Crank-Nicolson time stepping and variational discretization of control-constrained parabolic optimal control problems

Nikolaus von Daniels*
Michael Hinze*
Morten Vierling*

3-APR-2014

Abstract: We consider a control constrained parabolic optimal control problem and use variational discretization for its time semi-discretization, where the state equation is treated with a Petrov-Galerkin scheme using a piecewise constant Ansatz for the state and piecewise linear, continuous test functions. This results in variants of the Crank-Nicolson scheme for the state and the adjoint state. Exploiting a superconvergence result we prove second order convergence in time of the error in the controls. Moreover, the piecewise linear and continuous parabolic projection of the discrete state on the dual time grid provides a second order convergent approximation of the optimal state without further numerical effort. Numerical experiments confirm our analytical findings.

Mathematics Subject Classification (2010):

Keywords: Optimal control; Heat equation; Control constraints; Crank-Nicolson time stepping; Error estimates; Variational control discretization.

1 Introduction

The purpose of this paper is to prove optimal a priori error bounds for the variational time semi-discretization of a generic parabolic optimal control problem, where the state in time is approximated with a Petrov-Galerkin scheme. The key idea consists in choosing piecewise linear, continuous test functions and a discontinuous, piecewise constant Ansatz for the approximation of the state equation. With this Petrov Galerkin Ansatz variational discretization delivers a cG(1) time approximation of the optimal time semi-discrete adjoint state. The resulting time integration schemes for the state and the adjoint state are variants of the Crank-Nicolson scheme. Combining this setting with the supercloseness result of Corollary 4.2 for interval means we are able to prove second order in time convergence of the time discrete optimal control. Moreover, the piecewise linear and continuous parabolic projection of

*Schwerpunkt Optimierung und Approximation, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany.
the discrete state based on the the values of the discrete state on the dual time grid provides a second order convergent approximation of the optimal state without further numerical effort, see Lemma 4.5.

Our work is motivated by the work [MV11] of Meidner and Vexler, where under mild assumptions on the active set the same convergence order in time is shown for the post-processed piecewise linear, continuous parabolic projection of the piecewise constant in time optimal control obtained by a Petrov Galerkin scheme with variational discretization of the parabolic optimal control problem. In comparison to our work, they switched Ansatz and test space in their schemes.

Our work also shows that variational discretization of [Hin05] through the choice of Ansatz and test space in a Petrov Galerkin approximation of parabolic optimal control problems offers the possibility to control the discrete structure of variational optimal controls. Of course this fact also holds for other classes of PDE constrained optimal control problems.

In our note we only consider semi-discretization in time, since we are interested in the approximation of optimal controls, which in a realistic time-dependent scenario only depend on time, see the definition of the control operator $B$ below.

Our main result is proved in Theorem 5.2, where we show

$$\|\bar{u} - \bar{u}_k\| \leq Ck^2,$$

with $\bar{u}, \bar{u}_k$ denoting the optimal controls obtained from problems $(P)$ and $(P_k)$, respectively, where $k$ denotes the grid size of the time grid. This result could be compared to [MV11, Theorem 6.2], where under mild assumptions on the structure of the active set a similar bound is obtained for the post-processed parabolic projection of a piecewise constant optimal control. Our approach avoids such an assumption in the numerical analysis.

With $I := (0, T) \subset \mathbb{R}$ we consider the linear-quadratic optimal control problem

$$\min_{y \in Y, u \in U_{ad}} J(y, u) = \frac{1}{2}\|y - y_d\|^2_{L^2(I, L^2(\Omega))} + \frac{\alpha}{2}\|u\|^2_U,$$

s.t. $y = S(Bu, y_0),$

where with

$$W(I) := \{v \in L^2(I, H_0^1(\Omega)), v_t \in L^2(I, H^{-1}(\Omega))\} \hookrightarrow C([0, T], L^2(\Omega))$$

the operator $S : L^2(I, H^{-1}(\Omega)) \times L^2(\Omega) \to W(I)$, $(f, \kappa) \mapsto y := S(f, \kappa)$, denotes the weak solution operator associated with the parabolic problem

$$\begin{align*}
p_t y - \Delta y &= f \quad \text{in } I \times \Omega, \\
u &= 0 \quad \text{in } I \times \partial \Omega, \\
u(0) &= \kappa \quad \text{in } \Omega,
\end{align*}$$

i.e. for $(f, \kappa) \in L^2(I; H^{-1}(\Omega)) \times L^2(\Omega)$ the abstract function $y \in W(I)$ satisfies $y(0) = \kappa$ and

$$\int_0^T \langle y_t, v \rangle_{H^{-1}H_0^1} + a(y, v) \, dt = \int_0^T \langle f, v \rangle_{H^{-1}H_0^1} \, dt \quad \forall \, v \in L^2(I; H_0^1(\Omega)).$$

2
Here \( \Omega \subset \mathbb{R}^n, n = 2, 3 \), is a convex polygonal domain with boundary \( \partial \Omega \),

\[
a(y, v) := \int_\Omega \nabla y \nabla v \, dx,
\]

the control space is \( U = L^2(I, \mathbb{R}^D) \), \( D \in \mathbb{N} \), and the admissible set

\[
U_{ad} = \{ u \in U \mid a_i \leq u_i \leq b_i \text{ a.e., } i = 1, \ldots, D \}
\]
is closed and convex in \( U \), where \( a_i, b_i \in \mathbb{R}, a_i < b_i \) (\( i = 1, \ldots, D \)). The control operator

\[
B : U \to L^2(I, H^{-1}(\Omega)), \quad u \mapsto \sum_{i=1}^D u_i g_i,
\]
is linear and continuous with fixed functionals \( g_i \in H^{-1}(\Omega) \), whose regularity is specified in Assumption 1.1 below. The state space is \( Y = W(I) \). It is well known that the operator \( S \) is well defined, i.e. for every \((f, \kappa) \in L^2(I, H^{-1}(\Omega)) \times L^2(\Omega)\) a unique state \( y \in W(I) \) satisfying (1.2) exists. Furthermore, it fulfills

\[
\|y\|_{W(I)} \leq C \left\{ \|f\|_{L^2(I, H^{-1}(\Omega))} + \|\kappa\|_{L^2(\Omega)} \right\}.
\]

Now let \( y \in Y \) denote the unique solution of (1.2). Then it follows from integration by parts for abstract functions, that with the bilinear form \( A \) defined by

\[
A(y, v) := \int_0^T \langle v_t, y \rangle_{H^{-1}H^1_0} + a(y, v) dt + (y(T), v(T))_{L^2},
\]

the state \( y \) also satisfies

\[
A(y, v) = \int_0^T \langle f, v \rangle_{H^{-1}H^1_0} dt + (\kappa, v(0))_{L^2} \quad \forall \ v \in W(I).
\]

Furthermore, \( y \) is the only function in \( Y \) which satisfies (1.5). In the next section we use the bilinear form \( A \) to define our numerical approximation scheme for the state equation.

With \( O(k^2) \) error-bounds for the control in mind we follow [MV11] and make the following assumptions on the data.

**Assumption 1.1.** Let \( y_d \in H^1(I, L^2(\Omega)) \), and \( y_d(T) \in H^1_0(\Omega) \). Let further \( g_i \in H^1_0(\Omega), \) \( i = 1, \ldots, D \), and finally \( y_0 \in H^1_0(\Omega) \) with \( \Delta y_0 \in H^1_0(\Omega) \).

A lot of literature is available on optimal control problems with parabolic state equations. We refer to [HPUU09] for a comprehensive discussion, and also to [MV08a, MV08b, MV11, SV13] for the most recent developments related to optimal control with Galerkin methods in time. The paper is organized as follows. In section 2 we briefly summarize the solution theory of the optimal control problem. In section 3 we analyse the regularity of the state and the adjoint state, which plays an important role in the time discretization. In section 4 the time discretization of state and adjoint state is discussed in detail. In section 5 we introduce variational discretization of the optimal control problem \((P)\) and prove second order convergence of the variational discrete controls in time. In section 6 we present numerical results which confirm our analytical findings.
2 The continuous problem \((P)\)

It is well known that problem \((P)\) admits a unique solution \((\bar{y}, \bar{u}) \in Y \times U\), where \(\bar{y} = S(B\bar{u}, y_0)\). Moreover, using the orthogonal projection \(P_{U_{ad}} : L^2(I, \mathbb{R}^D) \to U_{ad}\), the optimal control is characterized by the first-order necessary and sufficient condition

\[
\bar{u} = P_{U_{ad}} \left( -\frac{1}{\alpha} B' \bar{p} \right),
\]

(2.1)

where \((\bar{p}, \bar{q}) \in L^2(I, H_0^1(\Omega)) \times L^2(\Omega)\) (here we use reflexivity of the involved spaces) denotes the adjoint variable which is the unique solution to

\[
\int_0^T \langle \bar{y}_t, \bar{p} \rangle_{H^{-1}H_0^1} + a(\bar{y}, \bar{p}) \, dt + (\bar{y}(0), \bar{q})_{L^2} = \int_0^T (\bar{y} - y_d)\bar{y} \, dx \, dt \quad \forall \bar{y} \in W(I).
\]

(2.2)

Here, \(B' : L^2(I, H_0^1(\Omega)) \to L^2(I, \mathbb{R}^D)\) denotes the adjoint operator of \(B\), which is characterized by

\[
B'q = \left( \langle g_1, q \rangle_{H^{-1}H_0^1}, \ldots, \langle g_D, q \rangle_{H^{-1}H_0^1} \right)^T.
\]

(2.3)

Furthermore we note that for \(v \in L^2(I, \mathbb{R}^D)\) there holds

\[
P_{U_{ad}}(v)(t) = \left( P_{[a_i, b_i]}(v_i(t)) \right)_{i=1}^D,
\]

where for \(a, b, z \in \mathbb{R}\) with \(a \leq b\) we set \(P_{[a, b]}(z) := \max\{a, \min\{z, b\}\}\).

Since \(\bar{y} - y_d \in L^2(I, L^2(\Omega))\) in (2.2), we have \(\bar{p} \in W(I)\), so that by integration by parts for abstract functions we conclude from (2.2) (compare (1.5))

\[
\int_0^T -\langle \bar{y}_t, \bar{p} \rangle_{H^{-1}H_0^1} + a(\bar{y}, \bar{p}) \, dt + (\bar{y}(0), \bar{q})_{L^2} + (\bar{y}(T), \bar{p}(T))_{L^2} - (\bar{y}(0), \bar{p}(0))_{L^2} = \int_0^T (\bar{y} - y_d)\bar{y} \, dx \, dt \quad \forall \bar{y} \in W(I),
\]

(2.4)

so that the function \(\bar{p}\) can be identified with the unique weak solution to the adjoint equation

\[
-\partial_t \bar{p} - \Delta \bar{p} = h \quad \text{in } I \times \Omega,
\]

\[
\bar{p} = 0 \quad \text{on } I \times \partial\Omega,
\]

\[
\bar{p}(T) = 0 \quad \text{on } \Omega,
\]

(2.5)

with \(h := \bar{y} - y_d\). Moreover, \(\bar{q} = \bar{p}(0)\).

3 Regularity results

In this section we summarize some existence and regularity results concerning equation (1.1) and (2.5), which can also be found in e.g. [MV11]. We abbreviate

\[
\|\cdot\|_I := \|\cdot\|_{L^2(I, L^2(\Omega))}, \quad |\cdot|_I := \|\cdot\|_{L^2(I, \mathbb{R}^D)}.
\]

For the unique weak solutions \(y\) to (1.1) and \(p\) to (2.5) we have from [Eva98] Theorems 7.1.5 and 5.9.4] the regularity results.
Lemma 3.1. For \( f, h \in L^2(I, L^2(\Omega)) \) and \( \kappa \in H^1_0(\Omega) \) the solutions \( y \) of (1.1) and \( p \) of (2.5) satisfy
\[
y, p \in L^2(I, H^2(\Omega) \cap H^1_0(\Omega)) \cap H^1(I, L^2(\Omega)) \rightarrow C([0, T], H^1_0(\Omega)).
\]
Furthermore, with some constant \( C > 0 \) there holds
\[
\|y\|_I + \|\partial_t y\|_I + \|\Delta y\|_I + \max_{t \in I} \|y(t)\|_{H^1(\Omega)} \leq C \left\{ \|f\|_I + \|\kappa\|_{H^1(\Omega)} \right\},
\]
and
\[
\|\partial_t p\|_I + \|\Delta p\|_I + \max_{t \in I} \|p(t)\|_{H^1(\Omega)} \leq C \|h\|_I.
\]
However, in order to achieve \( O(k^2) \)-convergence we need more regularity, i.e., at least second weak time derivatives. From [MV11, Proposition 2.1] we have

Lemma 3.2. Let \( f, h \in H^1(I, L^2(\Omega)) \), \( f(0), h(T) \in H^1_0(\Omega) \), and \( \kappa \in H^1_0(\Omega) \) with \( \Delta \kappa \in H^1_0(\Omega) \). Then the solutions \( y \) of (1.1) and \( p \) of (2.5) satisfy
\[
y, p \in H^1(I, H^2(\Omega) \cap H^1_0(\Omega)) \cap H^2(I, L^2(\Omega)).
\]
With some constant \( C > 0 \) we have the a priori estimates
\[
\|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I \leq C \left\{ \|f\|_{H^1(I, L^2(\Omega))} + \|f(0)\|_{H^1(\Omega)} + \|\kappa\|_{H^1(\Omega)} + \|\Delta \kappa\|_{H^1(\Omega)} \right\},
\]
and
\[
\|\partial_t^2 p\|_I + \|\partial_t \Delta p\|_I \leq C \left\{ \|h\|_{H^1(I, L^2(\Omega))} + \|h(T)\|_{H^1(\Omega)} \right\}.
\]
From Lemma 3.1 we conclude that the optimal state \( \bar{y} \) lives in \( H^1(I, L^2(\Omega)) \), and \( \bar{y}(T) \in H^1_0(\Omega) \). Thus, by Lemma 3.2 and Assumption 1.1 the optimal adjoint state \( \bar{p} \) is an element of \( H^2(I, L^2(\Omega)) \). It then follows from (2.3) that \( B' \bar{p} \in H^2(I, \mathbb{R}^D) \). Furthermore, for \( v \in W^{1,r}(I, \mathbb{R}^D), 1 \leq r \leq \infty \), one has
\[
\|\partial_t P_{a,b} (v)\|_{L^r(I, \mathbb{R}^D)} \leq \|\partial_t v\|_{L^r(I, \mathbb{R}^D)},
\]
so that (2.1) implies \( \bar{u} \in W^{1,\infty}(I, \mathbb{R}^D) \).

Hence, using our Assumption 1.1 Lemma 3.2 is applicable to the solution of the state equation and one obtains the following result.

Lemma 3.3 (Proposition 2.3 in [MV11]). Let Assumption 1.1 hold. For the unique solution \((\bar{y}, \bar{u})\) of (P) and the corresponding adjoint state \( \bar{p} \) there holds
\[
\bar{y}, \bar{p} \in H^1(I, H^2(\Omega) \cap H^1_0(\Omega)) \cap H^2(I, L^2(\Omega)), \quad \text{and} \quad \bar{u} \in W^{1,\infty}(I, \mathbb{R}^D).
\]

4 Time discretization

Let \( [0, T] = \bigcup_{m=1}^M I_m \), where the intervals \( I_m = [t_{m-1}, t_m] \) are defined through the partition \( 0 = t_0 < t_1 < \cdots < t_M = T \). Furthermore, let \( t^*_m = \frac{t_{m-1} + t_m}{2} \) denote the interval midpoints.

By \( 0 = t_0^* < t_1^* < \cdots < t_M^* < t_{M+1} := T \) we get the so-called dual partition of \([0, T]\), namely \( [0, T) = \bigcup_{m=1}^{M+1} I_m^* \), with \( I_m^* = [t_{m-1}^*, t_m^*] \). The grid width of the first (primal) partition is defined by the mesh-parameters \( k_m = t_m - t_{m-1} \) and
\[
k = \max_{1 \leq m \leq M} k_m.
\]
On these partitions we define the Ansatz and test spaces of our Petrov Galerkin scheme for the numerical approximation of the optimal control problem \( P \) w.r.t. time. We set

\[
P_k := \left\{ v \in C([0, T], H^1_0(\Omega)) \left| v|_{I_m} \in \mathcal{P}_1(I_m, H^1_0(\Omega)) \right\} \rightarrow W(I),
\]

\[
P_k^* := \left\{ v \in C([0, T], H^1_0(\Omega)) \left| v|_{I_m} \in \mathcal{P}_1(I_m, H^1_0(\Omega)) \right\} \rightarrow W(I)
\]

and

\[
Y_k := \left\{ v : [0, T] \rightarrow H^1_0(\Omega) \left| v|_{I_m} \in \mathcal{P}_0(I_m, H^1_0(\Omega)) \right\}. \]

We note that \( \dim(P_k) = \dim(Y_k) = M + 1 \), and that every \( v \in Y_k \) is also an element of \( L^2(I, H^1_0(\Omega)) \).

In what follows we frequently use the interpolation operators

1. \( \mathcal{P}_{Y_k} : L^2(I, H^1_0(\Omega)) \rightarrow Y_k \)

\[
\mathcal{P}_{Y_k} v|_{I_m} := \frac{1}{k_m} \int_{t_{m-1}}^{t_m} v(t) \, dt \text{ for } m = 1, \ldots, M, \text{ and } \mathcal{P}_{Y_k} v(T) := 0
\]

2. \( \Pi_{Y_k} : C([0, T], H^1_0(\Omega)) \rightarrow Y_k \)

\[
\Pi_{Y_k} v|_{I_m} := v(t^*_m) \text{ for } m = 1, \ldots, M, \text{ and } \Pi_{Y_k} v(T) := v(T).
\]

3. \( \pi_{P_k^*} : C([0, T], H^1_0(\Omega)) \cup Y_k \rightarrow P_k^* \)

\[
\pi_{P_k^*} v|_{I^*_1 \cup I^*_2} := v(t^*_1) + \frac{t^*_2 - t^*_1}{t^*_2 - t^*_1} (v(t^*_2) - v(t^*_1)),
\]

\[
\pi_{P_k^*} v|_{I_m} := v(t^*_{m-1}) + \frac{t - t^*_{m-1}}{t^*_m - t^*_{m-1}} (v(t^*_m) - v(t^*_{m-1})), \text{ for } m = 3, \ldots, M - 1,
\]

\[
\pi_{P_k^*} v|_{I^*_M \cup I^*_M+1} := v(t^*_M) + \frac{t - t^*_{M-1}}{t^*_M - t^*_{M-1}} (v(t^*_M) - v(t^*_{M-1})).
\]

To apply variational discretization to \( P \) we next introduce the Petrov-Galerkin scheme for the approximation of the states. For this purpose we extend the bilinear form \( A \) of \( 4.4 \) from \( W(I) \) to \( W(I) \cup Y_k \), i.e. we consider \( A \) as a mapping \( A : W(I) \cup Y_k \times W(I) \rightarrow \mathbb{R} \). Then, according to \( 4.5 \) we for \( (f, \kappa) \in L^2(I; H^{-1}(\Omega)) \times L^2(\Omega) \) consider the time-semidiscrte problem: Find \( y_k \in Y_k \), such that

\[
A(y_k, v_k) = \int_0^T \langle f, v_k \rangle_{H^{-1}_0} \, dt + \langle \kappa, v_k(0) \rangle_{L^2} \quad \forall \, v_k \in P_k.
\]

(4.1)

Then \( y_k \in Y_k \) is uniquely determined. This follows from the fact that with

\[
y_k = \alpha_M \chi_{[T]} + \sum_{i=0}^{M-1} \alpha_i \chi_{I_{i+1}}, \quad \alpha_i \in H^1_0(\Omega) \text{ for } i = 0, \ldots, M,
\]
the coefficients $\alpha_i$ for $i = 1, \ldots, M - 1$ are determined by a
half Rannacher smoothing step \cite{Ran84} for $\alpha_0$, and $\alpha_M$ is uniquely determined by $\alpha_{M-1}$.
For Petrov Galerkin approximations $y_k \in Y_k$ of states $y \in W(I)$ we can only expect $O(h)$
convergence, since $y_k$ is piecewise constant in time, compare \cite[Lemma 5.2]{MV11}. In order
to obtain $O(h^2)$ control approximations in our convergence analysis for problem (P) we rely
on the following super-convergence results for the projections $\Pi_k$ and $P_k$, see e.g. \cite[Lemma 5.3]{MV11}.

\textbf{Lemma 4.1.} Let $(f, \kappa)$ satisfy the regularity requirements of Lemma 3.2 and let $y, y_k$ solve
(1.2) and (4.1) with data $(f, \kappa)$, thus $y \in H^1(I, H^2(\Omega) \cap H^1_0(\Omega)) \cap H^2(I, L^2(\Omega))$. Then for some $C > 0$ there holds

$$\|y_k - \Pi_k y\|_I \leq C k^2 (\|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I).$$

Note that the proof of \cite[Lemma 5.3]{MV11} is applicable in our situation since the initial value
$\kappa$ is the same for both, the continuous problem (1.2) and (4.1).

\textbf{Corollary 4.2.} Let the assumptions of Lemma 4.1 hold. Then for some $C > 0$ there holds

$$\|y_k - P_k y\|_I \leq C k^2 (\|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I).$$

\textbf{Proof.} With the result of Lemma 4.1 at hand it suffices to show that

$$\|\Pi_k y - P_k y\|_I \leq k^2 \|\partial_t^2 y\|_I \quad (4.2)$$

holds. We prove this estimate for smooth functions $w \in C^2(I, L^2(\Omega)) \cap H^2(I, L^2(\Omega))$. The result
then follows by a density argument.

Suppose $w \in C^2(I, L^2(\Omega)) \cap H^2(I, L^2(\Omega))$. We use the Taylor expansion of $w$ at $t_m^*$ and obtain

$$\left\| \int_{t_m}^{t_{m+1}} w(t) - w(t_m^*) dt \right\|_{L^2(\Omega)}^2 = \left\| \int_{t_m}^{t_{m+1}} (t - t_m^*) \partial_t w(t_m^*) + \int_{t_m}^{t_{m+1}} (t - s) \partial_t^2 w(s) ds dt \right\|_{L^2(\Omega)}^2 \leq k_m \int_{t_m}^{t_{m+1}} \left\| (t - s) \partial_t^2 w(s) ds \right\|_{L^2(\Omega)}^2 dt \leq k_m^2 \int_{t_m}^{t_{m+1}} \left\| \partial_t^2 w(s) \right\|_{L^2(\Omega)}^2 ds \leq k_m^5 \int_{t_m}^{t_{m+1}} \left\| \partial_t^2 w(s) \right\|_{L^2(\Omega)}^2 ds, \quad (4.3)$$

where we have used the Cauchy-Schwarz inequality twice. This proves

$$\left( \sum_{m=1}^{M} k_m \left\| \frac{1}{k_m} \int_{t_{m-1}}^{t_m} w(t) - w(t_m^*) dt \right\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}} \leq k^2 \|\partial_t^2 w\|_I,$$

which is (4.2).

For the next Lemma, see \cite[Lemma 5.6]{MV11}, we need the following condition on the time
grid:

7
Assumption 4.3. There exists a constant $\kappa \geq 1$ independent of $k$ such that

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

holds for all $m = 1, 2, \ldots, M - 1$.

Lemma 4.4. Let the Assumption 4.3 be fulfilled. The interpolation operator $\pi_{P_k}$ has the following properties, where $C > 0$ is a constant independent of $k$.

1. $\|w - \pi_{P_k}w\|_I \leq C k^2 \|\partial_t^2 w\|_I \forall w \in H^2(I, L^2(\Omega))$, 
2. $\|\pi_{P_k}w\|_I \leq C \|w_k\|_I \forall w_k \in Y_k$.

Since the state is discretized piecewise constant, we can only expect first order convergence in time for its discretization error. The following Lemma shows that a projected version of the discretized state converges second order in time to the continuous state. The benefit of this result will be discussed in the numerics section.

Lemma 4.5. Let $y$ and $y_k$ be given as in Lemma 4.1. Then there holds

$$\|\pi_{P_k}y_k - y\|_I \leq C k^2 (\|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I).$$

Proof. Making use of the splitting

$$\|\pi_{P_k}y_k - y\|_I = \|\pi_{P_k}(y_k - \Pi_{Y_k}y)\|_I + \|\pi_{P_k}\Pi_{Y_k}y - y\|_I = \|\pi_{P_k}(y_k - \Pi_{Y_k}y)\|_I + \|\pi_{P_k}y - y\|_I,$$

the claim is an immediate consequence of Lemma 4.4 and Lemma 4.1.

In the numerical treatment of problem (P) we also need error estimates for discrete adjoint functions $p_k \in P_k \hookrightarrow W(I)$. For $h \in L^2(I, H^{-1}(\Omega))$ we consider the problem: Find $p_k \in P_k$ such that

$$A(\tilde{y}, p_k) = \int_0^T \langle h, \tilde{y}\rangle_{H^{-1}H^1_0} dt \quad \forall \tilde{y} \in Y_k. \quad (4.4)$$

This problem admits a unique solution $p_k \in P_k$. This follows from the fact that the solution operator of (4.4) is the adjoint of the solution operator associated to problem (4.1). If we write

$$p_k = \sum_{i=0}^M \beta_i b_i(t)$$

with coefficients $\beta_i \in H^1_0(\Omega)$ and $b_i \in C([0, T])$, $b_i(t_j) = \delta_{ij}$, for $i, j = 0, \ldots, M$, the coefficients $\beta_i$ are determined by a backward in time Crank-Nicolson scheme, starting with $\beta_M \equiv 0$. We have from [MV11, Lemma 4.7]

Lemma 4.6. Let $p_k \in P_k$ solve (4.4) with $h \in L^2(I, L^2(\Omega))$. Then there exists a constant $C > 0$ independent of $h$ such that

$$\|p_k\|_I + \|p_k(0)\|_{L^2(\Omega)} \leq C \|h\|_I.$$
Moreover, from [MV11, Lemma 6.3] we have the following convergence results for discrete adjoint approximations.

**Lemma 4.7.** Let $p, p_k$ solve (2.5) and (4.4), respectively, where $h \in L^2(I, L^2(\Omega))$. Then
\[
\|p_k - p\|_I \leq Ck^2 \left( \|\partial_t^2 p\|_I + \|\partial_t \Delta p\|_I \right)
\]
with a positive constant $C$ independent of the time mesh size $k$.

One essential ingredient of our convergence analysis is given by the following result.

**Lemma 4.8.** Let $y$ and $y_k$ as in Lemma 4.1 and let $p_k(h) \in P_k$ denote the solution to (4.4) with right hand side $h$. Then
\[
\|p_k(y_k - y)\|_I \leq C k^2 \left( \|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I \right).
\]

**Proof.** The function $p_k(y_k - y)$ solves (4.4) with $p_k(T) = 0$ and $h = y_k - y$. Since the test functions $\tilde{y}$ are elements of $Y_k$ we by Galerkin orthogonality obtain the same solution for $p_k$ with right hand side $h = y_k - \mathcal{P}Y_k y$, i.e. $p_k(y_k - y) = p_k(y_k - \mathcal{P}Y_k y)$. Hence by Lemma 4.1 and Corollary 4.2 we obtain
\[
\|p_k(y_k - y)\|_I = \|p_k(y_k - \mathcal{P}Y_k y)\|_I \leq C \|y_k - \mathcal{P}Y_k y\|_I \leq C k^2 \left( \|\partial_t^2 y\|_I + \|\partial_t \Delta y\|_I \right),
\]
which is the claim.

## 5 Variational discretization of the optimal control problem $\mathcal{P}$

To approximate the optimal control problem $\mathcal{P}$ we apply variational discretization of [Hin05] w.r.t. time, where the Petrov Galerkin state discretization introduced in the previous section is applied, i.e. we consider the optimal control problem
\[
\min_{y_k \in Y_k, u \in U_{ad}} J(y_k, u) = \frac{1}{2} \|y_k - y_d\|^2_{L^2(I, L^2(\Omega))} + \alpha \|u\|^2_{L^2(\Omega)},
\]
subject to $y_k = S_k(Bu, y_0)$,

where $S_k$ is the solution operator associated to (4.1). This problem admits a unique solution $(\tilde{y}_k, \tilde{u}_k) \in Y_k \times U_{ad}$, where $\tilde{y}_k = S_k(B\tilde{u}_k, y_0)$. The first order necessary and sufficient optimality condition for problem $\mathcal{P}_k$ reads
\[
\tilde{u}_k = P_{U_{ad}} \left( - \frac{1}{\alpha} B^t \tilde{p}_k \right),
\]
where $\tilde{p}_k \in P_k$ denotes the discrete adjoint variable, which is the unique solution to
\[
A(\tilde{y}_k, \tilde{p}_k) = \int_0^T \int_\Omega (\tilde{y}_k - y_d) \tilde{y} dx dt \text{ for all } \tilde{y}_k \in Y_k.
\]

Equation (5.1) is amenable to numerical treatment although the controls are not discretized explicitly, see [Hin05]. It is possible to implement a globalized semismooth Newton strategy in order to solve (5.1) numerically, see [HV12].

First let us establish an error estimate that resembles the standard estimate for variationally discretized problems. To begin with we for $v \in \mathcal{U}$ set $y(v) := \mathcal{S}(Bv, y_0)$ and denote with $y_k(v)$ the solution to (4.1) with $f := Bv$. Furthermore, we for $h \in L^2(I, H^{-1}(\Omega))$ denote with $p_k(h)$ the solution to (4.4).
Lemma 5.1. Let $\bar{u}$ and $\bar{u}_k$ solve $[P]$ and $[P_k]$, respectively. Then there holds
\[
\alpha |\bar{u}_k - \bar{u}|^2_I \leq \left\langle B'\left(p_k(\bar{y} - y_d) - \bar{p} + p_k(y_k(\bar{u})) - p_k(\bar{y})\right), \bar{u} - \bar{u}_k \right\rangle_{L^2(I,\mathbb{R}^p)}.
\]

Proof. We note that (2.1) and (5.1) can be equivalently expressed as
\[
\langle \alpha \bar{u} + B' \bar{p}, \bar{u} - u \rangle_{L^2(I,\mathbb{R}^p)} \leq 0 \quad \forall u \in U_{ad},
\]
(2.1)
\[
\langle \alpha \bar{u}_k + B' \bar{p}_k, \bar{u}_k - u \rangle_{L^2(I,\mathbb{R}^p)} \leq 0 \quad \forall u \in U_{ad}.
\]
(5.1)

Now inserting $\bar{u}_k$ into (2.1) and $\bar{u}$ into (5.1) and adding the resulting inequalities yields
\[
\left\langle \alpha (\bar{u}_k - \bar{u}) + B'(\bar{p}_k - \bar{p}), \bar{u}_k - \bar{u} \right\rangle_{L^2(I,\mathbb{R}^p)} \leq 0.
\]

After some simple manipulations we obtain
\[
\alpha |\bar{u}_k - \bar{u}|^2_I \leq \left\langle B'\left(p_k(\bar{y} - y_d) - \bar{p} + p_k(y_k(\bar{u})) - p_k(\bar{y})\right), \bar{u} - \bar{u}_k \right\rangle_{L^2(I,\mathbb{R}^p)}
\]
\[
+ \left\langle B'(\bar{p}_k - p_k(y_k(\bar{u}) - y_d)), \bar{u} - \bar{u}_k \right\rangle_{L^2(I,\mathbb{R}^p)}
\]
\[
- \|\bar{y}_k - y_k(\bar{u})\|^2_I
\]
\[
\leq \left\langle B'(p_k(\bar{y} - y_d) - \bar{p} + p_k(y_k(\bar{u})) - p_k(\bar{y})), \bar{u} - \bar{u}_k \right\rangle_{L^2(I,\mathbb{R}^p)}
\]
which is the desired estimate.

We are now in the position to formulate our main result.

Theorem 5.2. Let $\bar{u}$ and $\bar{u}_k$ denote the solutions to $[P]$ and $[P_k]$, respectively. Then there exists a constant $C > 0$ independent of $k$, such that
\[
\alpha |\bar{u}_k - \bar{u}|^2_I \leq C k^2 \left(\|\bar{u}\|_{H^1(I,\mathbb{R}^p)} + \|y_d\|_{H^1(I, L^2(\Omega))} + \|y_d(T)\|_{H^1(\Omega)} + \|y_0\|_{H^1(\Omega)} + \|\Delta y_0\|_{H^1(\Omega)}\right)
\]
(5.3)
is satisfied.

Proof. Making use of the continuity of $B$ and $B'$, compare (1.3) and (2.3), we directly infer from Lemma 5.1
\[
\alpha |\bar{u}_k - \bar{u}|^2_I \leq C \left(\|p_k(\bar{y} - y_d) - \bar{p}\|_I + \|p_k(y_k(\bar{u})) - p_k(\bar{y})\|_I\right)
\]
\[
\leq C k^2 \left(\|\partial^2 \bar{p}\|_I + \|\partial I \Delta \bar{p}\|_I + \|\partial I \Delta \bar{y}\|_I\right).
\]
The last estimate follows from the Lemmata 4.7 and 4.8. The claim is now a direct consequence of the Lemmata 3.1 and 3.2.

6 Numerical examples

We now construct numerical examples that validate our main result, i.e. Theorem 5.2. In both examples we make use of the fact that instead of the linear control operator $B$, given by (1.3), we can also use an affine linear control operator
\[
B : U \rightarrow L^2(I, H^{-1}(\Omega)), \quad u \mapsto g_0 + Bu.
\]
(6.1)
If we assume that $g_0$ is an element of $H^1(I, L^2(\Omega))$ with initial value $g_0(0) \in H^1_0(\Omega)$, all the preceding theory remains valid.
6.1 First example

The first example is taken from [MY11]. We recall it for convenience in our notation.

Given a space-time domain $\Omega \times I = (0, 1)^2 \times (0, 0.1)$, i.e. $D = 1$, we consider first the control operator $\tilde{B}$, which is fully characterized by means of the two functions

$$g_1(x_1, x_2) := \sin(\pi x_1) \sin(\pi x_2),$$

$$g_0(t, x_1, x_2) := -\pi^4 w_a(t, x_1, x_2) - BP_{U_{ad}} \left( -\frac{1}{4\alpha} \exp(a\pi^2 t) - \exp(a\pi^2 T) \right),$$

where

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

denote eigenfunctions of $\pm \partial_t - \Delta$. As a consequence we have

$$(B'z)(t) = \int_{\Omega} z(t, x_1, x_2) \cdot g_1(x_1, x_2) \, dx_1 dx_2,$$

compare (2.3). Note that we consider the adjoint of $B$, not of $\tilde{B}$. Furthermore we take

$$y_d(t, x_1, x_2) := \frac{a^2 - 5}{2 + a} \pi^2 w_a(t, x_1, x_2) + 2\pi^2 w_a(T, x_1, x_2),$$

and

$$u_0(x_1, x_2) := \frac{-1}{2 + a} \pi^2 w_a(0, x_1, x_2).$$

The admissible set $U_{ad}$ is defined by the bounds $a_1 := -25$ and $b_1 := -1$. Furthermore $\alpha := \pi^{-4}$ and $a := -\sqrt{5}$.

The exact solution of the optimal control problem (P) is given by

$$\bar{u}(t) = P_{U_{ad}} \left( -\frac{1}{4\alpha} \exp(a\pi^2 t) - \exp(a\pi^2 T) \right),$$

$$\bar{y}(t, x_1, x_2) = \frac{-1}{2 + a} \pi^2 w_a(t, x_1, x_2),$$

and

$$\bar{p}(t, x_1, x_2) = w_a(t, x_1, x_2) - w_a(T, x_1, x_2).$$

Note that this example fulfills the Assumption 1.1.

We solve this problem numerically using a fixpoint iteration on equation (5.1). We discretize in space with a fixed number of nodes $N_h = (2^7 + 1)^2 = 16641$. We examine the behavior of the temporal convergence by considering a sequence of meshes with $N_k = (2^\ell + 1)^2$ nodes at refinement levels $\ell = 1, 2, 3, 4, 5, 6$. Each fixpoint iteration is initialized by the starting value $u_{kh} := a_1$. As a stopping criterion we require

$$\|B' \left( p_{kh}^{\text{new}} - p_{kh}^{\text{old}} \right) \|_{L^\infty(\Omega \times I)} < t_0,$$

where $t_0 := 10^{-5}$ is a prescribed threshold.

Table 1 shows the behavior of several errors in time between the exact control $\bar{u}$ and its discretized computed counterpart $u_{kh}$, obtained by the fixpoint iteration. Furthermore, the
Table 1: First example: Errors and EOC in the control.

| ℓ | $\|\bar{u} - u_{kh}\|_{L^1(I,L^1(\Omega))}$ | $\|\bar{u} - u_{kh}\|_{L^2(L^2)}$ | $\|\bar{u} - u_{kh}\|_{L^\infty(L^\infty)}$ | EOC$_{L^1}$ | EOC$_{L^2}$ | EOC$_{L^\infty}$ |
|---|---|---|---|---|---|---|
| 1 | 0.07338346 | 0.31701554 | 1.97701729 | / | / | / |
| 2 | 0.01653824 | 0.08052755 | 0.66237792 | 2.15 | 1.98 | 1.58 |
| 3 | 0.00396507 | 0.01977927 | 0.19440662 | 2.06 | 2.03 | 1.77 |
| 4 | 0.00088306 | 0.00448012 | 0.05014900 | 2.17 | 2.14 | 1.95 |
| 5 | 0.00017870 | 0.00083749 | 0.00970228 | 2.31 | 2.42 | 2.37 |
| 6 | 0.00018581 | 0.00068442 | 0.00462541 | -0.06 | 0.29 | 1.07 |

Table 2: First example: Errors and EOC in the state.

| ℓ | $\|\bar{y} - y_{kh}\|_{L^1(I,L^1(\Omega))}$ | $\|\bar{y} - y_{kh}\|_{L^2(L^2)}$ | $\|\bar{y} - y_{kh}\|_{L^\infty(L^\infty)}$ | EOC$_{L^1}$ | EOC$_{L^2}$ | EOC$_{L^\infty}$ |
|---|---|---|---|---|---|---|
| 1 | 0.19002993 | 0.96055898 | 14.78668742 | / | / | / |
| 2 | 0.09429883 | 0.49287844 | 9.02297459 | 1.01 | 0.96 | 0.71 |
| 3 | 0.04706727 | 0.24798983 | 5.06528533 | 1.00 | 0.99 | 0.83 |
| 4 | 0.02352485 | 0.12419027 | 2.69722511 | 1.00 | 1.00 | 0.91 |
| 5 | 0.01176627 | 0.06216408 | 1.39374255 | 1.00 | 1.00 | 0.95 |
| 6 | 0.00588802 | 0.03119134 | 0.70870727 | 1.00 | 0.99 | 0.98 |

An *experimental order of convergence* (EOC) is given. The table indicates an error behavior of $O(k^2)$ for the $L^2$ error in the control, which is in accordance with Theorem 5.2. Furthermore, the error of the adjoint, see table 4, shows the same behavior, as expected. Since the state is discretized piecewise constant in time, the order of convergence is only one. However, without further numerical effort we obtain a second

Table 3: First example: Errors and EOC in the projected state.

| ℓ | $\|\bar{y} - \pi_{\ell} Y_{kh}\|_{L^1(I,L^1(\Omega))}$ | $\ldots$ | $\ldots$ | $\ldots$ | EOC$_{L^1}$ | EOC$_{L^2}$ | EOC$_{L^\infty}$ |
|---|---|---|---|---|---|---|---|
| 1 | 0.10937032 | 0.48300664 | 6.29978738 | / | / | / |
| 2 | 0.02713496 | 0.13212665 | 1.94510739 | 2.01 | 1.87 | 1.70 |
| 3 | 0.00720221 | 0.03723408 | 0.61346273 | 1.91 | 1.83 | 1.66 |
| 4 | 0.00183081 | 0.00982563 | 0.17399005 | 1.98 | 1.92 | 1.82 |
| 5 | 0.00042588 | 0.00242796 | 0.04646078 | 2.10 | 2.02 | 1.90 |
| 6 | 0.00009796 | 0.00054833 | 0.01201333 | 2.12 | 2.15 | 1.95 |

The results are consistent with the theoretical predictions.
order convergent approximation of the state with the projection $\pi P_k y_k$ of the discrete state $y_k$, compare Lemma 4.5 and see Table 3 for the corresponding numerical results. In practise this means that we can gain a better approximation of the state without further effort; we only have to interpret the discrete state vector $\bar{\vec{y}}_k$, i.e. the vector containing the value of $y_k$ on each interval $I_m$, in the right way, namely as a vector of values on the gridpoints of the dual grid $t^*_1 < ... < t^*_M$.

Figure 1 illustrates the convergence of $u_{kh}$ to $\bar{u}$. Note that the intersection points between active and inactive set need not coincide with time grid points since we use variational discretization. Let us further note that the number of fixpoint iterations does not depend on the fineness of the time grid size. In our example four iterations are needed to reach the above mentioned threshold of $t_0 := 10^{-5}$.

### 6.2 Second example

This example is a slight variant of the first one yielding more intersection points between the active and inactive set. With the space-time domain $\Omega \times I = (0,1)^2 \times (0,0.5)$, we set

$$\bar{y}(t,x_1,x_2) := w_a(t,x_1,x_2) := \cos \left( \frac{t}{I} \cdot 2\pi a \right) \cdot g_1(x_1,x_2),$$

$$\bar{p}(t,x_1,x_2) = w_a(t,x_1,x_2) - w_a(T,x_1,x_2),$$

Table 4: First example: Errors and EOC in the adjoint.

| $\ell$ | $\| \bar{\bar{p}} - p_{kh} \|_{L^1(I,L^1(\Omega))}$ | $\| \bar{\bar{p}} - p_{kh} \|_{L^2(L^2)}$ | $\| \bar{\bar{p}} - p_{kh} \|_{L^\infty(L^\infty)}$ | EOC$_{L^1}$ | EOC$_{L^2}$ | EOC$_{L^\infty}$ |
|-------|-------------------------------------------------|--------------------------------------|--------------------------------------|--------------|--------------|--------------|
| 1     | 0.00125773                                       | 0.00652756                           | 0.08119466                           | /            | /            | /            |
| 2     | 0.00029007                                       | 0.00166280                           | 0.02721190                           | 2.12         | 1.97         | 1.58         |
| 3     | 0.00006933                                       | 0.00040888                           | 0.00799559                           | 2.06         | 2.02         | 1.77         |
| 4     | 0.00001564                                       | 0.00009340                           | 0.00207127                           | 2.15         | 2.13         | 1.95         |
| 5     | 0.00000310                                       | 0.00001739                           | 0.00041008                           | 2.34         | 2.43         | 2.34         |
| 6     | 0.00000273                                       | 0.00001246                           | 0.00017768                           | 0.18         | 0.48         | 1.21         |

Figure 1: First example: Optimal control $\bar{u}$ (solid) and $u_{kh}$ (dashed) after refinement level $\ell$. 
where $g_1$ is defined in the first example. Consequently,

$$g_0 = g_1 2\pi \left( -\frac{a}{T} \sin \left( \frac{t}{T} 2\pi a \right) + \pi \cos \left( \frac{t}{T} 2\pi a \right) \right) - B\bar{u},$$

$$y_d = g_1 \left( \cos \left( \frac{t}{T} 2\pi a \right) \left( 1 - 2\pi^2 \right) - \frac{2\pi a}{T} \sin \left( \frac{t}{T} 2\pi a \right) + 2\pi^2 \right),$$

and

$$\bar{u} = P_{Uad} \left( -\frac{1}{4\alpha} \cos \left( \frac{t}{T} 2\pi a \right) + \frac{1}{4\alpha} \right).$$

Furthermore, we set $\alpha = 1$, $a_1 := 0.2$, $b_1 := 0.4$ and $a := 2$. Note that this example also fulfills the Assumption 1.1.

We now consider refinement levels $\ell = 1, 2, 3, 4, 5, 6, 7, 8$ and proceed as described in the first example. We obtain the same EOCs for control, state, and adjoint, see the tables 5, 6, 7, 8, and figure 2.

| $\ell$ | $\|\bar{u} - u_{kh}\|_{L_1^1(L_1^1(\Omega))}$ | $\|\bar{u} - u_{kh}\|_{L_2^2}$ | $\|\bar{u} - u_{kh}\|_{L_\infty(L_\infty)}$ | EOC$_{L_1}$ | EOC$_{L_2}$ | EOC$_{L_\infty}$ |
|-------|---------------------------------|-----------------|-----------------|-----------|-----------|-------------|
| 1     | 0.04925427                      | 0.09237138      | 0.20000000      | /        | /        | /           |
| 2     | 0.00256632                      | 0.01106114      | 0.07336869      | 4.26     | 3.06     | 1.45        |
| 3     | 0.00430215                      | 0.01144324      | 0.04704583      | -0.65    | -0.05    | 0.64        |
| 4     | 0.00069342                      | 0.00204495      | 0.00893696      | 2.54     | 2.48     | 2.40        |
| 5     | 0.00016762                      | 0.00050729      | 0.00249463      | 2.05     | 2.01     | 1.84        |
| 6     | 0.00003989                      | 0.00011939      | 0.00064497      | 2.07     | 2.09     | 1.95        |
| 7     | 0.00000948                      | 0.00003227      | 0.00020672      | 2.07     | 1.89     | 1.64        |
| 8     | 0.00000764                      | 0.00002142      | 0.00009457      | 0.31     | 0.59     | 1.13        |

Table 5: Second example: Errors and EOC in the control.
| ℓ | $∥\bar{y} - y_{kh}\parallel_{L^1(\Omega)}$ | $∥\bar{y} - y_{kh}\parallel_{L^2(\Omega)}$ | $∥\bar{y} - y_{kh}\parallel_{L^\infty(\Omega)}$ | EOC$_{L^1}$ | EOC$_{L^2}$ | EOC$_{L^\infty}$ |
|---|---|---|---|---|---|---|
| 1 | 0.19657193 | 0.41315218 | 2.24553307 | / | / | / |
| 2 | 0.13005269 | 0.25408123 | 1.25552256 | 0.60 | 0.70 | 0.84 |
| 3 | 0.05650537 | 0.11224959 | 0.6597254 | 1.20 | 1.18 | 0.93 |
| 4 | 0.02611675 | 0.05637041 | 0.38210207 | 1.11 | 0.99 | 0.79 |
| 5 | 0.01277289 | 0.02827337 | 0.19029296 | 1.03 | 1.00 | 1.01 |
| 6 | 0.00635223 | 0.01418903 | 0.09710641 | 1.01 | 0.99 | 0.97 |
| 7 | 0.00317298 | 0.00718111 | 0.04892792 | 1.00 | 0.98 | 0.99 |
| 8 | 0.00158730 | 0.00375667 | 0.02456764 | 1.00 | 0.93 | 0.99 |

Table 6: Second example: Errors and EOC in the state.

| ℓ | $∥\bar{y} - \pi P_k y_{kh}\parallel_{L^1(\Omega)}$ | $∥\cdots\parallel_{L^2(\Omega)}$ | $∥\cdots\parallel_{L^\infty(\Omega)}$ | EOC$_{L^1}$ | EOC$_{L^2}$ | EOC$_{L^\infty}$ |
|---|---|---|---|---|---|---|
| 1 | 0.19734452 | 0.42154165 | 2.65669891 | / | / | / |
| 2 | 0.13173168 | 0.25800727 | 1.39668789 | 0.58 | 0.71 | 0.93 |
| 3 | 0.03422500 | 0.07418402 | 0.40783930 | 1.94 | 1.80 | 1.78 |
| 4 | 0.01080693 | 0.02168391 | 0.15176831 | 1.66 | 1.77 | 1.43 |
| 5 | 0.00282859 | 0.00567595 | 0.04685968 | 1.93 | 1.93 | 1.70 |
| 6 | 0.00071212 | 0.00143268 | 0.01229008 | 1.99 | 1.99 | 1.93 |
| 7 | 0.00017551 | 0.00035509 | 0.00311453 | 2.02 | 2.01 | 1.98 |
| 8 | 0.00004104 | 0.00008530 | 0.00078765 | 2.10 | 2.06 | 1.98 |

Table 7: Second example: Errors and EOC in the projected state.

Acknowledgements

Christian Kahle provided some hints on the numerical implementation which are gratefully acknowledged.

References

[Ev98] Lawrence C. Evans. *Partial Differential Equations*. AMS, 1998.

[Hin05] Michael Hinze. A Variational Discretization Concept in Control Constrained Optimization: The Linear-Quadratic Case. *Computational Optimization and Applications*, 30(1):45–61, 2005.
\begin{table}
\begin{tabular}{ccccccc}
\hline
$\ell$ & $\|\bar{p} - p_{kh}\|_{L^1(L^1(\Omega))}$ & $\|\bar{p} - p_{kh}\|_{L^2(L^2)}$ & $\|\bar{p} - p_{kh}\|_{L^\infty(L^\infty)}$ & EOC$_{L^1}$ & EOC$_{L^2}$ & EOC$_{L^\infty}$ \\
\hline
1 & 0.20659855 & 0.46853028 & 2.86360259 & / & / & / \\
2 & 0.03491931 & 0.08118048 & 0.56829981 & 2.56 & 2.53 & 2.33 \\
3 & 0.01994220 & 0.04100552 & 0.20495644 & 0.81 & 0.99 & 1.47 \\
4 & 0.00440890 & 0.00895349 & 0.05815307 & 2.18 & 2.20 & 1.82 \\
5 & 0.00105993 & 0.00215639 & 0.01668075 & 2.06 & 2.05 & 1.80 \\
6 & 0.00026116 & 0.00053258 & 0.00447036 & 2.02 & 2.02 & 1.90 \\
7 & 0.00006984 & 0.00014824 & 0.00116014 & 1.90 & 1.85 & 1.95 \\
8 & 0.00004199 & 0.00008530 & 0.00046798 & 0.73 & 0.80 & 1.31 \\
\hline
\end{tabular}
\caption{Second example: Errors and EOC in the adjoint.}
\end{table}

[HPUU09] Michael Hinze, Rene Pinnau, Michael Ulbrich, and Stefan Ulbrich. Optimization with PDE Constraints. Springer, 2009.

[HV12] Michael Hinze and Morten Vierling. The semi-smooth Newton method for variationally discretized control constrained elliptic optimal control problems; implementation, convergence and globalization. Optimization Methods and Software, 27(6):933–950, 2012.

[MV08a] Dominik Meidner and Boris Vexler. A Priori Error Estimates for Space-Time Finite Element Discretization of Parabolic Optimal Control Problems Part I: Problems Without Control Constraints. SIAM Journal on Control and Optimization, 47(3):1150–1177, 2008.

[MV08b] Dominik Meidner and Boris Vexler. A Priori Error Estimates for Space-Time Finite Element Discretization of Parabolic Optimal Control Problems Part II: Problems with Control Constraints. SIAM Journal on Control and Optimization, 47(3):1301–1329, 2008.

[MV11] Dominik Meidner and Boris Vexler. A priori error analysis of the Petrov Galerkin Crank Nicolson scheme for parabolic optimal control problems. SIAM Journal on Control and Optimization, 49(5):2183–2211, 2011.

[Ran84] Rolf Rannacher. Finite Element Solution of Diffusion Problems with Irregular Data. Numerische Mathematik, 43:309–327, 1984.

[SV13] Andreas Springer and Boris Vexler. Third order convergent time discretization for parabolic optimal control problems with control constraints. Computational Optimization and Applications, pages 1–36, 2013.