_k)$ is a polynomial of degree one with respect to time on each interval I_m with $m \in \mathscr{R}$. Since the two point Radau quadrature formula integrates polynomials up to degree two exactly, we have the identity $(Gp_k, \Pi_k^0 v_k)_{I_m} = (G\tilde{p}_k, \Pi_k^0 v_k)_{I_m}$ and hence

$$(G(\hat{p}_k - p_k), \Pi_k^0 v_k)_I = \sum_{m \in \mathscr{R}} (G(\hat{p}_k - \tilde{p}_k), \Pi_k^0 v_k)_{I_m} \leq C\alpha^{-1} k^3 \left\| \partial_t^3 \bar{z} \right\|_I \|v_k\|_I .$$

The last step involves the interpolation estimate (26) which can be applied since for intervals with index in $\mathscr{R}$ every component of the control $\bar{q}$ is either constant or in $H^3(I_m, \mathbb{R})$. Using the stability estimate from Corollary 1 for the semidiscrete adjoint, we obtain as estimate for the third term

$$(G(\hat{p}_k - p_k), \Pi_k^0 v_k)_I \leq C\alpha^{-1} k^3 \left\| \partial_t^3 \bar{z} \right\|_I \|u_k(\bar{q}) - u_k(p_k)\|_I . \tag{57}$$

Putting Equations (52), (53), (54), (56), and (57) together and dividing everything by $\|u_k(\bar{q}) - u_k(p_k)\|_I$ gives

$$\begin{aligned} &\|u_k(\bar{q}) - u_k(p_k)\|_I \\ &\quad \leq C\alpha^{-1} k^3 \left\{ \gamma(k) \left\| \partial_t^2 \bar{z} \right\|_{L^\infty(I,H)} + \left\| \partial_t^2 \bar{u} \right\|_I + (1 + \alpha^{-\frac{1}{2}}) \left\| \partial_t^2 \bar{z} \right\|_I + \left\| \partial_t^3 \bar{z} \right\|_I \right\} . \end{aligned}$$

The desired estimate for $\|z_k(\bar{q}) - z_k(p_k)\|_I$ is obtained from this inequality by means of the stability estimate Corollary 1 for the semidiscrete adjoint. □

With these preparations we can prove the main result of this section:

Proof of Theorem 8 We start by splitting the error into the two parts

$$\|\bar{z} - \tilde{\pi}_k \bar{z}_k\|_I \leq \|\bar{z} - \tilde{\pi}_k z_k(\bar{q})\|_I + \|\tilde{\pi}_k (z_k(\bar{q}) - \bar{z}_k)\|_I .$$

For the first term, we use the estimate in Theorem 6. To bound the second term, we exploit the L^2 stability of the reconstruction operator $\tilde{\pi}_k$ and split further

$$\begin{aligned} \|\tilde{\pi}_k (z_k(\bar{q}) - \bar{z}_k)\|_I &\leq C \|z_k(\bar{q}) - \bar{z}_k\|_I \\ &\leq C \left(\|z_k(\bar{q}) - z_k(p_k)\|_I + \|z_k(p_k) - \bar{z}_k\|_I \right) . \end{aligned}$$

The first term on the right hand side can be estimated with Lemma 8 and for the second term the stability estimates from Corollary 1 and Theorem 1 give

$$\|z_k(p_k) - \bar{z}_k\|_I \leq C |p_k - \bar{q}_k|_I . \tag{58}$$

To estimate the term on the right hand side, we first observe that the inequality

$$\langle G^* \pi_k \bar{z} + \alpha p_k, \bar{q}_k - p_k \rangle_I \geq 0 \tag{59}$$

holds true. This can be seen as follows: we write

$$\begin{aligned}\langle G^*\pi_k\bar z+\alpha p_k,\bar q_k-p_k\rangle_I &= \sum_{m\in\mathscr{K}}\langle G^*\pi_k\bar z+\alpha p_k,\bar q_k-p_k\rangle_{I_m}\\ &\quad+\sum_{m\in\mathscr{R}}\sum_{i=1}^{D}((G^*\pi_k\bar z+\alpha p_k)_i,(\bar q_k-p_k)_i)_{L^2(I_m)}\end{aligned}$$

and show that each of the addends is non-negative. For the critical set $\mathscr{K}$ the factor $\bar q_k-p_k$ vanishes on the corresponding intervals due to the definition of p_k. For the remaining intervals we have to distinguish whether constraints are active or not, that is, for each component $\bar q_i$ of the control we have to consider the cases $m\in\mathscr{I}_i(\bar z)\setminus\mathscr{A}_i(\bar z)$, $m\in\mathscr{A}_i(\bar z)\setminus\mathscr{I}_i(\bar z)$, and $m\in\mathscr{E}_i(\bar z)$. If $m\in\mathscr{I}_i(\bar z)\setminus\mathscr{A}_i(\bar z)$, then we have point-wise $(G^*\bar z+\alpha\bar q)_i=0$ on the interval I_m. Therefore, on I_m the interpolant $(G^*\pi_k\bar z+\alpha p_k)_i=\pi_k(G^*\bar z+\alpha\bar q)_i$ vanishes.

In the other two cases, one of the constraints is active, that is, the component $\bar q_i$ has either the constant value q_i^a or the constant value q_i^b on I_m. So in particular we have $p_k=\bar q$. Since $\pi_k\bar z$ interpolates $\bar z$ in two points on I_m, we know that m is also in $\mathscr{A}_i(\pi_k\bar z)$ or $\mathscr{E}_i(\pi_k\bar z)$ respectively and therefore according to the definition of the set $\mathscr{R}$ not in $\mathscr{I}_i(\pi_k\bar z)$. From the optimality condition (8), the value of $-\alpha^{-1}(G^*\bar z)_i$ is either less or equal q_i^a or greater or equal q_i^b on I_m and hence we get for the projection

$$-\alpha^{-1}(G^*\pi_k\bar z)_i\begin{cases}\le q_i^a, & \text{if } \bar q_i=q_i^a,\\ \ge q_i^b, & \text{if } \bar q_i=q_i^b.\end{cases}$$

Therefore on the interval I_m, we have point-wise

$$(G^*\pi_k\bar z+\alpha p_k)_i\begin{cases}\ge 0, & \text{if } \bar q_i=q_i^a,\\ \le 0, & \text{if } \bar q_i=q_i^b\end{cases}$$

and, since $\bar q_k$ is in the admissible set and $p_k=\bar q$ for $m\in\mathscr{R}$,

$$(\bar q_k-p_k)_i\begin{cases}\ge 0, & \text{if } \bar q_i=q_i^a,\\ \le 0, & \text{if } \bar q_i=q_i^b.\end{cases}$$

So in total we have $((G^*\pi_k\bar z+\alpha p_k)_i,(\bar q_k-p_k)_i)_{L^2(I_m)}\ge 0$.

Testing the semidiscrete optimality condition (18) with $\delta q=p_k$ gives

$$\langle G^*\bar z_k+\alpha\bar q_k,p_k-\bar q_k\rangle_I\ge 0. \tag{60}$$

By adding up the relations (59) and (60) we get

$$\langle G^*(\pi_k\bar z-\bar z_k),\bar q_k-p_k\rangle_I-\alpha|\bar q_k-p_k|_I^2\ge 0.$$

We split the first term and obtain the estimate

$$\begin{aligned}\alpha|\bar q_k-p_k|_I^2 &\le \langle G^*(\pi_k\bar z-z_k(\bar q)),\bar q_k-p_k\rangle_I\\ &\quad+\langle G^*(z_k(\bar q)-z_k(p_k)),\bar q_k-p_k\rangle_I+\langle G^*(z_k(p_k)-\bar z_k),\bar q_k-p_k\rangle_I.\end{aligned} \tag{61}$$

The first term on the right hand side is estimated as in the proof of Theorem 6 giving

$$\begin{aligned}\langle G^*(\pi_k\bar{z}-z_k(\bar{q})),\bar{q}_k-p_k\rangle_I &\leq C\,\|\pi_k\bar{z}-z_k(\bar{q})\|_I\,|\bar{q}_k-p_k|_I\\ &\leq Ck^3\left(\left\|\partial_t^2\Delta z\right\|_I+\left\|\partial_t^3 z\right\|_I+\left\|\partial_t^2 u\right\|_I\right)|\bar{q}_k-p_k|_I.\end{aligned} \tag{62}$$

To bound the second term we use Lemma 8 to get

$$\begin{aligned}\langle G^*(z_k(\bar{q})-z_k(p_k)),\bar{q}_k-p_k\rangle_I &\leq C\,\|z_k(\bar{q})-z_k(p_k)\|_I\,|\bar{q}_k-p_k|_I\\ &\leq C(\alpha)k^3\Big[\gamma(k)\left\|\partial_t^2\bar{z}\right\|_{L^\infty(I,H)}+\left\|\partial_t^3\bar{z}\right\|_I\\ &\qquad+\left\|\partial_t^2\bar{u}\right\|_I+\left\|\partial_t^2\bar{z}\right\|_I\Big]|\bar{q}_k-p_k|_I\end{aligned} \tag{63}$$

with $C(\alpha)\leq C\alpha^{-1}\left(1+\alpha^{-\frac{1}{2}}\right)$.

Finally the third term on the right hand side of (61) is estimated by applying the semidiscrete state equation followed by the semidiscrete adjoint equation, which results in

$$\langle G^*(z_k(p_k)-\bar{z}_k),\bar{q}_k-p_k\rangle_I=-B(u_k(p_k)-\bar{u}_k,z_k(p_k)-\bar{z}_k)=-\|u_k(p_k)-\bar{u}_k\|_I^2\leq 0. \tag{64}$$

Plugging the estimates (62), (63) and (64) into (61), dividing by $\alpha|\bar{q}_k-p_k|_I$ and using Young's inequality to obtain $1+\alpha^{-1}+\alpha^{-\frac{3}{2}}\leq C\left(1+\alpha^{-\frac{3}{2}}\right)$ gives

$$|p_k-\bar{q}_k|_I\leq C(\alpha)k^3\left\{\gamma(k)\left\|\partial_t^2\bar{z}\right\|_{L^\infty(I,H)}+\left\|\partial_t^2\bar{u}\right\|_I+\left\|\partial_t^2\bar{z}\right\|_I+\left\|\partial_t^3\bar{z}\right\|_I+\left\|\partial_t^2\Delta\bar{z}\right\|_I\right\} \tag{65}$$

with $C(\alpha)\leq C\left(\alpha^{-1}+\alpha^{-\frac{5}{2}}\right)$. Plugging this estimate into (58), collecting all the resulting terms and estimating the factors involving α where appropriate with Young's inequality yields the claim. □

5.2 Spatial discretization

The error that the spatial discretization causes on the post-processed solution can be assessed independently. We show the following estimate.

Theorem 9 *For the error $\tilde{\pi}_k\bar{z}_k-\tilde{\pi}_k\bar{z}_{kh}$ between the reconstruction of the semidiscrete optimal adjoint and the reconstruction of the adjoint $\bar{z}_{kh}$ belonging to the discrete optimal control problem* (29) *the estimate*

$$\|\tilde{\pi}_k(\bar{z}_k-\bar{z}_{kh})\|_I\leq C\left(1+\alpha^{-1}\right)h^2\left\{\left\|\nabla^2\bar{u}_k\right\|_I+\left\|\nabla^2\bar{z}_k\right\|_I\right\}$$

holds true.

Proof To show the claim, we split

$$\|\tilde{\pi}_k(\bar{z}_k-\bar{z}_{kh})\|_I\leq\|\tilde{\pi}_k\bar{z}_k-\tilde{\pi}_k z_{kh}(\bar{q}_k)\|_I+\|\tilde{\pi}_k(z_{kh}(\bar{q}_k)-\bar{z}_{kh})\|_I. \tag{66}$$

Corollary 2 estimates the first term on the right hand side. For the second term we use first the stability estimate in Lemma 5 and subsequently the stability estimates in Theorem 3 for the fully discrete state and adjoint equations which give

$$\|\tilde{\pi}_k\,(z_{kh}(\bar{q}_k)-\bar{z}_{kh})\|_I \leq C\,\|z_{kh}(\bar{q}_k)-\bar{z}_{kh}\|_I \leq C\,\|u_{kh}(\bar{q}_k)-\bar{u}_{kh}\|_I \leq C|\bar{q}_k-\bar{q}_{kh}|_I. \quad (67)$$

To estimate the term $\bar{q}_k - \bar{q}_{kh}$ we test the optimality conditions (18) and (31) with $\delta q = \bar{q}_{kh}$ and $\delta q = \bar{q}_k$ respectively and add up the results. This gives

$$\langle \alpha\bar{q}_k + G^*\bar{z}_k - (\alpha\bar{q}_{kh} + G^*\bar{z}_{kh})\,, \bar{q}_{kh} - \bar{q}_k\rangle_I \geq 0.$$

Hence

$$\begin{aligned}\alpha|\bar{q}_k-\bar{q}_{kh}|_I^2 &\leq (\bar{z}_k-\bar{z}_{kh}, G\,(\bar{q}_{kh}-\bar{q}_k))_I \\ &= (\bar{z}_k - z_{kh}(\bar{q}_k), G\,(\bar{q}_{kh}-\bar{q}_k))_I + (z_{kh}(\bar{q}_k)-\bar{z}_{kh}, G\,(\bar{q}_{kh}-\bar{q}_k))_I\end{aligned} \quad (68)$$

For the second term, using the discrete state equation (28) followed by the discrete adjoint (30) we obtain

$$\begin{aligned}(z_{kh}(\bar{q}_k)-\bar{z}_{kh}, G\,(\bar{q}_{kh}-\bar{q}_k))_I &= B\,(z_{kh}(\bar{q}_k)-\bar{z}_{kh}, \bar{u}_{kh}-u_{kh}(\bar{q}_k)) \\ &= -\,\|\bar{u}_{kh}-u_{kh}(\bar{q}_k)\|_I^2 \leq 0.\end{aligned}$$

Therefore, Equation (68) gives the estimate

$$|\bar{q}_k-\bar{q}_{kh}|_I \leq C\alpha^{-1}\,\|\bar{z}_k - z_{kh}(\bar{q}_k)\|_I\,. \quad (69)$$

Plugging this result into (67), applying Theorem 7 and collecting the resulting terms on the right hand side of Equation (66) shows the assertion. □

Corollary 4 *For the reconstructed fully discrete solution* $\tilde{q}_{kh} = P_{Q_{ad}}\left(-\alpha^{-1}\tilde{\pi}_k\bar{z}_{kh}\right)$, *we have the estimate*

$$\begin{aligned}|\tilde{q}_{kh}-\bar{q}|_I \leq C_1(\alpha)k^3 &\left\{\gamma(k)\left\|\partial_t^2\bar{z}\right\|_{L^\infty(I,H)} + \left\|\partial_t^2\bar{u}\right\|_I + \left\|\partial_t^2\bar{z}\right\|_I + \left\|\partial_t^3\bar{z}\right\|_I + \left\|\partial_t^2\Delta\bar{z}\right\|_I\right\} \\ &+ C_2(\alpha)h^2\left\{\left\|\nabla^2\bar{u}_k\right\|_I + \left\|\nabla^2\bar{z}_k\right\|_I\right\}\end{aligned}$$

with $C_1(\alpha) \leq \left(\alpha^{-1}+\alpha^{-\frac{7}{2}}\right)$ *and* $C_2(\alpha) \leq \left(\alpha^{-1}+\alpha^{-2}\right)$.

Proof The error is split into

$$|\tilde{q}_{kh}-\bar{q}|_I \leq |\tilde{q}_{kh}-\tilde{q}_k|_I + |\tilde{q}_k-\bar{q}|_I$$

and we estimate the first term using Lipschitz continuity of $P_{Q_{\text{ad}}}$ and Theorem 9 and the second term by Corollary 3. □

6 Numerical results

For the numerical tests, we consider the case $D=1$, that is, a control consisting of one time dependent parameter. As spatial domain we use the unit square $\Omega=(0,1)^2$. The data and exact solutions of the test problems are stated in terms of the eigenfunctions

$$w_k(x)=2\sin(k\pi x_1)\sin(k\pi x_2)$$

of the Laplacian on the unit square. We denote the corresponding eigenvalues by $\lambda_k=2k^2\pi^2$. The operator G is defined through $(Gq)(t,x):=q(t)w_k(x)$ and for the right hand side of the state equation we set $f=0$.

For solving the discrete optimal control problem, we use a variant of the semi-smooth Newton method presented in [26]. To avoid complicated data structures, the variational control is not stored explicitly. Instead we store $-\frac{1}{\alpha}G^*z_{kh}$ and cut off on the fly whenever the control is evaluated. All computations were done using the software package RoDoBo [22].

Our test example is constructed such that the third derivative of the adjoint state with respect to time has a jump at the point where the control constraint becomes inactive. We consider the time interval $I=(\frac{1}{2},1)$ and the control constraints $q_a=-\frac{\sqrt{3}}{2\alpha}$ and $q_b=\frac{\sqrt{3}}{2\alpha}$. The remaining data are chosen as

$$u_d(t,x)=(\pi\cos(\pi t)-\lambda_k\sin(\pi t))\left(1+\frac{1}{\alpha(\lambda_k^2+\pi^2)}\right)w_k(x),$$

$$u\left(\frac{1}{2},x\right)=\left(\frac{\pi e^{\frac{1}{6}\lambda_k}\left(\pi\sqrt{3}-\lambda_k\right)}{2\lambda_k\alpha(\lambda_k^2+\pi^2)}-\frac{\sqrt{3}}{2\lambda_k\alpha}\right)w_k(x),$$

$k=1$, and $\alpha=0.1$. For this choice of data, the optimal control is given by

$$\bar{q}(t)=\begin{cases}-\frac{\sqrt{3}}{2\alpha}, & \text{if } t\le\frac{2}{3},\\ -\frac{\sin(\pi(1-t))}{\alpha}, & \text{otherwise.}\end{cases}$$

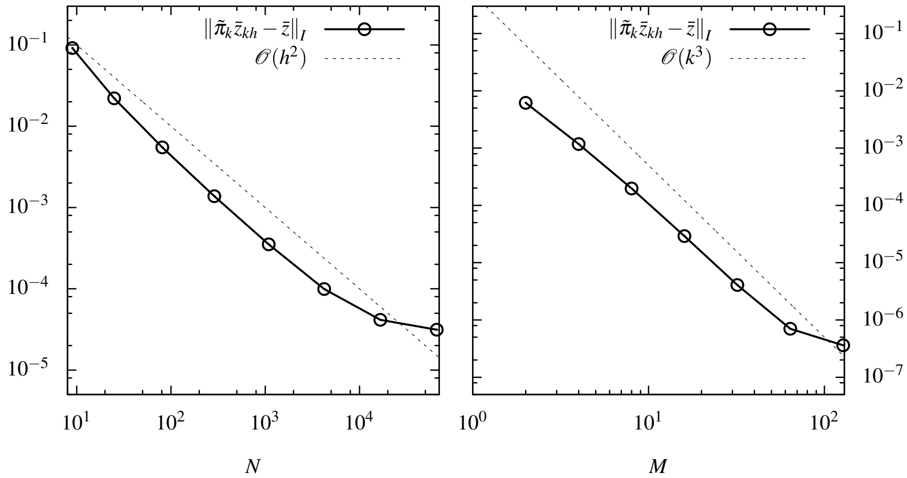
and for the optimal adjoint we obtain

$$\bar{z}(t,x)=\begin{cases}z_1(t)w_k(x), & \text{if } t\le\frac{2}{3},\\ \sin(\pi(1-t))w_k(x), & \text{otherwise,}\end{cases}$$

where

$$z_1(t)=\frac{\pi^2\sqrt{3}\cosh\left(\lambda_k\left(t-\frac{2}{3}\right)\right)+\pi\lambda_k\sinh\left(\lambda_k\left(t-\frac{2}{3}\right)\right)}{2\alpha\lambda_k^2\left(\lambda_k^2+\pi^2\right)} +\sin(\pi t)\left(1+\frac{1}{\alpha\left(\lambda_k^2+\pi^2\right)}\right)-\frac{\sqrt{3}}{2\alpha\lambda_k^2}.$$

We assess the $L^2(I,H)$ errors of the reconstructed adjoint for spatial and temporal discretization separately: to investigate the error of the spatial discretization, we fix the number of time steps at $M=16$, for the temporal error we consider a fixed uniform



(a) Refinement of the spatial grid for $M = 16$ time steps

(b) Refinement of the time steps for triangulation with $N = 1050625$ nodes

Fig. 2 Discretization error $\|\tilde{\pi}_k \bar{z}_{kh} - \bar{z}\|_I$ for spatial and temporal refinement

spatial triangulation with $N = 1050625$ nodes. In Figure 2(a) the error $\|\tilde{\pi}_k \bar{z}_{kh} - \bar{z}\|_I$ for a sequence of uniform refinements of the spatial grid is shown. Second order convergence with respect to the mesh width $h = \sqrt{\frac{1}{N}}$ is observed down to where the error contribution of the time discretization dominates.

Figure 2(b) shows the development of the error when refining the width $k = \frac{1}{M}$ of the time steps. The highest numerical order of convergence we observe is about 2.85, considerably less than predicted by Theorem 8. However, we note a slight increase of the observed order of convergence as the time steps decrease, up to the point where the spatial discretization error becomes dominant. Hence, it is reasonable to assume that the third order convergence will become apparent for smaller time steps and therefore can only be observed for an even finer spatial discretization. We substantiate this assumption by a computation using biquadratic finite elements in space, i. e., $s = 2$, which makes sense here since the solutions of our test problem are smooth with respect to the spatial variable. The result when refining the time step for a fixed spatial discretization with $N = 263169$ nodes is plotted in Figure 3. We observe a maximal estimated order of convergence of 2.97.

As evidence that the variational treatment of the control is in fact necessary to guarantee the almost third order error estimate for the post-processed solution, we consider an ODE version of the model problem (1). This avoids the spatial discretization error and therefore makes it easier to observe the influence of the control treatment on the time discretization. We compare our solution approach to a non-variational control treatment where the control space is discretized with piece-wise

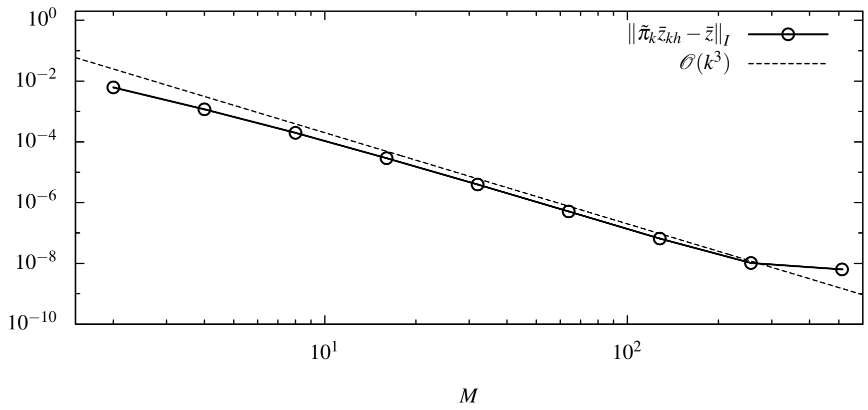


Fig. 3 Discretization error $\|\tilde{\pi}_k \bar{z}_{kh} - \bar{z}\|_I$ for biquadratic space discretization ($s = 2$) with $N = 263169$ nodes and uniform refinement of the time steps.

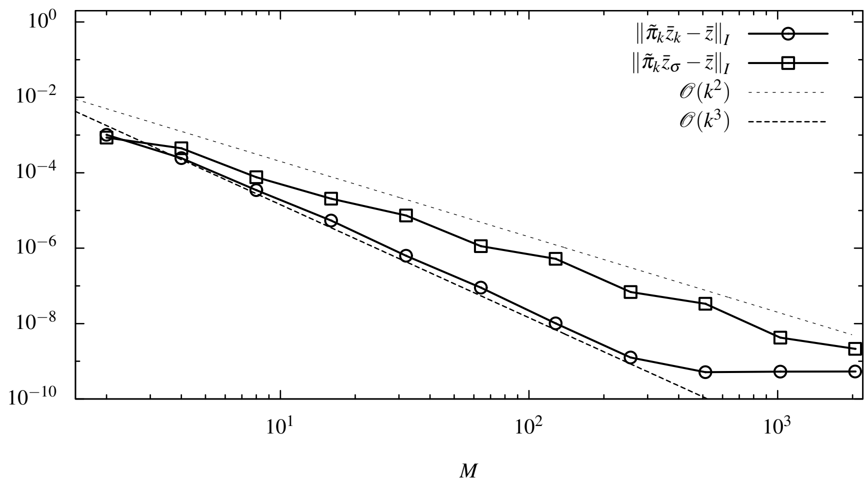


Fig. 4 Discretization errors $\|\tilde{\pi}_k \bar{z}_k - \bar{z}\|_I$ and $\|\tilde{\pi}_k \bar{z}_\sigma - \bar{z}\|_I$ for the ODE example

linear Ansatz functions in time. The scalar ODE problem reads: Minimize

$$J(q,u) = \frac{1}{2}\int_0^1 (u(t) - u_d(t))^2 \,\mathrm{d}t + \frac{\alpha}{2}\int_0^1 q(t)^2 \,\mathrm{d}t$$

subject to

$$\partial_t u + u = q, \quad u(0) = 0, \quad \text{and}$$
$$1 \le q(t) \le 2 \quad \text{a. e.}$$

with $q \in L^2((0,1))$ and $u \in H^1((0,1))$. We specify the data as

$$u_d(t) = \frac{3}{100}(2-t) + \begin{cases} 2 - 2e^{-t}, & t < \frac{1}{3}, \\ 6 - 3t - \left(3e^{\frac{1}{3}} + 2\right) e^{-t}, & \frac{1}{3} \le t \le \frac{2}{3}, \\ 1 + \left(3e^{\frac{2}{3}} - 3e^{\frac{1}{3}} - 2\right) e^{-t}, & t > \frac{2}{3}, \end{cases}$$

and $\alpha = 0.01$. Our regularity assumptions are obviously satisfied for the given data. The resulting optimal solutions are given by

$$\bar{q}(t) = \begin{cases} 2, & t < \frac{1}{3} \\ 3 - 3t, & \frac{1}{3} \le t \le \frac{2}{3} \\ 1, & t > \frac{2}{3} \end{cases},$$

$$\bar{u}(t) = u_d(t) + \frac{3}{100}(t-2), \quad \text{and}$$

$$\bar{z}(t) = \frac{3}{100}(t-1).$$

To solve the problem with variational control treatment, we use the same semi-smooth Newton method as above. For piece-wise linear control discretization, the discrete equations are solved by a primal-dual active set strategy, see, e. g., [5, 11]. We denote the discrete optimal control, state, and adjoint resulting from the discretized control by $\bar{q}_\sigma$, $\bar{u}_\sigma$, and $\bar{z}_\sigma$ respectively.

We examine the $L^2(I)$ error of the reconstructed adjoint for both treatments of the control on a sequence of uniformely refined temporal grids. In Figure 4 the development of the errors $\|\tilde{\pi}_k \bar{z}_k - \bar{z}\|_{L^2(I)}$ and $\|\tilde{\pi}_k \bar{z}_\sigma - \bar{z}\|_{L^2(I)}$ with respect to the number of time steps M is shown. It can be seen that while the reconstruction obtained from variational control treatment converges with third order, the error of $\tilde{\pi}_k \bar{z}_\sigma$ decreases only with second order.

Remark 6 In the example the control is piece-wise linear, so the main contribution to the observed error stems from the two discretization intervals containing kinks. For more general examples, it can be difficult to observe the different behaviour of the two discretizations because frequently the error caused by the kinks is dominated by the error contribution from approximating the control on the rest of the interval. Since the latter seems to be of third order for both, the error for the discretized control will be dominated by the second order component originating from the kinks only for very small time steps.

Acknowledgements The first author gratefully acknowledges financial support from the Munich Centre of Advanced Computing and the International Graduate School of Science and Engineering at the Technische Universität München. Furthermore, we would like to thank Konstantin Pieper for his help with the implementation of variational control discretization, Dominik Meidner for his helpful comments, and the anonymous referee for improving the estimate given in Lemma 6.

A Proof of Theorem 2

The proof is similar to Theorem 5 in [6, Chapter 7.1] where higher order stability estimates for a continuous parabolic problem are shown. Just as there, a Galerkin approximation with respect to the spatial variable is used. Let $\{v_n\}_{n\in\mathbb{N}}$ be an orthonormal basis of H consisting of eigenfunctions of $-\Delta$ defined on V. Then we define the spaces V_N and X^r_{kN} by

$$V_N := \operatorname{span}\{v_n \,|\, n \le N\},$$
$$X^r_{kN} := \left\{ v \in L^2(I,V_N) \,\middle|\, v|_{I_m} \in \mathscr{P}_r(I_m,V_N),\ m=1,\dots,M \right\}.$$

Replacing the test and trial spaces in Equation (20) by X^r_{kN} leads to a sequence of Galerkin approximations y_{kN} of the semidiscrete solution y_k. In a first step we have to show that for those approximations the stated stability estimates hold.

Lemma 9 *For the Galerkin approximations y_{kN} as defined above we have the stability estimate*

$$\left\|\Delta^2 y_{kN}\right\|_I + \left(\sum_{m=1}^{M} \|\partial_t \Delta y_{kN}\|^2_{I_m} \right)^{\frac{1}{2}} \le C \|\Delta w\|_I$$

with a constant C independent of N.

Proof To get the estimate for the first term we test with $\varphi = \Delta^3 y_{kN}$, which exists since y_{kN} is a linear combination of eigenvectors of Δ, resulting in

$$-\sum_{m=1}^{M} (\Delta^3 y_{kN}, \partial_t y_{kN})_{I_m} - (\Delta^3 y_{kN}, \Delta y_{kN})_I - \sum_{m=1}^{M-1} (\Delta^3 y^-_{kN,m}, [y_{kN}]_m) - (\Delta^3 y^-_{kN,M}, y^-_{kN,M}) = (\Delta^3 y_{kN}, w)_I.$$

We apply Green's formula to each term and get

$$\sum_{m=1}^{M} (\nabla\Delta y_{kN}, \partial_t \nabla\Delta y_{kN})_{I_m} - \left\|\Delta^2 y_{kN}\right\|^2_I + \sum_{m=1}^{M-1} (\nabla\Delta y^-_{kN,m}, [\nabla\Delta y_{kN}]_m) + \left\|\nabla\Delta y^-_{kN,M}\right\|^2 = (\Delta^2 y_{kN}, \Delta w)_I.$$

With the two identities

$$(\nabla\Delta y_{kN}, \partial_t \nabla\Delta y_{kN})_{I_m} = \frac{1}{2}\left\|\nabla\Delta y^-_{kN,m}\right\|^2 - \frac{1}{2}\left\|\nabla\Delta y^+_{kN,m-1}\right\|^2$$

and

$$(\nabla\Delta y^-_{kN,m}, [\nabla\Delta y_{kN}]_m) = \frac{1}{2}\left\|\nabla\Delta y^+_{kN,m}\right\|^2 - \frac{1}{2}\left\|\nabla\Delta y^-_{kN,m}\right\|^2 - \frac{1}{2}\|[\nabla\Delta y_{kN}]_m\|^2$$

we obtain

$$-\left\|\Delta^2 y_{kN}\right\|^2_I - \sum_{m=1}^{M-1} \frac{1}{2}\|[\nabla\Delta y_{kN}]_m\|^2 - \frac{1}{2}\left\|\nabla\Delta y^-_{kN,M}\right\|^2 = (\Delta^2 y_{kN}, \Delta w)_I,$$

which immediately gives the estimate for the first term

$$\left\|\Delta^2 y_{kN}\right\|_I \le \|\Delta w\|_I.$$

In order to obtain the second estimate, we test with the interval-wise defined function φ where $\varphi\big|_{I_m} = (t-t_m)\partial_t \Delta^2 y_{kN}$ for a fixed index m and $\varphi = 0$ otherwise. Using the dual formulation (17) of the bilinear form, we note that the jump terms vanish and we get

$$-((t-t_m)\partial_t \Delta^2 y_{kN}, \partial_t y_{kN})_{I_m} - ((t-t_m)\partial_t \Delta^2 y_{kN}, \Delta y_{kN})_{I_m} = ((t-t_m)\partial_t \Delta^2 y_{kN}, w)_{I_m}.$$

We apply Green's formula with respect to the spatial variable on each of the three terms and obtain after reordering

$$\int_{I_m} (t_m - t)\,\|\partial_t \Delta y_{kN}\|^2\,\mathrm{d}t = \int_{I_m} (t_m - t)(\partial_t \Delta y_{kN}, -\Delta w - \Delta^2 y_{kN})\,\mathrm{d}t$$
$$\leq \left(\int_{I_m} (t_m - t)\,\|\partial_t \Delta y_{kN}\|^2\,\mathrm{d}t\right)^{\frac{1}{2}} \left(\int_{I_m} (t_m - t)\,\left\|-\Delta w - \Delta^2 y_{kN}\right\|^2\,\mathrm{d}t\right)^{\frac{1}{2}}.$$

Together with the inverse estimate (4.5) from [17], which reads in our case

$$\|y_{kN}\|_{I_m}^2 \leq C k_m^{-1} \int_{I_m} (t_m - t)\,\|y_{kN}\|^2\,\mathrm{d}t$$

with C independent of N, we obtain the estimate

$$\|\partial_t \Delta y_{kN}\|_{I_m}^2 \leq C k_m^{-1} \int_{I_m} (t_m - t)\,\|\partial_t \Delta y_{kN}\|^2\,\mathrm{d}t$$
$$\leq C k_m^{-1} \int_{I_m} (t_m - t)\,\left\|-\Delta w - \Delta^2 y_{kN}\right\|^2\,\mathrm{d}t$$
$$\leq C\left\|-\Delta w - \Delta^2 y_{kN}\right\|_{I_m}^2 \leq C\left(\|\Delta w\|_{I_m}^2 + \left\|\Delta^2 y_{kN}\right\|_{I_m}^2\right).$$

Summing over all time intervals yields

$$\sum_{m=1}^{M} \|\partial_t \Delta y_{kN}\|_{I_m}^2 \leq C\left(\|\Delta w\|_I^2 + \left\|\Delta^2 y_{kN}\right\|_I^2\right)$$

which shows the second estimate. □

Proof of Theorem 2 From Lemma 9 we have

$$\left\|\Delta^2 y_{kN}\right\|_I + \left(\sum_{m=1}^{M} \|\partial_t \Delta y_{kN}\|_{I_m}^2\right)^{\frac{1}{2}} \leq C\,\|\Delta w\|_I$$

with C independent of N. Therefore the sequence $\{y_{kN}\}_{n\in\mathbb{N}}$ is bounded with respect to the norm $\|\cdot\|_Y$ given by

$$\|y_k\|_Y^2 = \|y_k\|_I^2 + \left\|\Delta^2 y_k\right\|_I^2 + \sum_{m=1}^{M} \|\partial_t \Delta y_k\|_{I_m}^2$$

and there exists a sub-sequence $\left(y_{kN_j}\right)_{j\in\mathbb{N}}$ that converges weakly with respect to the Y norm to a limit $\tilde{y}_k$ which satisfies the estimate

$$\left\|\Delta^2 \tilde{y}_k\right\|_I + \|\partial_t \Delta \tilde{y}_k\|_I \leq C\,\|\Delta w\|_I.$$

To complete the proof, we need to show that $\tilde{y}_k$ is in fact the solution y_k of the semidiscrete problem (20). Therefore we note that the stability estimate in Corollary 1 also works for the Galerkin approximations y_{kN} with the constant C independent of N. We fix $\bar{N}$, then for any $\varphi \in X_{k\bar{N}}^r$ and for any $N_j \geq \bar{N}$ the identity

$$-\sum_{m=1}^{M} (\varphi, \partial_t y_{kN_j})_{I_m} - (\varphi, \Delta y_{kN_j})_I - \sum_{m=1}^{M} (\varphi^-, [y_{kN_j}]_m) = (\varphi, w)_I \tag{70}$$

holds true. Since $\sum_{m=1}^{M}\left\|\partial_t y_{kN_j}\right\|_I^2$, $\left\|\Delta y_{kN_j}\right\|_I$ and $\sum_{m=1}^{M}\left\|[y_{kN_j}]_m\right\|^2$ are bounded by the stability estimate we can extract a subsequence such that (70) holds for the weak limit which has to be $\tilde{y}_k$ again. Passing to the limit $\bar{N} \to \infty$ shows that in fact $\tilde{y}_k = y_k$. □

References

1. Abramowitz, M., Stegun, I.A. (eds.): Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables, *Applied Mathematics Series*, vol. 55. National Bureau of Standards (1972). 10^{th} printing
2. Adjerid, S., Devine, K.D., Flaherty, J.E., Krivodonova, L.: A posteriori error estimation for discontinuous Galerkin solutions of hyperbolic problems. Comput. Methods Appl. Mech. Eng. **191**(11-12), 1097 – 1112 (2002)
3. Akrivis, G., Makridakis, C., Nochetto, R.: Galerkin and Runge-Kutta methods: unified formulation, a posteriori error estimates and nodal superconvergence. Numer. Math. **118**, 429–456 (2011)
4. Becker, R., Vexler, B.: Optimal control of the convection-diffusion equation using stabilized finite element methods. Numer. Math. **106**(3), 349–367 (2007)
5. Bergounioux, M., Ito, K., Kunisch, K.: Primal-dual strategy for constrained optimal control problems. SIAM J. Control Optimization **37**(4), 1176–1194 (1999)
6. Evans, L.C.: Partial differential equations, *Graduate Series in Mathematics*, vol. 19, second edn. AMS, Providence (2010)
7. Fidkowski, K.J.: Output error estimation strategies for discontinuous Galerkin discretizations of unsteady convection-dominated flows. Int. J. Numer. Methods Eng. **88**(12), 1297–1322 (2011)
8. Gascoigne: The finite element toolkit. URL http://www.gascoigne.uni-hd.de
9. Grisvard, P.: Elliptic problems in nonsmooth domains., *Monogr. Stud. Math.*, vol. 24. Pitman Publishing Inc (1985)
10. Hinze, M.: A variational discretization concept in control constrained optimization: The linear-quadratic case. Comput. Optim. Appl. **30**(1), 45–61 (2005)
11. Kunisch, K., Rösch, A.: Primal-dual active set strategy for a general class of constrained optimal control problems. SIAM J. Optim. **13**(2), 321–334 (2002)
12. Lasiecka, I., Malanowski, K.: On discrete-time Ritz-Galerkin approximation of control constrained optimal control problems for parabolic systems. Control Cybernet. **7**(1), 21–36 (1978)
13. Makridakis, C., Nochetto, R.H.: A posteriori error analysis for higher order dissipative methods for evolution problems. Numer. Math. **104**, 489–514 (2006)
14. Malanowski, K.: Convergence of approximations vs. regularity of solutions for convex, control-constrained optimal-control problems. Appl. Math. Optim. **8**, 69–95 (1982)
15. Matthies, G., Schieweck, F.: Higher order variational time discretizations for nonlinear systems of ordinary differential equations. Preprint (2011)
16. Maz'ya, V., Rossmann, J.: Elliptic Equations in Polyhedral Domains., *Math. Surv. Monogr.*, vol. 162. AMS (2010)
17. Meidner, D., Vexler, B.: 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**(3), 1150–1177 (2008)
18. Meidner, D., Vexler, B.: 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**(3), 1301–1329 (2008)
19. Meidner, D., Vexler, B.: A Priori Error Analysis of the Petrov–Galerkin Crank–Nicolson Scheme for Parabolic Optimal Control Problems. SIAM J. Control Optim. **49**(5), 2183–2211 (2011)
20. Meyer, C., Rösch, A.: Superconvergence properties of optimal control problems. SIAM J. Control Optim. **43**(3), 970–985 (2004)
21. Michels, H.H.: Abscissas and weight coefficients for Lobatto quadrature. Math. Comp. **17**, 237–244 (1963)
22. RoDoBo: A C++ library for optimization with stationary and nonstationary PDEs with interface to GASCOIGNE [8]. URL http://www.rodobo.uni-hd.de
23. Rösch, A.: Error estimates for parabolic optimal control problems with control constraints. Z. Anal. Anwend. **23**(2), 353–376 (2004)
24. Rösch, A., Simon, R.: Superconvergence properties for optimal control problems discretized by piecewise linear and discontinuous functions. Numer. Funct. Anal. Optim. **28**(3-4), 425–443 (2007)
25. Rösch, A., Vexler, B.: Optimal control of the Stokes equations: a priori error analysis for finite element discretization with postprocessing. SIAM J. Numer. Anal. **44**(5), 1903–1920 (2006)
26. Schiela, A.: A simplified approach to semismooth Newton methods in function space. SIAM J. Optim. **19**(3), 1417–1432 (2008)

27. Schötzau, D., Wihler, T.: A posteriori error estimation for hp-version time-stepping methods for parabolic partial differential equations. Numer. Math. **115**, 475–509 (2010)
28. Thomée, V.: Galerkin Finite Element Methods for Parabolic Problems, second edn. Springer Ser. Comput. Math. Springer-Verlag, Berlin (2006)
29. Wloka, J.: Partial differential equations. Cambridge University Press, Cambridge (1987)