Optimal Finite Element Error Estimates for an Optimal Control Problem governed by the Wave Equation with controls of Bounded Variation.

S. Engel †, P. Trautmann ‡, B. Vexler §,

[Received on ...]


This work discusses the finite element discretization of an optimal control problem for the linear wave equation with time-dependent controls of bounded variation. The main focus lies on the convergence analysis of the discretization method. The state equation is discretized by a space-time finite element method. The controls are not discretized. Under suitable assumptions optimal convergence rates for the error in the state and control variable are proven. Based on a conditional gradient method the solution of the semi-discretized optimal control problem is computed. The theoretical convergence rates are confirmed in a numerical example.

Keywords: BV-Functions; optimal control of a wave equation; error bounds; finite elements.

1. Introduction

In this paper we derive a priori error estimates for a finite element discretization of the following optimal control problem governed by the linear wave equation:

$$\min_{u \in BV(0,T)^m} \frac{1}{2} \|y_u - y_d\|_{L^2(\Omega_T)}^2 + \sum_{j=1}^m \alpha_j \|D_i u_j\|_{M(I)} =: J(y,u)$$

subject to

$$\left\{ \begin{array}{ll}
\partial_t y - \Delta y = f + \sum_{j=1}^m u_j g_j & \text{in } I \times \Omega \\
y = 0 & \text{on } I \times \partial \Omega \\
(y, \partial_t y) = (y_0, y_1) & \text{in } \{0\} \times \Omega,
\end{array} \right.$$

where $\Omega \subset \mathbb{R}^n$, with $n \in \{1,2,3\}$, is a convex, polygonal/polyhedral bounded domain. For $T \in (0,\infty)$ we denote $I = (0,T)$. The desired state $y_d$ is assumed to satisfy $y_d \in C^1(I;L^2(\Omega))$. The time depending controls $u$ are given by $u = (u_1, \ldots, u_m) \in BV(0,T)^m$, and $BV(0,T)^m$ is endowed with the norm $\|u\|_{BV(I)^m} = \sum_{j=1}^m (||u_j||_{L^1(I)} + ||D_i u_j||_{M(I)})$. Here $M(I)$ is the space of Borel measures, endowed with the total variation norm $\| \cdot \|_{M(I)}$. Further, let $(g_j)_{j=1}^m \subset L^\infty(\Omega) \setminus \{0\}$ with pairwise disjoint supports and $\alpha_j > 0$. The initial data is chosen as $(y_0, y_1) \in H_0^1(\Omega) \times L^2(\Omega)$. Finally, we set $\Omega_T := I \times \Omega$.

In this work we focus on controls of bounded variation in time. By using the total variation norm in $(P)$, sparsity in the derivative of the controls is promoted, resulting in locally constant controls. This

1†Technical University of Munich, Chair for Optimal Control, Department of Mathematics, Boltzmannstr. 3, 85748 Garching b. München, Germany, (sebastian.engel@ma.tum.de). Funding: The first author’s work was supported by the International Research Training Group IGDK, funded by the German Science Foundation (DFG) and the Austrian Science Fund (FWF)

2‡University of Graz, Institute of Mathematics and Scientific Computing, Heinrichstraße 36, 8010 Graz, Austria (philip.trautmann@uni-graz.at)

3§Technical University of Munich, Chair for Optimal Control, Department of Mathematics, Boltzmannstr. 3, 85748 Garching b. München, Germany, (vexler@ma.tum.de)

© The author 2019. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.
is in particular the case if the derivative of the optimal control is a linear combination of Dirac functions. Optimal control problems with $BV$-controls are already analyzed for elliptic and parabolic state equations in Casas and Kunisch [2017], Casas et al. [2017], Hafemeyer et al. [2019], Casas et al. [1998, 1999], Clason and Kunisch [2011].

Since our article deals with a priori error estimates of a finite element discretization for the control problem $(P)$, we briefly discuss previous works on error estimates for PDE control problems with $BV$-controls. In Casas et al. [2017] the authors discretize the time-dependent $BV-$controls by cellwise constant functions. The state equation is discretized by piecewise constant finite elements in time and linear continuous finite elements in space. Based on this discretization approach, the authors show that the optimal value of the cost functional and the states converge with an order of $\sqrt{\tau}$ in time and linear in space. However, numerical experiments in Casas et al. [2017] indicate better results. In Hafemeyer et al. [2019] the authors analyze a finite element discretization of an elliptic control problem with $BV$-controls in a one dimensional setting. As in our case the controls are not discretized. The main contribution of this work is the derivation of optimal error estimates for the control variable in the $L^1$-norm. Their analysis relies on the one dimensional setting and on structural assumption on the optimal adjoint state which guarantee that the optimal control is piecewise constant and has finitely many jumps. In our work we derive similar optimal error estimates also for the problem with a multi-dimensional wave equation and our analysis relies partially on techniques developed in the former work.

Next we briefly address the difficulties in the derivation of finite element error estimates for optimal control problems with PDEs and $BV$-controls. Standard techniques for the derivation of finite element error estimates, see e.g. Casas and Tröltzsch [2012], cannot be applied due to the non-smoothness of the cost functional and the non-reflexivity of $BV(I)$. In the last years several papers concerning the derivation of finite element error estimates for optimal control problems with measure-valued controls appeared, see e.g. Pieper and Vexler [2013], Trautmann et al. [2018b]. Using the fact that for one dimensional controls, $BV(I)$ is isomorphic to $M(I) \times \mathbb{R}$, several techniques from these works are used to derive error estimates for $BV$-controls. Finally, we mention that the literature on finite element error estimates for optimal control problems governed by the wave equation is very limited. To our knowledge the only existing work in this context is Trautmann et al. [2018b] which uses the space-time finite element discretization developed and analyzed in Zlotnik [1994]. Our work also relies on this discretization method for the state equation and its error analysis. The main contribution of this work is the derivation of an optimal error estimate of the control variable in the $L^1(I)$-norm and of the state variable in the $L^2(\Omega_T)$-norm. The state equation is discretized by a space-time finite element method with piecewise linear and continuous Ansatz- and test-functions from Zlotnik [1994]. The weak formulation of the discrete state equation is augmented with a stabilization term involving the stabilization parameter $\sigma$. Stability of the method depends on the value of this parameter. Moreover, for certain values of this parameter the method is equivalent to wellknown time stepping schemes, like the Crank-Nicolson scheme or the Leap-Frog scheme. The $BV$-controls are not discretized.

Due to fact that the controls are only time-dependent the problem under consideration can be reformulated as a measure-valued control problem. Based on the optimality conditions of the continuous and discrete optimal control problem the error in the state variable in the $L^2(\Omega_T)$-norm can be represented in terms of the finite element error of the state and adjoint state equation in the $L^2(\Omega_T)$-norm resp. the $L^\infty(I;L^2(\Omega))$ as well as the error in the control variable in the $L^1(I)$-norm. The convergence rates for the finite element error of the state and adjoint state are obtained from Zlotnik [1994]. Under the assumption that the continuous and time dependent function

$$\tilde{p}_{1,t}: I \mapsto - \int_t^T \int_\Omega p g_i \, dx \, ds,$$
We consider $\Omega$ where weak solutions of the wave equation. In section 3 the space-time finite element method from Zlotnik this work has the following structure. Section 2 summarizes several needed results on the regularity of variable in the $L$ velocity $y$ a weak solution of $(2.1)$ $(\Omega_f)$ it follows that the continuous optimal control is piecewise constant and has finitely many jumps. To obtain this information about the form of the optimal BV control, using $\tilde{p}_{1,i}$, is particularly easy because we consider controls in one dimension. Furthermore, it is proven that the optimal control of the discrete problem has the same number jumps which are located close to the jumps of the continuous optimal control. Using these properties the error of the optimal control in the $L^1(\Omega)$-norm is estimated in terms of the error of the state variable in the $L^2(\Omega_f)$-norm. Using a bootstrapping argument optimal rates for the error in the state and control variable as well as for the optimal value of the cost are proven. These rates are confirmed by a numerical example with known solution.

This work has the following structure. Section 2 summarizes several needed results on the regularity of weak solutions of the wave equation. In section 3 the space-time finite element method from Zlotnik [1994] is presented. Moreover, important stability results as well as a priori error estimates are stated. Section 4 deals with the reformulation of the BV-control problem as a measure-valued control problem and with the analysis of this problem. In particular, first order optimality conditions are derived. The next section 5 is concerned with discretization of the control problem. It is based on the mentioned space-time finite element method and the variational discretization concept. In section 6 the error estimates for the optimal state and control variable as well as the optimal functional value are derived. Finally, in section 7 a generalized conditional gradient method is introduced which applicable in the context of controls which are not discretized. Based on this method a problem with known solution is solved and the theoretical error estimates are confirmed.

2. Preliminaries on the Wave Equation

We consider $\Omega \subset \mathbb{R}^d$, $d = 1, 2, 3$ as convex, polygonal/polyhedral domain. Let $\{\lambda_k\}_{k \in \mathbb{N}}$ be the non-decreasing eigenvalues of the Laplace operator $-\Delta$ with homogeneous boundary conditions and let $\{\mu_k\}_{k \in \mathbb{N}}$ be the corresponding system of eigenfunctions, which are orthonormal complete in $L^2(\Omega)$, and orthogonal complete in $H^1_0(\Omega)$. Hence, let us introduce for $\alpha \geq 0$ the Hilbert spaces

\[ H^\alpha = \left\{ w \in L^2(\Omega) \left| \| w \|_{H^\alpha}^2 := \sum_{k \geq 1} \left( \lambda_k \right)^\alpha \| w \|_{L^2(\Omega)}^2 < \infty \right\} \right. \]

For $\alpha = 0, 1$ we get $L^2(\Omega)$ respectively $H^1_0(\Omega)$. The convexity of $\Omega$ implies that $H^2 = H^2(\Omega) \cap H^1_0(\Omega)$. In general holds $H^\alpha \hookrightarrow H^\beta$ for $\alpha \geq \beta$. We denote the dual space of $H^\alpha$ by $H^{-\alpha}$. Next we introduce the weak solution of the wave equation with the forcing function $f$, initial displacement $y_0$, and initial velocity $y_1$.

**Definition 2.1** [Ladyzenskaya 1973, Chap.IV, Sec.4]

Let $(f, y_0, y_1) \in L^1(I; L^2(\Omega)) \times H^1_0(\Omega) \times L^2(\Omega)$. We call a function $y \in C(I; H^1_0(\Omega))$ with $\partial_t y \in C(I; L^2(\Omega))$ a weak solution of $(2.1)$, if

\[ \int_0^T -\langle \partial_t y, \partial_t \eta \rangle_{L^2(\Omega)} + \langle \nabla y, \nabla \eta \rangle_{L^2(\Omega)} dt = \langle y_1, \eta(0) \rangle_{L^2(\Omega)} + \int_0^T \langle f, \eta \rangle_{L^2(\Omega)} dt \quad (2.1) \]

for any $\eta \in L^1(I; H^1_0(\Omega))$ such that $\partial_t \eta \in L^1(I; L^2(\Omega))$, $\eta|_{t=T} = 0$, and $y$ satisfies the initial condition $y|_{t=0} = y_0$.

For the following existence and regularity results of weak solutions of the wave equation we refer to Zlotnik [1994, Proposition 1.1., 1.3.]:
We discretize the time interval \( I \) provided \( f \), solution piecewise linear and continuous functions with respect to \( y \) and its nodal basis by \( y \).

**Proof.** The proof can be found in [Zlotnik (1994), Proposition 1.3, Remark 1.2].

**Definition 2.3** Let us define the following continuous linear operators:

\[
L: L^2(\Omega_T) \rightarrow L^2(\Omega_T), \quad f \mapsto y(f) \quad \text{and} \quad Q: H^1_0(\Omega) \times L^2(\Omega) \rightarrow L^2(\Omega_T), \quad (y_0, y_1) \mapsto y(y_0, y_1)
\]

The function \( y(f) \) denotes the weak solution of the wave equation with \( y_0 = y_1 = 0 \) and forcing function \( f \). The function \( y(y_0, y_1) \) denotes the weak solution of the wave equation with initial datum \( y_0 \) and \( y_1 \) and \( f = 0 \).

**Lemma 2.1** The adjoint operator \( L^* : L^2(\Omega_T) \rightarrow L^2(\Omega_T) \) of \( L \) is given by \( w \mapsto p(w) \) where \( p(w) \in C(T; H^1_0(\Omega)) \cap C(T; L^2(\Omega)) \) is the weak solution of the backwards in time equation

\[
(\mathcal{W}^*) \begin{cases} \partial_t p - \Delta p = w & \text{in} \quad I \times \Omega \\ p = 0 & \text{on} \quad I \times \partial \Omega \\ (p, \partial_t p) = (0, 0) & \text{in} \quad \{T\} \times \Omega. \end{cases}
\]

**Lemma 2.2** Let \( y \) be a weak solution of \( \mathcal{W} \) for \( (y_0, y_1, 0) \) and \( p \) of \( \mathcal{W}^* \) for \( w \). There holds

\[
\int_0^T (y, w)_{L^2(\Omega)} \, dt = \left( y_0, \int_0^T w \, dt \right)_{L^2(\Omega)} - \left( \nabla y_0, \nabla \int_0^T p \, dt \right)_{L^2(\Omega)} + (y_1, p(0))_{L^2(\Omega)}.
\]

**Proof.** This proven by testing (2.1) for \( p \) with \( \tilde{y} = y - y_0 \).

### 3. Approximation of the Wave Equation

In the following we introduce the space-time finite element method for the discretization of the wave equation. This method can be found in [Zlotnik (1994)]. We consider a mesh \( \mathcal{T}_h \) consisting of a finite set of triangles (for \( d = 2 \)) or tetrahedra (for \( d = 3 \)) \( K \) with \( h = \max_{K \in \mathcal{T}_h} \rho(K) \), where \( \rho(K) \) denotes the diameter of \( K \). We assume that the family of meshes \( (\mathcal{T}_h)_h \) is admissible, shape regular and quasi-uniform. Since \( \Omega \) is polygonal and convex, we require that \( \Omega = \bigcup_{K \in \mathcal{T}_h} K \) holds. We denote the space of piecewise linear and continuous finite elements based on the triangulation by \( \mathcal{T}_h \) by \( S_h \subset H^1_0(\Omega) \cap C(\Omega) \) and its nodal basis by \( (\phi_i)_{i=1}^N \).

#### 3.1 Space-Time Finite Element Method

We discretize the time interval \( I \) uniformly with the time nodes \( 0 = t_0 < \ldots < t_M = T \) and the stepsize \( \tau = T/M \). We denote the set of time nodes by \( \mathcal{T} = \{t_0, \ldots, t_M\} \). Then we introduce the space of piecewise linear and continuous functions with respect to \( \mathcal{T} \) by

\[
S_{\mathcal{T}} := \left\{ w \in C(\mathcal{T}) \mid w|_{[t_{k-1}, t_k]} \text{ linear}, \ 1 \leq k \leq M \right\}.
\]
Let $\sigma \geq 0$. We call $y_\theta \in \tilde{S}_\theta := \text{span}\{v_h \cdot v_r | v_h \in S_h, v_r \in S_r\}$ a discrete solution of (2.1) if $y_\theta$ satisfies:

$$
\int_0^T -\langle \partial_t y_\theta, \partial_t \eta \rangle_{L^2(\Omega)} - \left( \sigma - \frac{1}{6} \right) \tau^2 \langle \nabla \partial_t y_\theta, \nabla \partial_t \eta \rangle_{L^2(\Omega)} + \langle \nabla y_\theta, \nabla \eta \rangle_{L^2(\Omega)} dt = \langle y_1, \eta(0) \rangle_{L^2(\Omega)} + \int_0^T \langle f, \eta \rangle_{L^2(\Omega)} dt \tag{3.1}
$$

for all $\eta \in \tilde{S}_\theta$ with $\eta(T) = 0$ and initial condition $y_\theta(0) := R_h y_0$, where $R_h$ is the Ritz projection on $S_h$, i.e.

$$
\langle \nabla R_h y_0, \nabla \phi \rangle_{L^2(\Omega)} = \langle \nabla y_0, \nabla \phi \rangle_{L^2(\Omega)} \quad \forall \phi \in S_h.
$$

**Remark 3.1** Here $\sigma$ plays the role of a stabilization parameter. With an increasing value of $\sigma$ the method becomes more stable. For $\sigma \geq 1/4$ the method is unconditionally stable, see [Zlotnik 1994].

**Lemma 3.1** Let $y_\theta \in \tilde{S}_\theta$ be a solution of (3.1) for $(y_0, y_1, 0)$ and $p_\theta \in \tilde{S}_\theta$ a corresponding discrete solution of $W^*$ for $w$. There holds

$$
\int_0^T \langle y_\theta, w \rangle_{L^2(\Omega)} dt = \left( R_h y_0, \int_0^T w dt \right)_{L^2(\Omega)} - \left( \nabla R_h y_0, \nabla \int_0^T p_\theta dt \right)_{L^2(\Omega)} + \langle y_1, p_\theta(0) \rangle_{L^2(\Omega)}
$$

**Proof.** This is proven by testing (3.1) for $p_\theta$ with $\tilde{y}_\theta = y_\theta - R_h y_0$. \hfill $\Box$

### 3.2 A Priori Error Estimates for the Space-Time Finite Element Method

Next we make an assumption on the relationship between $\tau$ and $h$ which ensures stability of the method for $0 \leq \sigma < 1/4$.

**Assumption 1** Let $\varepsilon_0 \in (0, 1]$ be arbitrary and fixed. Moreover, let $c_1$ be the smallest constant in the inverse inequality $\| \nabla \phi \|_{L^2(\Omega)} \leq c_1 h^{-1} \| \phi \|_{L^2(\Omega)}$ for all $\phi \in S_h$. Moreover, let a $c_2$ be the constant in this a priori estimate for the Ritz projection $\| w - R_h w \|_{L^2(\Omega)} \leq c_2 h \| \nabla w \|_{L^2(\Omega)}$. From now on it is assumed that

1. $\sigma \geq \frac{1}{4} - \frac{c_1 h^2 (1-\varepsilon_0^2)}{\tau^2}$,

2. $\sigma \geq \frac{1+\varepsilon_0^2}{4} - \frac{c_1 h^2}{\tau^2}$,

3. $|\sigma|\tau^2 \leq 2(c_2 h^2 + \tau^2)$.

**Remark 3.2** This space-time finite element method is related to well-known time-stepping schemes. For $\sigma = 0$ it is related to the explicit Leap-Frog-method and for $\sigma = \frac{1}{4}$ to the Crank-Nicolson scheme, see also [Trautmann et al. 2018a, Remark 5.1, 5.4]. A more detailed discussion can be found in [Zlotnik 1994].

Let us further define $\| u \|_{C(T; L^2(\Omega))} := \max_{t \in \tau} \| u(t) \|_{L^2(\Omega)}$ for all $u \in C(T; L^2(\Omega))$. 

The standard hat functions form a basis $e_m(t_k) = \delta_{mk}$, $m, k = 0, \ldots, M$ of this discrete space. Furthermore, let us define the mesh operators $\delta_i w_m = \frac{w_m - w_{m-1}}{\tau}$. Finally, we use the notation $\theta := (\tau, h)$ with $\tau, h > 0$.

**Definition 3.1** Let $\sigma \geq 0$. We call $\tilde{y}_\theta(\tau, h) := \{y_\theta = \sum_{i=0}^N y_i \varphi_i(\tau, h) | y_i \in S_h \}$ a space-time finite element solution of (2.1) if $\tilde{y}_\theta$ satisfies:

$$
\int_0^T \langle \partial_t y_\theta, \partial_t \eta \rangle_{L^2(\Omega)} - \left( \sigma - \frac{1}{6} \right) \tau^2 \langle \nabla \partial_t y_\theta, \nabla \partial_t \eta \rangle_{L^2(\Omega)} + \langle \nabla y_\theta, \nabla \eta \rangle_{L^2(\Omega)} dt = \langle y_1, \eta(0) \rangle_{L^2(\Omega)} + \int_0^T \langle f, \eta \rangle_{L^2(\Omega)} dt \tag{3.1}
$$

for all $\eta \in \tilde{S}_\theta$ with $\eta(T) = 0$ and initial condition $y_\theta(0) := R_h y_0$, where $R_h$ is the Ritz projection on $S_h$, i.e.

$$
\langle \nabla R_h y_0, \nabla \phi \rangle_{L^2(\Omega)} = \langle \nabla y_0, \nabla \phi \rangle_{L^2(\Omega)} \quad \forall \phi \in S_h.
$$
Lemma 3.2  The solution $y_\theta$ of (3.1) for $(f, y_0, y_1) \in L^1(I, L^2(\Omega)) \times H^1_0(\Omega) \times L^2(\Omega)$ satisfies the following inequality

$$
\|y_\theta\|_{C(T; L^2(\Omega))} \leq c \left( \|y_0\|_{H^1_0(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|f\|_{L^1(I; L^2(\Omega))} \right)
$$

with a constant $c$ independent of $h, f, y_0$ and $y_1$.

Proof. The result follows directly from [Zlotnik, 1994] Theorem 2.1, Remark 2.1.

Theorem 3.2  The following error estimate holds

$$
\|y - y_\theta\|_{C(T; L^2(\Omega))} \leq c(h^2 + \tau^2)^{\frac{\alpha}{2}} \left( \|y_0\|_{H^\alpha} + \|y_1\|_{H^{\alpha-1}} + \|f\|_{W^{\alpha-1}(I, H^{\alpha})} \right)
$$

for $\alpha_1, \alpha_2 \in \{0, 1\}, \alpha_1 + \alpha_2 = \alpha - 1$ and $\alpha_1 \leq \alpha_2$, provided $(f, y_0, y_1) \in W^{\alpha-1}(I, H^{\alpha}) \times H^{\alpha} \times H^{\alpha-1}$.

Proof. The result follows directly from [Zlotnik, 1994] Theorem 4.1, 4.3.

Corollary 3.1  The following error estimate holds

$$
\|y - y_\theta\|_{L^2(\Omega_T)} \leq c(\tau^2 + h^2) \left( \|y_0\|_{H^3} + \|y_1\|_{H^2} + \|f\|_{L^1(I; H^2)} \right).
$$

Proof. This is shown by using that

$$
\|y - y_\theta\|_{L^2(\Omega_T)} \leq \|y - i_\tau y\|_{L^2(\Omega_T)} + \|i_\tau y - y_\theta\|_{L^2(\Omega_T)}
$$

where $i_\tau : C(T; L^2(\Omega)) \to S_\tau \otimes L^2(\Omega)$ is the nodal interpolant. According to [Zlotnik, 1994] Theorem 4.1] we have

$$
\|i_\tau y - y_\theta\|_{L^2(\Omega_T)} \leq c\|i_\tau y - y_\theta\|_{C(T; L^2(\Omega))} \\
\leq c\|y - y_\theta\|_{C(T; L^2(\Omega))} \leq c(\tau^2 + h^2) \left( \|y_0\|_{H^3} + \|y_1\|_{H^2} + \|f\|_{L^1(I; H^2)} \right).
$$

Moreover, we have according to Theorem 2.2

$$
\|y - i_\tau y\|_{L^2(\Omega_T)} \leq c\tau^2 \|y\|_{H^2(I; L^2(\Omega))} \leq c\tau^2 \left( \|y_0\|_{H^3} + \|y_1\|_{H^2} + \|f\|_{L^2(I; H^2)} \right).
$$

This proves the assertion.

4. Equivalent Problem ($\hat{P}$)

In this section we introduce a specific isomorphism between $BV(I)^m \otimes \{(g_j)_{j=1}^m\}$ and $M(I)^m \times \mathbb{R}^m$. Based on this isomorphism ($P$) is equivalently formulated as a measure valued control problem. First of all we prove existence and uniqueness of a solution to ($P$).

Theorem 4.1  Problem ($P$) has a unique solution in $BV(I)^m$.

Proof. Utilizing the fact, that the forward mapping is continuous from $L^2(I)^m$ to $L^2(\Omega_T)$, the proof can be carried out along the line of [Casas et al., 2017] Theorem 3.1].
Next we introduce several linear and continuous operators and discuss its properties. The operator $B: M(I)^m \times \mathbb{R}^m \to L^2(\Omega_T)$ is given by

$$
(v, c) \mapsto \sum_{j=1}^m \left( \int_0^T dv_j(s) - \frac{1}{T} \int_0^T \int_0^t dv_j(s) \, ds + c_j \right) g_j.
$$

(4.1)

The measures $v_j$ are the derivatives of the generated BV-function and $c_j$ are the mean values. Next, we define the predual operator of $B$ given by $B^*: L^2(\Omega_T) \to C_0(I)^m \times \mathbb{R}^m$

$$
B^*: q \mapsto \left( w_1', \ldots, w_m', \int_0^T \int_\Omega qg_1 \, dx \, dt, \ldots, \int_0^T \int_\Omega qg_m \, dx \, dt \right)
$$

where $w \in H^2(I)$ solves

$$
\begin{align*}
-\omega'' &= \int_\Omega q(\cdot, x)g_j(x) \, dx - \frac{1}{T} \int_\Omega q(t, x)g_j(x) \, dx \, dt \quad \text{in } (0, T), \\
\omega'_j(0) &= \omega'_j(T) = 0 \quad \text{with } \int_0^T w_j(t) \, dt = 0 \quad \text{for } j = 1, \ldots, m.
\end{align*}
$$

(4.2)

**Proposition 4.2** The operator $B^*: L^2(\Omega_T) \to C_0(I)^m \times \mathbb{R}^m$ is well defined and the predual of $B$, i.e. the following holds

$$
\int_{\Omega_T} B(v, c)q \, dx \, dt = \langle (v, c), B^*(q) \rangle
$$

for all $(v, c) \in M(I)^m \times \mathbb{R}^m$ and for all $q \in L^2(\Omega_T)$.

**Proof.** The equation (4.2) has a unique solution $w_j \in H^2(I)$, since $\int_\Omega q(\cdot, x)g_j \, dx - \frac{1}{T} \int_\Omega q(t, x)g_j \, dx \, dt \in L^2(I)$ and has zero mean. Moreover, we have $w_j' \in H^0_1(I) \hookrightarrow C_0(I)$. Thus, the operator $B^*$ is well defined. Moreover, there holds

$$
\langle (v, c), B^*(q) \rangle = \sum_{j=1}^m \int_0^T w'_j \, dv_j + \sum_{j=1}^m c_j \int_0^T \int_\Omega qg_j \, dx \, dt
$$

$$
= \sum_{j=1}^m \int_0^T -w'_j \, dv_j + \sum_{j=1}^m c_j \int_0^T \int_\Omega qg_j \, dx \, dt
$$

$$
= \sum_{j=1}^m \int_0^T \left( \int_\Omega qg_j \, dx - \frac{1}{T} \int_\Omega \int_0^T qg_j \, dx \, dt \right) \int_0^T dv_j + \sum_{j=1}^m c_j \int_0^T \int_\Omega qg_j \, dx \, dt
$$

$$
= \int_0^T \int_\Omega q \sum_{j=1}^m \left( \int_0^T dv_j - \frac{1}{T} \int_\Omega \int_0^T dv_j \, dt + c_j \right) g_j \, dx \, dt = \int_{\Omega_T} B(v, c)q \, dx \, dt
$$

for all $(v, c) \in M(I)^m \times \mathbb{R}^m$ and for all $q \in L^2(\Omega_T)$. The use of integration by parts is justified by the density of $C_c^\infty(I)$ in $H^0_1(I)$.

**Proposition 4.3** Let $w_j \in H^2(I)$, $j = 1, \ldots, m$ be the solution of (4.2). Then there holds

$$
w'_j(t) = \int_t^T \int_\Omega q(s, x)g_j(x) \, dx \, ds + \frac{(t-T)}{T} \int_0^T \int_\Omega q(t, x)g_j(x) \, dx \, dt.
$$

**Proposition 4.4** The operator $B: M(I)^m \times \mathbb{R}^m \to BV(I)^m \otimes \{g_j\}_{j=1}^m$ is an isomorphism.
Proof. The inverse of $B$ is given by

$$B^{-1}: \sum_{j=1}^{m} u_j g_j \mapsto \left( u'_1, \ldots, u'_m, \frac{1}{T} \int_0^T u_1 \, dt, \ldots, \frac{1}{T} \int_0^T u_m \, dt \right).$$

Next we introduce the operator $D: M(I)^m \times \mathbb{R}^m \to M(I)^m$ defined by $(v, c) \mapsto v$. Its predual operator is given by $D^*: C_0(I)^m \to C_0(I)^m \times \mathbb{R}^m$ with $D^*: h \mapsto (h, 0)$. Finally, let us introduce $P_1: M(I)^m \to M(I)$ defined by $v \mapsto v_1$. The predual operator has the form $P_1^*: C_0(I) \to C_0(I)^m$ with $P_1^*: h \mapsto (0, \ldots, h, \ldots, 0)$. Using $B$ we can rewrite $(P)$ into the equivalent problem

$$\begin{aligned}
\min_{v \in M(I)^m} \left\{ S(v, c) - y_d \right\}_{L^2(\Omega_T)} + \sum_{j=1}^{m} \alpha_j \|v_j\|_{M(I)} =: f(v, c),
\end{aligned}$$

with $S: M(I)^m \times \mathbb{R}^m \to L^2(\Omega_T)$ defined by $(v, c) \mapsto L(B(v, c)) + Q(y_0, y_1)$.

### 4.1 First-Order optimality condition of $(P)$

In the following a necessary and sufficient first-order optimality condition of $(P)$ is presented as well as sparsity results for the derivative of the optimal control. Let $(\overline{v}, \overline{c})$ be the unique optimal pair. We define the quantities $\overline{p} = L^* (S(\overline{v}, \overline{c}) - y_d)$ and $\overline{p}_1 \in C(\overline{T})^m$ by

$$\begin{aligned}
\overline{p}_{1,i} := - \int_{T} \int_{\Omega} \overline{p}_i \, dx \, ds
\end{aligned}$$

for $i = 1, \ldots, m$.

**Theorem 4.5** The pair $(\overline{v}, \overline{c}) \in M(I)^m \times \mathbb{R}^m$ is an optimal control of $(P)$ if and only if

$$\begin{aligned}
\overline{p}_{1,i} &\in \alpha_i \partial \|v_i\|_{M(I)}, & i = 1, \ldots, m, \\
\overline{p}_1(0) &= 0.
\end{aligned}$$

Equivalently it holds

$$\begin{aligned}
\langle v - \overline{v}_i, \overline{p}_{1,i} \rangle_{M(I)}, &\leq \alpha_i \|v_i\|_{M(I)} \quad \forall v \in M(I) \quad \text{and} \quad i = 1, \ldots, m.
\end{aligned}$$

and $\overline{p}_1(0) = 0$.

**Proof.** The proof of Theorem 4.5 is done along the lines of the proof of [Casas et al., 2017, Theorem 3.3]. By the convexity of $(P)$ we have, that $(\overline{v}, \overline{c}) \in M(I)^m \times \mathbb{R}^m$ is an optimal control of $(P)$ if and only if

$$\begin{aligned}
0 &\in \partial \left( \frac{1}{2} \|S(\overline{v}, \overline{c}) - y_d\|_{L^2(\Omega_T)}^2 + \sum_{j=1}^{m} \alpha_j \|v_j\|_{M(I)} \right) \subseteq (M(I)^m \times \mathbb{R}^m)^*.
\end{aligned}$$

Define the following function $F(v, c) := \frac{1}{2} \|S(v, c) - y_d\|_{L^2(\Omega_T)}^2$ for $(v, c) \in M(0, T)^m \times \mathbb{R}^m$. Its Gateaux derivative has the form

$$\begin{aligned}
DF(v, c)(v, c) &= B^* L^* (S(v, c) - y_d) \in C_0(I)^m \times \mathbb{R}^m
\end{aligned}$$
According to the theory of convex analysis, e.g. [Ekeland and Témam, 1999], we have
\[ 0 \in DF(v,c)(\varphi,\tilde{v}) + \partial \left( \sum_{i=1}^{m} \alpha_i \|v_i\|_{M(I)} \right) \subseteq (M(I)^m \times \mathbb{R}^m)^*. \]  
(4.6)

Using
\[ \partial \left( \sum_{i=1}^{m} \alpha_i \|P_i \tilde{v}(\varphi,\tilde{v})\|_{M(I)} \right) = \sum_{i=1}^{m} \alpha_i \partial^* P_i^* \partial \|v_i\|_{M(I)} = \left( \alpha_i \partial \|v_i\|_{M(I)} \right)_{i=1}^{m}, \]
and (4.6) as well as Proposition 4.3 imply
\[ \bar{y}_{1,i} \in \alpha_i \partial \|v_i\|_{M(I)} \quad \forall i = 1, \ldots, m, \quad \bar{y}_1(0) = 0. \]  
(4.7)

The following proposition is a consequence of [Casas and Kunisch, 2014, Proposition 3.2.]:

**Proposition 4.6** Let \((\varphi, \tilde{v}) \in M(I)^m \times \mathbb{R}^m\) be an optimal control of \((\bar{P})\), then for all \(i = 1, \ldots, m\) and \(\bar{y}_{1,i} = (\bar{y}_{1,i})_i \subseteq \mathbb{R}^m\) given in (4.3) holds
\begin{enumerate}
  \item \(\|\bar{y}_{1,i}\|_{C_0(I)} \leq \alpha_i\),
  \item \(\int_t^T \frac{\bar{y}_{1,i}}{\alpha_i} dt = \int_t^T d[\bar{v}_i] = \|v_i\|_{M(I)}\),
  \item supp(\(\bar{v}^+_i\)) \subseteq \{t \in I | \bar{y}_{1,i}(t) = \pm \alpha_i\}, \quad \text{where} \quad \bar{v}_i = \bar{v}^+_i - \bar{v}^-_i \quad \text{is the Jordan decomposition of} \quad \bar{v}_i.
\end{enumerate}

**Remark 4.1** Let us note that the boundary property of \(\bar{y}_1\), i.e. \(\bar{y}_1(0) = \bar{y}_1(T) = 0\), and the continuity of \(\bar{y}_1\) imply with Proposition 4.6(c), that there exists a \(\varepsilon > 0\) such that \(\text{dist}(\text{supp}(\bar{v}^+_i), \{0, T\}) > \varepsilon_i\).

5. The Variationally Discretized Problem

In this section we introduce a discretized version of \((\bar{P})\) and discuss its properties. We use the concept of variational discretization in which the control is not discretized. In particular, we consider the problem \((\bar{P}_{\text{semi}})\):

\[
(\bar{P}_{\text{semi}}) \begin{cases}
\min_{\nu \in M(I)^m} \frac{1}{2} \|S_{\omega}(\nu, c) - y_d\|_{L^2(\Omega_T)}^2 + \sum_{j=1}^{m} \alpha_j \|v_j\|_{M(I)} =: J_\omega(\nu, c)
\end{cases}
\]

with \(S_{\omega} : M(I)^m \times \mathbb{R}^m \rightarrow L^2(\Omega_T)\) defined by \((\nu, c) \mapsto L_{\omega}(B(\nu, c)) + Q_{\omega}(y_0, y_1)\). Here \(L_{\omega} : L^2(\Omega_T) \rightarrow L^2(\Omega_T)\) is defined by \(f \mapsto y_{\omega}(f)\), where \(y_{\omega}(f)\) solves (3.2) for a source \(f\) and \((y_0, y_1) = (0, 0)\). The operator \(Q_{\omega} : H^1_0(\Omega) \times L^2(\Omega) \rightarrow L^2(\Omega_T)\) is defined by \((y_0, y_1) \mapsto y_{\omega}(y_0, y_1)\), where \(y_{\omega}(y_0, y_1)\) solves (3.1) with \((y_0, y_1)\) as initial datum and \(f = 0\).

**Remark 5.1** We can represent the adjoint of \(L_{\omega}\) in the form \(w \mapsto L^*_{\omega}(w)(t, x) = L_{\omega}(w \circ \tilde{\omega})(\tilde{\phi}(t, x))\) with \(\tilde{\omega}(t, x) = (T - t, x)\), and \(w \in L^2(\Omega_T)\). This is true since \(L_{\omega}(f)(0) = 0\) and \(L^*_{\omega}(w)(T) = 0\) and thus \(L_{\omega}(f)\) and \(L^*_{\omega}(w)\) can be used in (3.1) as test functions for the forwards and backwards equation. Hence, Theorem 3.3, Corollary 3.1 and Lemma 3.2 are valid for \(L^*_{\omega}(w)\) as well.

**Theorem 5.1** The problem \((\bar{P}_{\text{semi}})\) has a solution in \(M(I)^m \times \mathbb{R}^m\).

**Proof.** The existence of an optimal control for \((\bar{P}_{\text{semi}})\) can be similarly shown as in the proof of Theorem 4.1. \(\square\)
Note that a BV-representation of the solutions $(\overline{\nu}, \overline{c})$, $(\overline{\nu}_\varphi, \overline{c}_\varphi)$ of $(\overline{P})$, respectively $(\tilde{\overline{P}}^\text{semi})$ are defined by

$$\overline{u}(t) := \int_0^t d\overline{\nu}(s) - \frac{1}{T} \int_0^T d\overline{\nu}(s) \, dt + \overline{c}, \quad \text{and} \quad \overline{u}_\varphi(t) := \int_0^t d\overline{\nu}_\varphi(s) - \frac{1}{T} \int_0^T d\overline{\nu}_\varphi(s) \, dt + \overline{c}_\varphi. \quad (5.1)$$

Next we define the quantities $\overline{p}_\varphi = L^*_\varphi(S_\varphi(\varphi, \varphi) - y_d)$ and

$$\overline{p}_{1, \varphi, j} := - \int_t^T \int_\Omega \overline{p}_\varphi g_j \, dx \, ds \quad \text{for} \quad j = 1, \ldots, m,$$

which is continuously differentiable and piecewise quadratic in time.

**Theorem 5.2** The pair $(\overline{\nu}_\varphi, \overline{c}_\varphi) \in M(I)^m \times \mathbb{R}^m$ is an optimal control of $(\tilde{\overline{P}}^\text{semi})$ if and only if

$$\overline{p}_{1, \varphi, j} \in \alpha_i \partial \|\overline{\nu}_{\varphi, i}\|_{M(I)} \quad i = 1, \ldots, m, \quad (5.2)$$

$$\overline{p}_{1, \varphi}(0) = 0. \quad (5.3)$$

Equivalently it holds

$$\langle \hat{\nu} - \overline{\nu}_\varphi, \overline{p}_{1, \varphi, j} \rangle_{M(I), C_0(I)} + \alpha_i \|\overline{\nu}_{\varphi, i}\|_{M(I)} \leq \alpha_i \|\hat{\nu}\|_{M(I)} \quad \forall \hat{\nu} \in M(I), \quad i = 1, \ldots, m, \quad (5.4)$$

and $\overline{p}_{1, \varphi}(0) = 0$.

**Proof.** The proof is similar to Theorem 4.5

**Remark 5.2** Due to Theorem 5.2, we can show that Proposition 4.6 holds analogously for $(\tilde{\overline{P}}^\text{semi})$.

### 6. A Priori Error Estimates

In this section error estimates of problem $(\tilde{\overline{P}}^\text{semi})$ for the optimal control, optimal state and optimal cost functional value are presented. Under specific assumptions, we proof optimal rates for the optimal control, state and cost. For reason of convenience, the following notation is introduced. For an optimal control $(\overline{\nu}_\varphi, \overline{c}_\varphi) \in M(I)^m \times \mathbb{R}^m$ of $(\tilde{\overline{P}}^\text{semi})$ and the optimal control $(\nu, \varphi) \in M(I)^m \times \mathbb{R}^m$ of $(\overline{P})$ we introduce the corresponding optimal states by $\overline{y}_\varphi := S_\varphi(\overline{\nu}_\varphi, \overline{c}_\varphi)$ and $\overline{y} := S(\nu, \varphi)$. Further, we define the mixed state by $\overline{y}_\varphi := L_\varphi(B(\varphi, \varphi)) + Q_\varphi(y_0, y_1)$. The mixed adjoint state is chosen as $\tilde{L} := L_\varphi(\overline{y} - y_d)$.

In the proofs of following the Lemmata and Theorem, we use similar steps as in the proof of [Pieper and Vexler 2013 Theorem 4.4].

**Lemma 6.1** There holds

$$\langle \overline{p}_{1, \varphi} - \overline{p}_{1, \varphi, \varphi} - \nu \rangle = 0 \quad (6.1)$$

with $(\nu, \varphi)$ as the optimal control of $(\overline{P})$ and $(\overline{\nu}_\varphi, \overline{c}_\varphi)$ as an solution of $(\tilde{\overline{P}}^\text{semi})$.

**Proof.** Inequality (6.1) follows from monotonicity of the subdifferential.

**Lemma 6.2** Consider optimal control $(\nu, \varphi)$ of $(\overline{P})$, and $(\overline{\nu}_\varphi, \overline{c}_\varphi)$ of $(\tilde{\overline{P}}^\text{semi})$, as well as their BV-representations $\overline{\nu}$, and $\overline{p}_\varphi$. For the optimal states $\overline{\nu}$ and $\overline{\nu}_\varphi$ of problem $(\overline{P})$, respectively $(\tilde{\overline{P}}^\text{semi})$, we have

$$\|\overline{\nu}_\varphi - \overline{\nu}\|_{L^2(\Omega_T)} \leq C \|\overline{\nu}_\varphi - \overline{\nu}\|_{L^2(\Omega_T)} + C \|\overline{\nu}_\varphi - \overline{\nu}\|_{L^1(I;L^2(\Omega))} \quad (6.2)$$

with a constant $C > 0$ depending on $g$. 

Proof. Lemma 6.1, the properties of $B$ and $B^*$ and the fact that $\overline{p}_1(0) = \overline{p}_{1,0}(0) = 0$ imply the following

$$0 \geq \langle p_{1,\theta} - p_1, v_\theta - \overline{v} \rangle = \langle B^*(\overline{p}_\theta - \overline{p}), (\overline{v}_\theta - v, \xi_\theta - \xi) \rangle = \langle \overline{p}_\theta - \overline{p}_0, (\xi_\theta - \xi) \cdot g \rangle_{L^2(\Omega_T)} + \langle \hat{\overline{p}}_\theta - \overline{p}, (\bar{\overline{u}}_\theta - \overline{u}) \cdot g \rangle_{L^2(\Omega_T)}$$

$$= \langle \overline{v}_\theta - \overline{v}, L_\theta((\bar{\overline{u}}_\theta - \bar{\overline{u}})g) \rangle_{L^2(\Omega_T)} + (\hat{\overline{p}}_\theta - \overline{p}, (\bar{\overline{u}}_\theta - \bar{\overline{u}}) \cdot g \rangle_{L^2(\Omega_T)}$$

$$= \langle \xi_\theta - \xi, \xi_\theta - \xi \rangle_{L^2(\Omega_T)} + \langle \overline{v}_\theta - \overline{v}, \overline{v}_\theta - \overline{v} \rangle_{L^2(\Omega_T)} + (\hat{\overline{p}}_\theta - \overline{p}, (\bar{\overline{u}}_\theta - \bar{\overline{u}}) \cdot g \rangle_{L^2(\Omega_T)}.$$  

From these calculations we obtained (6.2) by

$$\|\theta - \eta\|_{L^2(\Omega_T)} \leq \|\overline{v}_\theta - \overline{v}\|_{L^2(\Omega_T)} + (\overline{p} - \overline{p}_0, (\bar{\overline{u}}_\theta - \bar{\overline{u}}) \cdot g \rangle_{L^2(\Omega_T)}$$

$$\leq \frac{1}{2} \|\overline{v}_\theta - \overline{v}\|_{L^2(\Omega_T)} + \frac{1}{2} \|\overline{v}_\theta - \overline{v}\|_{L^2(\Omega_T)} + c \|\bar{\overline{u}}_\theta - \bar{\overline{u}}\|_{L^2(\Omega_T)} + \|\bar{\overline{u}}_\theta - \bar{\overline{u}}\|_{L^2(\Omega_T)}.$$  

\[ \square \]

**Lemma 6.3** The sequence of the BV representatives $(\bar{\overline{u}}_\theta)_{\theta}$ of the optimal controls of $(\bar{\overline{p}}^\text{gen})_{\theta}$ are bounded in $BV(I)^m$ with respect to $\theta \to 0$.

Proof. At first, we show that

$$\bar{\overline{u}}_\theta = \int_0^t d\bar{\overline{v}}_\theta(s) - \frac{1}{T} \int_0^T d\bar{\overline{v}}_\theta(s) ds + \tau_\theta = \hat{\bar{\overline{u}}}_\theta + \tau_\theta$$

is bounded in $BV(I)^m$ for $\theta \to 0$. Due to the optimality of $\bar{\overline{u}}_\theta$, holds the inequality $J_\theta(\bar{\overline{u}}_\theta) \leq J_\theta(0)$ for all considered $\theta$. Define $\gamma_\theta := S_\theta(0,0)$ and $\gamma = S(0,0)$. Using Lemma 3.2, we have

$$\|\gamma_\theta\|_{C([t,T];L^2(\Omega))} \leq c \left(\|\gamma_0\|_{B_0^0(\Omega)} + \|\gamma_1\|_{L^2(\Omega)}\right).$$

Thus, the discrete states $\gamma_\theta$ are bounded in $L^2(\Omega_T)$. Hence $\{J_\theta(0)\}_{\theta}$ is bounded in $R$. This implies that $J_\theta(\bar{\overline{u}}_\theta)$ is bounded and thus, $(\gamma_\theta)_{\theta}$ and $(\tau_\theta)_{\theta}$ are bounded in $L^2(\Omega_T)$, and $\tau_\theta$ respectively. Now it suffices to show that $\tau_\theta \in \mathbb{R}^m$ is bounded in order to get the boundedness of $(\bar{\overline{u}}_\theta)_{\theta}$ in $BV(I)^m$. Assume that $\tau_\theta \in \mathbb{R}^m$ is unbounded for $\theta \to 0$. It holds

$$\sum_{j=1}^m \alpha_j \|D_i \hat{\bar{\overline{u}}}_\theta, j\|_{L^1(I)} = \sum_{j=1}^m \alpha_j \|D_i \bar{\overline{u}}_{\theta, j}\|_{L^1(I)} \leq J_\theta(\bar{\overline{u}}_\theta) \leq J_\theta(0)$$

and with the Poincare inequality for BV(I) functions ([Ambrosio et al. 2000 p. 152]), we get that $(\hat{\bar{\overline{u}}}_\theta)_{\theta}$ is bounded in $BV(I)^m$. Consider $z_\theta = \bar{\bar{v}}_\theta - \overline{v}_\theta$ with $\overline{v}_\theta = L_\theta(\hat{\bar{\overline{u}}}_\theta \cdot g + Q_\theta(\gamma_0, \gamma_1))$. The $BV$ boundedness of $(\hat{\bar{\overline{u}}}_\theta)_{\theta}$, and therefore the boundedness in $L^2(I)^m$, implies by Lemma 3.2 that $(\bar{\overline{v}}_\theta)_{\theta}$ is bounded in $L^2(\Omega_T)$. The boundedness of $(\bar{\overline{v}}_\theta)_{\theta}$ and $(\gamma_\theta)_{\theta}$ lead to the boundedness of $(z_\theta)_{\theta}$. The linearity of $L_\theta(B(\cdot, \cdot))$, implies $z_\theta = L_\theta(B(0, \tau_\theta))$. Consider now $p_\theta := \max_{1 \leq j \leq m} |\tau_\theta, j|$, with $\tau_\theta \to \infty$, $\xi_\theta := \frac{1}{p_\theta} z_\theta$, and $\bar{\overline{u}}_\theta = \frac{\bar{\overline{r}}_\theta}{p_\theta}$. There exists a $\theta_0 > 0$ such that for all $\theta < \theta_0$ the sequence $\bar{\overline{u}}_\theta$ is bounded by definition in $\mathbb{R}^m$. Thus, let us now consider $\theta \leq \theta_0$ such as in this case all proofs hold. Hence, there exists a subsequence of $\bar{\overline{u}}_\theta$, which converges to some $\bar{\overline{u}}$. Denote this converging subsequence again by $\bar{\overline{u}}_\theta$. The linear structure of $L_\theta(B(\cdot, \cdot))$ gives us $\bar{\overline{v}}_\theta = L_\theta(B(0, \bar{\overline{u}}_\theta))$. Define by $\xi_\theta$ the solution $L_\theta(B(0, \bar{\overline{u}}))$. Next we show that Lemma 3.2 leads to $||\xi_\theta - \bar{\overline{v}}_\theta||_{L^2(\Omega_T)} \to 0$. Thus, we have

$$||\xi_\theta - \bar{\overline{v}}_\theta||_{L^2(\Omega_T)} = ||L_\theta(B(0, \bar{\overline{u}} - \bar{\overline{v}}))||_{L^2(\Omega_T)} \leq c |\bar{\overline{u}} - \bar{\overline{v}}|_{\mathbb{R}^m} \to 0.$$
Define $\xi$ by $L(B(0, \bar{a}))$. Then we have
\[ \|\xi - \xi_0\|_{C(T; L^2(\Omega))} = \|L(B(0, \bar{a}) - L_0(B(0, \bar{a}))\|_{C(T; L^2(\Omega))} \to 0 \]
according to Theorem 3.2. With the boundedness of $\bar{z}_0$ in $L^2(\Omega_T)$, the unboundedness of $|p_0|$, and the definition of $\xi_0 = \frac{m_0}{p_0}$, we can deduce that $\xi_0 \to 0$ in $L^2(\Omega_T)$. Hence, $\xi = \xi_0 + (\xi - \xi_0) \to 0$ in $L^2(\Omega_T)$. Thus, we obtain that $\xi = 0$, which implies $\sum_{j=1}^m \pi_j = 0$. Because $g_j \in L^m(\Omega) \setminus \{0\}$ have pointwise disjoint supports, we get that $\bar{a}_j = 0$, which is a contradiction. Thus, it is shown that $\bar{e}_\theta$ is bounded, and hence $(\bar{e}_\theta)_\theta$ is bounded in $BV(I)^m$. □

The next theorem states an a priori error estimate for the optimal state. Under additional assumptions on the structure of the optimal adjoint state an improved rate for the optimal state is proven. Furthermore, an optimal convergence for the control in the $L^1(I)$-norm is proven.

**Theorem 6.1** For $y_d \in C^1(T; L^2(\Omega))$ and $(y_0, y_1) \in H_0^1(\Omega) \times L^2(\Omega)$, the following a priori error estimate holds
\[ \|y - y_0\|_{L^2(\Omega_T)} \leq c(h^{2/3} + \tau^{2/3}) \left(\|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|y_d\|_{C^1(T; L^2(\Omega))}\right). \] (6.3)

For $y_d \in C^1(T; H_0^1(\Omega))$, $g \in (H^2)^m$, and $(y_0, y_1) \in H^3 \times H^2$, the following error rate holds
\[ \|y - y_0\|_{L^2(\Omega_T)} \leq c(h + \tau) \left(\|y_0\|_{H^3} + \|y_1\|_{H^2} + \|y_d\|_{C^1(T; H_0^1(\Omega))}\right). \] (6.4)

**Proof.** First we consider $\|p - \hat{p}_\theta\|_{L^1(T; L^2(\Omega_T))}$ in (6.2). By the regularity of $B(\bar{y}, \bar{\zeta}) \in L^2(\Omega_T)$, we get that $\bar{y} - y_d \in C^1(T; L^2(\Omega))$ from Theorem 2.2. The error estimate in (3.3) with $\alpha_1 = 0$ and $\alpha_2 = 1$ implies then
\[ \|p - \hat{p}_\theta\|_{L^1(T; L^2(\Omega_T))} \leq c(h^2 + \tau^2)^{1/3} \|y - y_d\|_{W^{1,1}(T; L^2(\Omega))}. \] (6.5)

Consider now the term $\|\bar{y} - \hat{y}_\theta\|_{L^2(\Omega_T)}$ in (6.2). Using (3.3) with $\alpha_1 = 0$ and $\alpha_2 = 1$ implies
\[ \|\bar{y} - \hat{y}_\theta\|_{L^2(\Omega_T)} \leq c(h^2 + \tau^2) \left(\|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|B(\bar{y}, \bar{\zeta})\|_{L^1(T; L^2(\Omega))}\right). \] (6.6)

Hence, Lemma 6.3, (6.2), (6.5) and (6.6) imply (6.3). Assume that $y_d \in C^1(T; H_0^1(\Omega))$, $g \in (H^2)^m$ and $(y_0, y_1) \in H^3 \times H^2$ hold. By Theorem 2.2 we have $\bar{y} - y_d \in C^1(T; H_0^1(\Omega))$. Using (3.3) with $\alpha_1 = 1$ and $\alpha_2 = 1$, implies
\[ \|p - \hat{p}_\theta\|_{L^1(T; L^2(\Omega_T))} \leq c(h^2 + \tau^2)^{1/2} \|y - y_d\|_{W^{1,1}(T; H_0^1(\Omega))}. \] (6.7)

Corollary 3.1 leads to $\|\bar{y} - \hat{y}_\theta\|_{L^2(\Omega_T)} = O(\tau^2 + h^2)$. Hence, (6.7) and the convergence rate of $\|\bar{y} - \hat{y}_\theta\|_{L^2(\Omega_T)}$ give us (6.4). □

**Lemma 6.4** Let $p$ be the weak solution of $\mathcal{W}^*$ for $w \in C^1(T; H^2(\Omega))$ with $\gamma = 0$, and $p_\theta$ its discrete counterpart. Then there holds
\[ \left\|R_h \int_0^T p \, dt - \int_0^T p_\theta \, dt\right\|_{H_0^1(\Omega)} \leq c(h^2 + \tau^2)^{1/3} \|w\|_{C^1(T; H^2(\Omega))}. \]
Proof. This follows directly from [Zlotnik, 1994, Theorem 4.1].

To avoid repetitions and increase readability, we introduce the following constant κ, which is used to specify the convergence rates depending on the regularity of \((y_d, y_0, y_1, g)\). To address the convergence rate later, the following data dependent constant is defined

$$\kappa := \begin{cases} 
\frac{2}{3}, & \text{if } y_d \in C^1(\mathcal{I}, L^2(\Omega)), \quad (y_0, y_1) \in H_0^1(\Omega) \times L^2(\Omega), \quad g \in L^m(\Omega), \\
2, & \text{if } y_d \in C^1(\mathcal{I}, H_0^1(\Omega)), \quad (y_0, y_1) \in H^3 \times H^2, \quad g \in (H^2)^m.
\end{cases}$$

**Theorem 6.2** For the optimal control \((v, \tau)\) of \((\bar{P})\) and solutions \((v_\theta, \tau_\theta)\) of \((\bar{P}_\theta^\text{semi})\) the following a priori error estimates hold:

$$\|J(v, \tau) - J_\theta(v_\theta, \tau_\theta)\| = \begin{cases} 
O(\tau + h^2), & \text{if } \kappa = 2/3, \\
O(\tau^2 + h^3), & \text{if } \kappa = 2,
\end{cases} \quad (6.8)$$

$$\|\|v\|_{M(\mathcal{I})^m} - \|v_\theta\|_{M(\mathcal{I})^m}\| = \begin{cases} 
O(\tau^{2/3} + h^{2/3}), & \text{if } \kappa = 2/3, \\
O(\tau + h), & \text{if } \kappa = 2.
\end{cases} \quad (6.9)$$

The following proof is a modified version of the proof of [Pieper and Vexler, 2013, Theorem 4.2.].

Proof. Optimality leads to the following two inequalities

$$J(v, \tau) \leq J(v_\theta, \tau_\theta) \quad \text{and} \quad J_\theta(v_\theta, \tau_\theta) \leq J_\theta(v, \tau).$$

This implies \(J(v, \tau) - J_\theta(v, \tau) \leq J(v_\theta, \tau) - J_\theta(v_\theta, \tau_\theta) \leq J(v_\theta, \tau_\theta) - J_\theta(v_\theta, \tau_\theta).\) So it remains to estimate the error with respect to the cost functionals for a fixed \((v, c)\), i.e. \((v, \tau)\) and \((v_\theta, \tau_\theta)\). From the cost functionals of \((\bar{P})\) and \((\bar{P}_\theta^\text{semi})\) follows the following identity

$$J(v, c) - J_\theta(v, c) = -\frac{1}{2} \|S(v, c) - S_\theta(v, c)\|^2_{L^2(\Omega_\tau)} + \langle B(v, c), L^*(S(v, c) - y_d) - L_\theta^*(S(v, c) - y_d) \rangle_{L^2(\Omega_\tau)}$$

$$+ \langle Q(y_0, y_1) - Q_\theta(y_0, y_1), S(v, c) - y_d \rangle_{L^2(\Omega_\tau)} \quad (6.10)$$

Lemmas 2.2 and 3.1 imply

$$\langle Q(y_0, y_1) - Q_\theta(y_0, y_1), S(v, c) - y_d \rangle_{L^2(\Omega_\tau)}$$

$$= \langle y_1, p(0) - \rho_\theta(0) \rangle_{L^2(\Omega)} + \left( y_0 - R_h y_0, \int_0^T S(v, c) - y_d \, dt \right)_{L^2(\Omega)}$$

$$+ \langle \nabla (R_h y_0) - y_0, \nabla \int_0^T p \, dt \rangle_{L^2(\Omega)} + \langle \nabla R_h y_0, \nabla \int_0^T p_\theta - p \, dt \rangle_{L^2(\Omega)} \quad (6.11)$$

We set \(\alpha = 1\) for \(\kappa = 2/3\) and \(\alpha = 2\) for \(\kappa = 2\). Then we have according to Theorems 2.2 and 3.2

$$\|y_1\|_{L^2(\Omega)} \|p - \rho_\theta \|_{C(\mathcal{I}, L^2(\Omega))} \leq c(\tau^\alpha + h^\alpha) \|y_1\|_{L^2(\Omega)} \|S(v, c) - y_d\|_{C(\mathcal{I}, L^2(\Omega))}.$$
Further we have
\[
\left( y_0 - R_h y_0, \int_0^T S(v, c) - y_d \, dt \right)_{L^2(\Omega)} \leq \| y_0 - R_h y_0 \|_{L^2(\Omega)} \| S(v, c) - y_d \|_{L^2(\Omega)} \\
\leq c h^\alpha \| y_0 \|_{H^\alpha} \| S(v, c) - y_d \|_{L^2(\Omega)}.
\]

Since \( S(v, c) - y_d \in C^1(\bar{T}; L^2(\Omega)) \) we have \( p \in C(\bar{T}; \mathbb{H}^2) \) according to Theorem 2.2 and thus
\[
\left( \nabla (R_h y_0 - y_0), \nabla \int_0^T p \, dt \right)_{L^2(\Omega)} = \left( R_h y_0 - y_0, -\Delta \int_0^T p \, dt \right)_{L^2(\Omega)} \\
\leq c \| y_0 - R_h y_0 \|_{L^2(\Omega)} \| p \|_{L^2(\bar{T}; \mathbb{H}^2)} \\
\leq c h^\alpha \| y_0 \|_{H^\alpha} \| S(v, c) - y_d \|_{C^1(\bar{T}; L^2(\Omega))}.
\]

Finally, we have according to Lemma 6.4
\[
\left( \nabla R_h y_0, \nabla \int_0^T p_{\varphi} - p \, dt \right)_{L^2(\Omega)} = \left( \nabla R_h y_0, \nabla \left( \int_0^T p_{\varphi} - R_h \int_0^T p \, dt \right) \right)_{L^2(\Omega)} \\
\leq c \| y_0 \|_{L^2(\Omega)} \| \int_0^T p_{\varphi} - R_h \int_0^T p \, dt \|_{H^1(\bar{T}; \mathbb{H}^2)} \leq c \| y_0 \|_{L^2(\Omega)} \| (\varphi^\alpha + h^\alpha) \|_{C^1(\bar{T}; \mathbb{H}^\alpha)} \| S(v, c) - y_d \|_{C^1(\bar{T}; L^2(\Omega))}.
\]

For \((B(v, c), y_0, y_1, y_d) \in L^2(\bar{T}; \Omega) \times H^1(\bar{T}; \mathbb{H}^2) \times L^2(\bar{T}; \mathbb{H}^2) \times C^1(\bar{T}; L^2(\Omega)), \) the function \( S(v, c) - y_d \) is an element of \( C^1(\bar{T}; L^2(\Omega)) \) according to Theorem 2.2. If \((g_1, y_0, y_1, y_d) \in (\mathbb{H}^2)^m \times \mathbb{H}^2 \times \mathbb{H}^2 \times C^1(\bar{T}; H^1(\Omega)) \) we get that \( S(v, c) - y_d \in C^1(\bar{T}; H^1(\Omega)). \) Hence, the a priori estimate (3.3) implies
\[
|B(v, c, L^\star(S(v, c) - y_d) - L^\star_0(S(v, c) - y_d))|_{L^2(\Omega_T)} = O(\varphi^\alpha + h^\alpha).
\]

Inequality (3.3) and Corollary 3.1 imply \( \frac{1}{2} \| S(v, c) - S_0(v, c) \|_{L^2(\Omega_T)} = O(\varphi^\alpha + h^\alpha). \) Thus, the assertion is proven.

### 6.1 Optimal Convergence Rates for the Optimal Controls of \((\bar{P}_{\varphi\alpha})_{\alpha})\)

Under certain assumptions we show that the BV-representations \( \bar{\pi}_{\varphi\alpha} \) of the optimal controls of \((\bar{P}_{\varphi\alpha})_{\alpha})\) converge with a specific rate in the \(L^1\) norm to the optimal control \( \pi \) of \((P)\). Further, define the following functions:
\[
z(t) := \varphi_{11}(P(t)) = \int_{\Omega} \left( S(\bar{v}, \bar{\pi}) - y_d \right) (t) g \, dx \quad \text{ (6.13)}
\]
\[
z_{\varphi\alpha}(t) := \varphi_{11}(P(t)) = \int_{\Omega} \left( S_{\varphi\alpha}(\pi_{\varphi\alpha}, \pi_{\varphi\alpha}) - y_d \right) (t) g \, dx \quad \text{ (6.14)}
\]

with \((\bar{v}, \bar{\pi})\) as the optimal control of \((\bar{P})\) and \((\pi_{\varphi\alpha}, \pi_{\varphi\alpha})\) as optimal control of \((\bar{P}_{\varphi\alpha})_{\alpha})\). Due to Proposition 4.6 and Remark 5.2 it holds that \( \text{supp}(\bar{v}_i) \subseteq \{ t \mid z_i(t) = 0 \} \) and \( \text{supp}(\pi_{\varphi\alpha, i}) \subseteq \{ t \mid z_{\varphi\alpha, i}(t) = 0 \} \).

**Lemma 6.5** The matrix \( G := (L_{g_{i,j}}(g_{i,j})_{L^2(\Omega_T)})_{i,j=1}^m \in \mathbb{R}^{m \times m} \) is symmetric and positive definite.

**Proof.** The matrix \( G \) is a Gramian-matrix, which is a consequence of the uniqueness of solutions of the wave equation the fact that \( \{ g_{i,j} \}_{i=1}^m \) is a linear independent system.
Theorem 6.3 \( \overline{u}_\vartheta \) converges weakly* in \( BV(0, T)^m \) to the solution \( \overline{u} \) for \( \vartheta \to 0 \).

**Proof.** Let \( (\vartheta_m, h_m) \) be a null sequence such that \( (\vartheta_m)_{m=1}^\infty \subset \mathbb{R}^+ \), \( (h_m)_{m=1}^\infty \subset \mathbb{R}^+ \). Let \( (\vartheta_m, T_m) \) be a bounded sequence in \( BV(I)^m \) where \( \{v_\vartheta, \vartheta_\vartheta\} \) are optimal controls of \( (\vartheta_m, h_m) \). The weak* compactness of closed and bounded sets in \( BV(I)^m \) implies the existence of a subsequence \( \{v_\vartheta, \vartheta_\vartheta\}_k \) which converges weakly* to some \( \tilde{u} \in BV(I)^m \). Hence, \( \{v_\vartheta, \vartheta_\vartheta\}_k \) converges in \( L^2(I)^m \) to \( \tilde{u} \) and \( D_t v_\vartheta \) converges weakly* in \( I^m \) to \( D_t \tilde{u} \). There exists a unique element \( (\tilde{v}, \tilde{c}) \in M(I)^m \times \mathbb{R}^m \) such that \( \tilde{u} = \int_0^T d\tilde{v}(s) - \frac{1}{2} \int_0^T \int_0^t d\tilde{v}(s) \, dt + \tilde{c} \). Due to the weak* l.s.c. of \( \| \cdot \|_{M(I)} \) in \( M(I) \), we get

\[
\liminf_{k \to \infty} \sum_{i=1}^m \alpha_i \|v_{\vartheta,i} - \tilde{u}_i\|_{M(I)} \geq \sum_{i=1}^m \alpha_i \|D_t \tilde{u}_i\|_{M(I)}. \tag{6.15}
\]

Let us show that

\[
\lim_{k \to \infty} \|S_{\vartheta_k}(v_{\vartheta_k}, \vartheta_{\vartheta_k}) - y_d\|_{L^2(\varOmega_T)}^2 = \|S(\tilde{v}, \tilde{c}) - y_d\|_{L^2(\varOmega_T)}^2 \tag{6.16}
\]

holds. Theorem 3.2 the stability of \( L_{\vartheta_k} \), see Lemma 3.2 and the strong convergence of \( \vartheta_{\vartheta_k} \) in \( L^2(I) \) lead to

\[
\|S_{\vartheta_k}(v_{\vartheta_k}, \vartheta_{\vartheta_k}) - S(\tilde{v}, \tilde{c})\|_{L^2(\varOmega_T)} \leq \|S_{\vartheta_k}(v_{\vartheta_k}, \vartheta_{\vartheta_k}) - S_{\vartheta_k}(\tilde{v}, \tilde{c})\|_{L^2(\varOmega_T)} + \|S_{\vartheta_k}(\tilde{v}, \tilde{c}) - S(\tilde{v}, \tilde{c})\|_{L^2(\varOmega_T)} \leq c\|\vartheta_{\vartheta_k} - \tilde{u}\|_{L^2(I)^m} + c(h_{\vartheta_k}^2 + \vartheta_{\vartheta_k}^2)^{1/2}(\|y_0\|_{H^1(\varOmega)} + \|y_1\|_{L^2(\varOmega)} + \|B(\tilde{v}, \tilde{c})\|_{L^1(\varOmega)}). \tag{6.17}
\]

This leads to (6.16). With (6.15), (6.16) and Theorem 6.2 we get

\[
J(\vartheta, \vartheta) = \liminf_{k \to \infty} J_{\vartheta_k}(v_{\vartheta_k}, \vartheta_{\vartheta_k}) \geq J(\tilde{v}, \tilde{c}).
\]

The uniqueness of the optimal control of \( (P) \) leads to the desired result. \( \square \)

**Corollary 6.1** There holds \( \vartheta_{\vartheta} \to \vartheta \) in \( L^2(I) \) for \( \vartheta \to 0 \).

Next we prove pointwise convergence of \( z_{\vartheta} \) and \( \vartheta_{\vartheta} \).

**Lemma 6.6** For \( \vartheta \to 0 \) we have \( \|z_{\vartheta} - z\|_{L^\infty(I)^m} \to 0 \).

**Proof.** By Theorem 2.2 and Definition 3.1, we have that \( z_{\vartheta} \in C(\bar{T}) \) and \( z \in C^1(\bar{T}) \). Hence, \( \|z_{\vartheta} - z\|_{L^\infty(I)^m} \) is well-defined. There holds that

\[
\|z_{\vartheta} - z\|_{L^\infty(I)^m} \leq c \sup_{t \in \bar{T}} \|L^*_\vartheta(S_\vartheta(\vartheta_\vartheta, \vartheta_\vartheta) - y_d)(t, \cdot) - L^*(S(\vartheta, \vartheta_\vartheta) - y_d)(t, \cdot)\|_{L^1(\varOmega)} \leq c \left[ \|L^*_\vartheta(S_\vartheta(\vartheta_\vartheta, \vartheta_\vartheta)) - L^*(S(\vartheta, \vartheta_\vartheta))\|_{C(\bar{T}, L^2(\varOmega))} + \|L^*_\vartheta(y_d) - L^*(y_d)\|_{C(\bar{T}, L^2(\varOmega))} \right]. \tag{6.18}
\]

Next we show that \( \|z_{\vartheta} - z\|_{L^\infty(I)^m} \to 0 \) holds. According to Theorem 3.2 we obtain

\[
\|L^*_\vartheta(y_d) - L^*(y_d)\|_{C(\bar{T}, L^2(\varOmega))} \leq c(h^2 + \vartheta_{\vartheta}^2)^{1/2}\|y_d\|_{L^1(\varOmega)} \tag{6.19}
\]

Furthermore, we have

\[
\|L^*_\vartheta(S_\vartheta(\vartheta_\vartheta, \vartheta_\vartheta)) - L^*(S(\vartheta, \vartheta_\vartheta))\|_{C(\bar{T}, L^2(\varOmega))} \leq \|L^*_\vartheta(S_\vartheta(\vartheta_\vartheta, \vartheta_\vartheta) - S(\vartheta, \vartheta_\vartheta))\|_{C(\bar{T}, L^2(\varOmega))} + \|L^*_\vartheta(S(\vartheta, \vartheta_\vartheta) - S(\vartheta, \vartheta_\vartheta))\|_{C(\bar{T}, L^2(\varOmega))}. \tag{6.20}
\]
For the second term on the right hand side of (6.19), we have using Theorem 3.2

$$
\|L_0^s(S(\bar{v}, \bar{c})) - L^s(\bar{v}, \bar{c})\|_{C(\tilde{\Omega}^2)_{L^1(\Omega)}} \lesssim c(h^2 + \tau^2)^{\frac{1}{2}} \|S(\bar{v}, \bar{c})\|_{L^1(\tilde{\Omega}^2(\Omega))}.
$$

(6.20)

For the first term on the right hand side of (6.19), we use the stability of $L_0$ and $L_0^s$ from Lemma 3.2 to obtain

$$
\|L_0^s(S(\bar{v}, \bar{c})) - L_0^s(\bar{v}, \bar{c})\|_{C(\tilde{\Omega}^2)_{L^1(\Omega)}} \lesssim c \|S_0(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{C(\tilde{\Omega}^2(\Omega))} \\
\lesssim c \left( \|S_0(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{C(\tilde{\Omega}^2(\Omega))} + \|S(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{C(\tilde{\Omega}^2(\Omega))} \right) \\
\lesssim c \|\bar{v}_\theta - \bar{v}\|_{L^2(I)} + \|S_0(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{C(\tilde{\Omega}^2(\Omega))}.
$$

(6.21)

The strong convergence of $\bar{u}_\theta$ to $\bar{u}$ in $L^2(I)$, see Corollary 6.1 and Theorem 3.2 imply that $\|L_0^s(S(\bar{v}_\theta, \bar{c}_\theta)) - L_0^s(S(\bar{v}, \bar{c}))\|_{C(\tilde{\Omega}^2(\Omega))}$ converges to 0 for $\theta \to 0$. Hence, $\|\bar{v}_\theta - \bar{v}\|_{L^2(I)}$ converges to 0 for $\theta \to 0$. □

**Lemma 6.7** There exists a constant $c > 0$ such that the following inequality holds for all $f \in L^1(I; L^2(\Omega))$

$$
\|\bar{d}_t L_0^s(f)\|_{C(\tilde{\Omega}^2)_{L^1(\Omega)}} \lesssim c \|f\|_{L^1(I; L^2(\Omega))}.
$$

(6.22)

**Proof.** This follows directly from [Zlotnik 1994, Theorem 2.1]. □

**Lemma 6.8** The following a priori error estimate

$$
\|\bar{d}_t(L^s(f) - L_0^s(f))\|_{C(\tilde{\Omega}^2)_{L^1(\Omega)}} \lesssim (h^2 + \tau^2)^{\frac{1}{2}} \|f\|_{W^{1,1}(I; L^2(\Omega))}
$$

holds for all $f \in W^{1,1}(I; L^2(\Omega))$.

**Proof.** This follows directly from [Zlotnik 1994, Theorem 4.2]. □

**Lemma 6.9** We have

$$
\|\bar{d}_t(z_{\theta} - z)\|_{C(\tilde{\Omega}^2)_{L^1(\Omega)}} \to 0 \quad \text{for} \quad \theta \to 0.
$$

(6.23)

**Proof.** Lemma 6.7 implies

$$
\|\bar{d}_t(z_{\theta} - z)\|_{C(\tilde{\Omega}^2)_{L^1(\Omega)}} = \sum_{l=1}^m \sup_{i \in I^2} \left| \bar{d}_t \int_\Omega \left( L_0^s(S(\bar{v}_\theta, \bar{c}_\theta) - y_d)(t_i) - L^s(S(\bar{v}, \bar{c}) - y_d)(t_i) \right) g_{l - i} \, dx \right|
$$

$$
\lesssim c \|\bar{d}_t(L_0^s(S(\bar{v}_\theta, \bar{c}_\theta) - y_d) - L^s(S(\bar{v}, \bar{c}) - y_d))\|_{C(\tilde{\Omega}^2(\Omega))} \\
\lesssim c \|\bar{d}_t(L_0^s(S(\bar{v}_\theta, \bar{c}_\theta) - y_d) - L^s(S(\bar{v}, \bar{c}) - y_d))\|_{C(\tilde{\Omega}^2(\Omega))} \\
+ c \|\bar{d}_t(L_0^s(S(\bar{v}, \bar{c}) - y_d) - L^s(S(\bar{v}, \bar{c}) - y_d))\|_{C(\tilde{\Omega}^2(\Omega))} \\
\lesssim c \|S_0(\bar{v}_\theta, \bar{c}_\theta) - S(\bar{v}, \bar{c})\|_{L^1(I; L^2(\Omega))} \\
+ c \|\bar{d}_t(L_0^s(S(\bar{v}, \bar{c}) - y_d) - L^s(S(\bar{v}, \bar{c}) - y_d))\|_{C(\tilde{\Omega}^2(\Omega))}
$$

(6.24)

The first term on the right hand side of the last inequality converges to 0 for $\theta \to 0$, e.g. see (6.21). Because $y_d$ and $S(\bar{v}, \bar{c}) \in C^1(I; L^2(\Omega))$ holds Lemma 6.8 implies that the last term in the last inequality converges to 0 for $\theta \to 0$. This proves the assertion. □

**Lemma 6.10** We have

$$
\|\bar{d}_t(z_{\theta} - z)\|_{L^2(I; L^1(\Omega))} \to 0 \quad \text{for} \quad \theta \to 0.
$$

(6.25)
Proof. At first we define a cell-wise discretization of the derivative of \( z \) as follows
\[
\overline{\delta} z := \sum_{i=1}^{M} \frac{z(t_i) - z(t_{i-1})}{\tau} 1_{I_i}, \quad \text{with } I_i := (t_{i-1}, t_i), \quad i = 1, \ldots, M.
\] (6.26)
Then we proceed with
\[
\| \partial_t (z_\theta - \overline{\delta} z) \|_{L^\infty(I)\cap P_{\text{NE}}} \leq \| \partial_t z_\theta - \overline{\delta} z \|_{L^\infty(I)\cap P_{\text{NE}}} + \| \overline{\delta} z - \partial_t \overline{\delta} z \|_{L^\infty(I)\cap P_{\text{NE}}}.
\]
Using the disjoint supports of the characteristic functions in the definition of \( \overline{\delta} z \) leads to
\[
\| \partial_t z_\theta - \overline{\delta} z \|_{L^\infty(I)\cap P_{\text{NE}}} = \sum_{j=1}^{m} \max_{i=1,\ldots,M} \left| \frac{z'(t_j) - z'(t_{j-1})}{\tau} - \frac{z_j(t_j) - z_j(t_{j-1})}{\tau} \right| = \sum_{j=1}^{m} \left| \frac{z_j(t_j) - z_j(t_{j-1})}{\tau} - \partial_t z_j(t_j) \right| + \max_{i=1,\ldots,M} \sup_{t_j \in I_i} \left| \partial_t z_j(t_j) - \partial_t z_j(t_j) \right|.
\]
In the last equation, we directly see that the first term converges to 0 due to [Anastassiou, 2017, Theorem 1.11]. The second term converges to 0 due to the uniform continuity of \( \partial_t z_j(t) \) in \( I \). Hence, the result follows for \( \theta \to 0 \), which implies the claim.

\[ \square \]

**Lemma 6.11** The convergence \( \| p_{1, \theta} - p_1 \|_{L^\infty(I)\cap \mathcal{W}_{\text{NE}}} \to 0 \) holds for \( \theta \to 0 \).

**Proof.** This follows directly from
\[
\| p_{1, \theta} - p_1 \|_{L^\infty(I)\cap \mathcal{W}_{\text{NE}}} \leq c \| L_\delta'(S_\theta(\bfv_\theta, \r_{\bfv_\theta}) - y_d) - L_\delta'(S(\bfv, \r) - y_d) \|_{C(I, L^2(\Omega))},
\]
which converges to 0 using the same steps as in Lemma [6.6].

In order to proof a priori error estimates for the control in the \( L^1(I) \)-norm and higher convergence rates for the state variable we have to make the following assumption.

**Assumption** (A1) \( \{ j \in I \mid |p_{1,j}(t)| = \alpha \} = \{ t_{1,j}, \ldots, t_{m_j,i} \} \) for \( m_j \in \mathbb{N} \), with \( i = 1, \ldots, m_j \).

(A2) \( \partial_t z_{i,j}(t, \theta) \neq 0 \), for \( j = 1, \ldots, m_j \) and \( i = 1, \ldots, m \).

**Remark 6.1** The assumption (A1) enforces finitely many jumps for the optimal control of \( (P) \), i.e. \( \text{supp}(D_t \bfv_\theta) \subseteq \{ t_{1,j}, \ldots, t_{m,j} \} \) for \( \bfv \in BV(I)^m \).

**Lemma 6.12** Let \( (\bfv_\theta, r_{\bfv_\theta}) \) be an optimal control of \( (P_\text{opti}) \). Under the assumptions (A1) and (A2) above, there exists a \( \delta > 0 \), and \( \theta_0 > 0 \) such that for all \( 0 < \theta \leq \theta_0 \) holds
\[
\bfv_\theta = \sum_{j=1}^{m} c_{i,j,\theta} \delta_{i,j,\theta} \quad \text{with } c_{i,j,\theta} \in \mathbb{R} \quad \text{and } t_{j,i} \in B_\delta(t_{j,i}),
\]
where \( B_\delta(t_{j,i}) \) are pairwise disjoint for a fixed \( i = 1, \ldots, m \) with respect to the index \( j = 1, \ldots, m \) and with \( 0 < \theta \to 0 \). The coefficients in front of the Dirac measures of \( \bfv_{i,j} \), i.e. \( c_{i,j,\theta} \) for \( l = 1, \ldots, m \), are possibly 0.
Proof. Let us begin with the case \( m = 1, m_1 = 1 \), i.e. \( \{ t \in I \mid |p_1(t)| = \alpha \} = \{ \bar{t}_1 \} \). First of all we know that \( |p_1(t)| \leq \alpha \) for all \( t \in \bar{I} \) holds and since \( p_1 \in C^1(\bar{I}) \) as well as that \( \bar{t}_1 \) is an interior point follows
\[
z(\bar{t}_1) = -\bar{\omega}_1 p_1(\bar{t}_1) = 0.\]
Moreover, due to (A2) there exists a \( \delta > 0 \) and \( c_1 > 0 \) such that \( |\bar{\omega}_1 z(t)| > c_1 \) for all \( t \in B^s(\bar{t}_1) \subset I \). Since \( \bar{\omega}_1 z \) is continuous, \( \bar{\omega}_1 z \) does not change its sign on \( B^s(\bar{t}_1) \) and hence \( z \) is strictly monotone in \( B^s(\bar{t}_1) \). Therefore \( \bar{t}_1 \) is the only root of \( z \) in \( B^s(\bar{t}_1) \). Moreover, there exist \( t_- < t_+ \in B^s(\bar{t}_1) \) with \( z(t_-) < 0 < z(t_+) \). By Lemma 6.6 there exists a \( \delta_0 = (\tau_0, h_0) \) such that \( z(\delta_0) < 0 < z(\delta_0) \) for all \( \delta \neq \delta_0 \). Since \( z_0 \) is continuous there exists a \( t_0 \in (t_-, t_+) \) such that \( z_0(t_0) = 0 \) for all \( \delta \neq \delta_0 \). Next we show that there exists a \( \delta_0 \neq \delta_0 \) such that \( \delta_0 \) is the only root of \( z_0 \) in \( B^s(\bar{t}_1) \) for all \( \delta \neq \delta_0 \). Lemma 6.10 implies existence of a \( \delta_0 < \delta_0 \) that \( \bar{\omega}_1 z_0 \) is either strictly positive or strictly negative on \( B^s(\bar{t}_1) \). Now let \( \bar{t}_0 \) be a second root of \( z_0 \) in \( B^s(\bar{t}_1) \). Then it holds
\[
0 = z_0(t_0) - z_0(\bar{t}_0) = \int_{t_0}^{\bar{t}_0} \bar{\omega}_1 z_0(t) \, dt \neq 0.
\]
Hence, there is no second root of \( z_0 \) in \( B^s(\bar{t}_1) \). Next we show that \( t \neq t_0 \) and \( z_0(t) = 0 \) imply the existence of \( \delta_0 < \delta_0 \) such that \( |p_1(t)| < \alpha \) for all \( \delta < \delta \) and thus \( \tau_0 = c_1 \delta \bar{\omega}_0 \) with \( c_1 \) possibly zero. Such a \( t \) can only exist in \( I \setminus B^s(\bar{t}_1) \). Due to Assumption 6.1 and the condition \( |p_1(\delta)| \leq \alpha \) for all \( \delta \in I \) there exists a \( \delta > 0 \) such that \( |p_1(\delta)| < \alpha - \varepsilon \) for all \( \delta \in I \setminus B^s(\bar{t}_1) \). Lemma 6.11 implies the existence of a \( \delta_0 < \delta \) with \( |p_1(\delta)| < \alpha - \varepsilon / 2 \) for all \( \delta < \delta_0 \) and \( t \in I \setminus B^s(\bar{t}_1) \). In the case of \( m = 1 \) and \( \{ t \in I \mid |p_1(t)| = \alpha \} = \{ \bar{t}_1, \bar{t}_1 \} \) with \( m_1 > 1 \), we can find for each \( \bar{t}_i \) a \( \delta_0 > 0 \) with \( \cap_{i=1}^{m_i} B^s(\bar{t}_i) = \emptyset \) and \( \bar{t}_i \) such that there exists a \( t_0^{\bar{t}_i} \in B^s(\bar{t}_i) \) with \( \tau_0 = |B^s(\bar{t}_i)| = c_1 \delta_0 \bar{\omega}_0 \) and \( \tau_0 T \cap \cup_{i=1}^{m_i} B^s(\bar{t}_i) = \emptyset \). Then choose \( \delta_0 < \min_{i=1, \ldots, m_1} \delta_0 \). In the case of \( m > 1 \), one has to consider the same proof as above with respect to an additional subindex \( i = 1, \ldots, m \) and the smallest \( \delta_0 > 0 \) used in the proofs of each component \( i = 1, \ldots, m \).

From now on, we will assume that \( \delta \leq \delta_0 \) holds with \( \delta_0 \) from Lemma 6.12. Furthermore, without loss of generality, we assume that \( \delta > 0 \) in Lemma 6.12 is considered to be small enough such that there exists a \( \delta > 0 \) for which \( \bar{\delta} < \text{dist}(B^s(\bar{t}_j i), (0, T)) \), and \( \delta < \text{dist}(B^s(\bar{t}_j i), B^s(t_j, i), j = 1, \ldots, m_i, j_1 \neq j_2 \) for \( i = 1, \ldots, m \). Let us note that Remark 4.1 and Lemma 6.12 guarantee that such a \( \delta > 0 \) exists. Under these assumptions, we can work with the following definition.

**Definition 6.4** Let us define the BV representations of the optimal controls of \( P \) and \( (P^\text{sem}_\delta) \) in a more explicit form
\[
\bar{u}_i = c_i + \sum_{j=1}^{m_i} c_j^i \left( 1_{(t_{j, i}, T)}(t) - \frac{T - t_{j, i}}{T} \right), \quad \bar{u}_i = c_i, \bar{\omega} + \sum_{j=1}^{m_i} c_j^i \left( 1_{(T_i, t_{j, i})} - \frac{T - t_{j, i}}{T} \right)
\]
for \( i = 1, \ldots, m \).

**Lemma 6.13** The following inequality holds
\[
\|\bar{u} - \bar{u}_i \|_{L^1(T)} \leq c \left( \sum_{i=1}^m \left( |c_i - c_i| + \sum_{j=1}^{m_i} |c_j^i| \cdot |t_{j, i} - t_{j, i}| + |c_j^i - c_j^i| \right) \right) \quad (6.27)
\]
for some constant \( c \) which depends only on \( T \).

We can prove (6.27) by using \( \|1_{(t_{j, i}, T)} - 1_{(T_i, t_{j, i})}\|_{L^1(T)} = |t_{j, i} - t_{j, i}| \).

**Lemma 6.14** For each \( t_{j, i} \) there is a function \( f_{j, i} \in C^1_c(\Omega_T) \), with \( j = 1, \ldots, m_i \) and \( i = 1, \ldots, m \), such that the function
\[
g^i(t, x) := \int_t^T L^*(f_{j, i})(s, x) \, ds \in C_0(I; L^2(\Omega)) \quad (6.28)
\]
fulfills the properties

a) \( L^*(f_{j,i})(t,x) = h_{i,j}(t)f_i(x) \) for some \( h_{i,j} \in C_c^\infty(I), f_i \in C_c^\infty(\Omega) \), and \( f_{j,i} = f_i \partial_n h_{i,j} - h_{i,j} \Delta f_i, \)

b) \( 0 \leq \int_0^T h_{i,j}(s) \, ds < 1 \) in \( [t_{j,i} - \frac{\delta}{2}, t_{j,i} + \frac{\delta}{2}] \cap B_\delta(t_{j,i}), \)

c) \( \int_0^T h_{i,j}(s) \, ds = 1 \) in \( B_\delta(t_{j,i}), \)

d) \( \text{supp} \left( \int_0^T h_{i,j}(s) \, ds \right) \subseteq [t_{j,i} - \frac{\delta}{2}, t_{j,i} + \frac{\delta}{2}] \subset I, \text{ i.e. } \int_\Omega g_i^j \, dx \in C_0(I), \)

e) \( \langle f_i, g_i \rangle_{L^2(\Omega)} = \delta_{i,j}, \)

with \( \delta_{i,j} = 0, \) if \( l \neq 0 \) and 1 else.

**Proof.** For all \( I_{j,i} := [t_{j,i} - \frac{\delta}{2}, t_{j,i} + \frac{\delta}{2}], \) \( j = 1, \ldots, m_i \) and \( i = 1, \ldots, m, \) there exists \( \tilde{p}_{j,i} \in C_c^\infty(I) \) such that \( 0 \leq \tilde{p}_{j,i} \leq 1 \) with

\[
\tilde{p}_{j,i} = 0 \text{ in } I \backslash I_{j,i}, \quad \tilde{p}_{j,i} = 1 \text{ in } [t_{j,i} - \frac{\delta}{2}, t_{j,i} + \frac{\delta}{2}].
\]

(6.29)

Let us define \( h_{i,j} = -\partial_n \tilde{p}_{j,i}. \) For each \( g_i, i = 1, \ldots, m, \) we can find a \( f_i \in C_c^\infty(\Omega) \) such that \( (f_i, g_k)_{L^2(\Omega)} \)

is 0 for \( i \neq k \) and 1 else. One can show that \( L^*(f_{j,i}) = h_{i,j}f_i, \) and \( f_{j,i} := f_i \partial_n h_{i,j} - h_{i,j} \Delta f_i \) holds. Hence, \( g_i^j, \) defined in (6.28), fulfills the desired properties a)-e). \( \square \)

**Lemma 6.15** There exists a constant \( c > 0 \) independent of \( \Theta \) such that

\[
|c_j^i - c_{j,i,\Theta}| \leq c \left( \tau^K + h^K + \|\nabla \tilde{r}_{\Theta}\|_{L^2(\Omega,T)} \right)
\]

with \( j = 1, \ldots, m_i \) and \( i = 1, \ldots, m. \)

(6.30)

**Proof.** Let \( i = 1, \ldots, m, \) \( j \in \{1, \ldots, m_i\} \) and consider from Lemma 6.14 the function \( g_i^j = \int_0^T h_{i,j,i}f_i \, ds = \int_0^T L^*(f_{j,i}) \, ds. \) Hence, we have

\[
c_j^i - c_{j,i,\Theta} = \int_0^T \int_0^T h_{i,j}(s) \, ds \, dF_i(t_{j,i} - \frac{\delta}{2}, t_{j,i} + \frac{\delta}{2})
\]

\[
= \sum_{l=1}^m \int_0^T h_{i,j}(t) \, dt \int_\Omega f_i \, dx = \int_0^T \int_\Omega -\partial_n g_i^j \left( \sum_{l=1}^m (\nabla - \nabla_{\Theta}) g_i \right) \, dt \, dx
\]

(6.31)

Thus, it follows

\[
c_j^i - c_{j,i,\Theta} = \int_0^T \int_\Omega -\partial_n g_i^j \left( \sum_{l=1}^m (\nabla - \nabla_{\Theta}) g_i \right) \, dt \, dx = \int_0^T \int_\Omega \left( L^*(f_{j,i}) - L^*_{\Theta}(f_{j,i}, \nabla \tilde{r}_{\Theta}) \right) \, dx
\]

\[
= \left( L^*(f_{j,i}) - L^*_{\Theta}(f_{j,i}, \nabla \tilde{r}_{\Theta}) \right)_{L^2(\Omega,T)} + \left( L^*_{\Theta}(f_{j,i}, \nabla \tilde{r}_{\Theta}) - B(\nabla \tilde{r}_{\Theta}) \right)_{L^2(\Omega,T)}.
\]

By Theorem 3.2 and the boundedness of \( (\nabla \tilde{r}_{\Theta}) \) we obtain

\[
\left( L^*(f_{j,i}) - L^*_{\Theta}(f_{j,i}, \nabla \tilde{r}_{\Theta}) \right)_{L^2(\Omega,T)} \leq \|L^*(f_{j,i}) - L^*_{\Theta}(f_{j,i})\|_{L^2(\Omega,T)} \|B(\nabla \tilde{r}_{\Theta}) - B(\nabla \tilde{r}_{\Theta})\|_{L^2(\Omega,T)} = O(\tau^2 + h^2).
\]
Moreover, there holds
\[
(L_\theta^*(f_{j,i}), B(\bar{v}, \bar{\tau}) - B(\bar{v}_\theta, \bar{\tau}_\theta))_{L^2(\Omega_T)} = (f_{j,i}, L_\theta(B(\bar{v}, \bar{\tau}) - B(\bar{v}_\theta, \bar{\tau}_\theta)))_{L^2(\Omega_T)}
\]
\[
\leq \|f_{j,i}\|_{L^2(\Omega_T)} \|S_\theta(\bar{v}, \bar{\tau}) - S_\theta(\bar{v}_\theta, \bar{\tau}_\theta)\|_{L^2(\Omega_T)}
\]
\[
\leq c \|S_\theta(\bar{v}, \bar{\tau}) - S(\bar{v}, \bar{\tau})\|_{L^2(\Omega_T)} + c \|S(\bar{v}, \bar{\tau}) - S_\theta(\bar{v}_\theta, \bar{\tau}_\theta)\|_{L^2(\Omega_T)}
\]
\[
\leq c(\tau^k + h^k + ||\bar{y} - \bar{y}_\theta||_{L^2(\Omega_T)})
\]
according to Theorem 3.2 and Corollary 3.1

**Lemma 6.16** There holds that
\[
|t_{j,i}^r - t_{j,i}^\prime| \leq c \left( \tau^k + h^k + ||\bar{y} - \bar{y}_\theta||_{L^2(\Omega_T)} \right)
\]
with \( j = 1, \ldots, m_i \) and \( i = 1, \ldots, m \). (6.32)

**Proof.** Using that \( z_i(t_{j,i}^r) = z_{i,j}(t_{j,i}^\prime) = 0 \) and \( \xi_j = \partial_i p_{1,i} \in C^1(\bar{T}) \) gives us \( z_i(t_{j,i}^r) - z_{i,j}(t_{j,i}^\prime) = z_i(t_{j,i}^r) + \partial_i z(\xi_j)(t_{j,i}^r - t_{j,i}^\prime) \) for some \( \xi_j \in B_\delta(t_{j,i}^r) \). In the proof of Lemma 6.12 we have shown that \( |\partial_i z(\xi_j)| > 0 \) for all \( \xi_j \in B_\delta(t_{j,i}^r) \) and therefore we have \( \partial_i z_i(\xi) \neq 0 \). Then Lemma 3.2 and Theorem 3.2 imply
\[
|t_{j,i}^r - t_{j,i}^\prime| \leq c \|z_i - z_{i,j}\|_{L^\infty(\bar{T})} \leq c \|z - z_0\|_{L^\infty(\bar{\Omega})}
\]
\[
\leq c \|L^*(S(\bar{v}, \bar{\tau}) - \bar{y}_d) - L^*_{\theta}(S_\theta(\bar{v}_\theta, \bar{\tau}_\theta) - \bar{y}_d)\|_{C(\bar{T};L^2(\Omega))}
\]
\[
\leq c \|L^*(S(\bar{v}, \bar{\tau}) - \bar{y}_d) - L^*_{\theta}(S_\theta(\bar{v}_\theta, \bar{\tau}_\theta) - \bar{y}_d)\|_{C(\bar{T};L^2(\Omega))} + c \|L^*_{\theta}(S_\theta(\bar{v}_\theta, \bar{\tau}_\theta) - S_\theta(\bar{v}_\theta, \bar{\tau}_\theta))\|_{C(\bar{T};L^2(\Omega))}
\]
\[
\leq c \|L^*(S(\bar{v}, \bar{\tau}) - \bar{y}_d) - L^*_{\theta}(S_\theta(\bar{v}_\theta, \bar{\tau}_\theta) - \bar{y}_d)\|_{C(\bar{T};L^2(\Omega))} + c \|S(\bar{v}, \bar{\tau}) - S_\theta(\bar{v}_\theta, \bar{\tau}_\theta)\|_{L^1(\bar{T};L^2(\Omega))}
\]
\[
\leq c(\tau^k + h^k + ||\bar{y} - \bar{y}_\theta||_{L^2(\Omega_T)}).
\]
(6.33)

**Lemma 6.17** Let \( \theta_0 \) be small enough such that
\[
\left( (L_{\theta_0}(g_i), L_{\theta_0}(g_j))_{L^2(\Omega_T)} \right)_{i,j = 1, \ldots, m} > 0 \text{ holds.}
\]

Then we obtain
\[
|c_i - c_{i,\theta}| \leq c(\tau^k + h^k + ||\bar{y} - \bar{y}_\theta||_{L^2(\Omega_T)}).
\]
(6.34)

**Proof.** First, define the function \( h(t) = 1_{(t,T)} - \frac{T-t}{T-t} \). The optimality conditions of the continuous and discrete problem lead to
\[
0 = p_{1,i}(0) = \sum_{l=1}^m c_i \int_0^T \int_\Omega L^*(L(g_l)) g_i \, dx \, dt
\]
\[
+ \sum_{l=1}^m \sum_{j=1}^{m_l} c_{i,l} \int_0^T \int_\Omega L^*(L(h(t_{j,l})g_l)) g_i \, dx \, dt + \int_0^T \int_\Omega L^*(Q(\eta_0, \eta_1) - \eta_d) g_i \, dx \, dt
\]
as well as in the discrete case to
\[
0 = p_{1,\theta,i}(0) = \sum_{l=1}^m c_{i,\theta} \int_0^T \int_\Omega L^*_{\theta}(L_\theta(g_l)) g_i \, dx \, dt
\]
\[
+ \sum_{l=1}^m \sum_{j=1}^{m_l} c_{i,\theta,l} \int_0^T \int_\Omega L^*_{\theta}(L_\theta(h(t_{j,l,\theta})g_l)) g_i \, dx \, dt + \int_0^T \int_\Omega L^*_{\theta}(Q_\theta(\eta_0, \eta_1) - \eta_d) g_i \, dx \, dt.
\]
Again, by Theorem 2.2 we have

\[ \sum_{l=1}^{m} (c_{l,0} - c_l) \int_{0}^{T} \int_{\Omega} L^* (L(g_l)) g_i \, dx \, dt = \sum_{l=1}^{m} c_{l,0} \int_{0}^{T} \int_{\Omega} [L^* (L(g_l)) - L^*_0 (L_\theta (g_l))] g_i \, dx \, dt \]

\[ + \sum_{l=1}^{m} \sum_{j=1}^{m_l} \left( c_{j} - c_{j,0} \right) \int_{0}^{T} \int_{\Omega} L^* (L (h(t_{j,l}, g_l))) g_i \, dx \, dt \]

\[ + \sum_{l=1}^{m} \sum_{j=1}^{m_l} c_{j,0} \left( \int_{0}^{T} \int_{\Omega} L^* \left( L (h(t_{j,\theta}, g_l)) \right) - L^*_0 \left( L_\theta \left( h(t_{j,\theta}, g_l) \right) \right) \right) g_i \, dx \, dt \]

\[ + \int_{0}^{T} \int_{\Omega} [L^* \left( Q(y_0, y_1) - y_0 \right) - L^*_0 \left( Q_\theta (y_0, y_1) - y_\theta \right) ] g_i \, dx \, dt. \tag{6.35} \]

Next we consider the following inequality

\[ \sum_{l=1}^{m} \sum_{j=1}^{m_l} \left( c_{j} - c_{j,0} \right) \int_{0}^{T} \int_{\Omega} L^* (L (h(t_{j,l}, g_l))) g_i \, dx \, dt \]

\[ \leq c \sum_{l=1}^{m} \sum_{j=1}^{m_l} \left| c_{j} - c_{j,0} \right| \leq c \left( \tau^\kappa + h^\kappa \right). \tag{6.36} \]

For the following we remark that \( \tilde{u}_\theta \) is bounded \( BV(I) \), see Lemma 6.3. Then we consider the first term in (6.35) on the righthand side. The regularity of \( g_l \) implies that \( L(g_l) \in C^1(\overline{I}; \mathbb{H}^\alpha) \) with \( \alpha = 0, 1 \) depending on \( \kappa \) according to Theorem 2.2. Thus, with (6.3), Lemma 3.2 and Corollary 3.1 it follows,

\[ \sum_{l=1}^{m} \sum_{j=1}^{m_l} \left( c_{j} - c_{j,0} \right) \int_{0}^{T} \int_{\Omega} L^* (L (h(t_{j,l}, g_l))) g_i \, dx \, dt \]

\[ \leq c \sum_{l=1}^{m} \left( \| L^* (L(g_l)) - L^*_0 (L_\theta (g_l)) \|_{L^2(\Omega_\tau)} \right) \]

\[ + \| L(g_l) - L_\theta (g_l) \|_{L^2(\Omega_\tau)} = O(\tau^\kappa + h^\kappa). \]

By Lemma 6.15 we obtain:

\[ \sum_{l=1}^{m} \sum_{j=1}^{m_l} \left( c_{j} - c_{j,0} \right) \int_{0}^{T} \int_{\Omega} L^* (L (h(t_{j,l}, g_l))) g_i \, dx \, dt \]

\[ \leq c \sum_{l=1}^{m} \sum_{j=1}^{m_l} \| L^* (L (h(t_{j,l}, g_l))) - L^*_0 (L_\theta (h(t_{j,\theta}, g_l))) \|_{L^1(\Omega_\tau)} \]

\[ + \| L (h(t_{j,l}, g_l)) - L_\theta (h(t_{j,\theta}, g_l)) \|_{L^1(\Omega_\tau)} \tag{6.37} \]

Again, by Theorem 2.2 we have \( L(h(\tilde{t}, g_l)) \in C^1(\overline{I}; \mathbb{H}^\alpha) \) with \( \alpha = 0, 1 \) and any \( \tilde{t} \in I \). Hence, with (6.3) we get

\[ \sum_{l=1}^{m} \sum_{j=1}^{m_l} \| L^* (L (h(t_{j,l}, g_l))) - L^*_0 (L (h(t_{j,l}, g_l))) \|_{L^1(\Omega_\tau)} = O(\tau^\kappa + h^\kappa). \]

Next we consider the following inequality

\[ \| L (h(t_{j,l}, g_l)) - L_\theta (h(t_{j,\theta}, g_l)) \|_{L^1(\Omega_\tau)} \]

\[ \leq \| L (h(t_{j,l}, g_l)) - L_\theta (h(t_{j,l}, g_l)) \|_{L^1(\Omega_\tau)} + \| L_\theta (h(t_{j,l}, g_l)) - L_\theta (h(t_{j,\theta}, g_l)) \|_{L^1(\Omega_\tau)}. \tag{6.38} \]
Due to (3.3) and Corollary 3.1, the first term on the right hand side of (6.38) possess the asymptotic rate $O(\tau^k + h^k)$. By Lemma 3.2 and Lemma 6.16 we obtain for the second term an estimate in terms of $c(\tau^k + h^k + \|y - \tilde{y}_\theta\|_{L^2(\Omega_T)})$. Finally, we consider the last term in (6.35). We have

$$
\int_0^T \int_\Omega [L^*(Q(y_0, y_1) - y_d) - L^*_\theta (Q_\theta(y_0, y_1) - y_d)] g_i \, dx \, dt \leq c\|L^*(Q(y_0, y_1) - y_d) - L^*_\theta (Q_\theta(y_0, y_1) - y_d)\|_{L^2(\Omega_T)} + c\|L^*_\theta (Q_\theta(y_0, y_1) - y_d)\|_{L^2(\Omega_T)}. \quad (6.39)
$$

The first term converges in (6.39) with a rate $(\tau^k + h^k)$ according to Theorem 3.2 since $Q(y_0, y_1) - y_d \in C^1(I, H^\alpha)$ with $\alpha = 0.1$. The prescribed regularity of $(y_0, y_1)$, Lemma 3.2 and the error estimates in (3.3) give us an estimate in terms of order $(\tau^k + h^k)$ of the last term in (6.39). Thus, we have

$$
\sum_{i=1}^m (c_i - c_i, \theta) \int_0^T \int_\Omega [L^*(L(g_i))] g_i \, dx \, dt \leq c(\tau^k + h^k + \|y - \tilde{y}_\theta\|_{L^2(\Omega_T)}). \quad (6.40)
$$

Next we recall the symmetric positive definiteness of the matrix $G$ from Lemma 6.5. It holds

$$
G(\tilde{\tau} - \tilde{\tau}_\theta) = \left(\sum_{i=1}^m (c_i - c_i, \theta) \int_0^T \int_\Omega [L^*(L(g_i))] g_i \, dx \, dt\right)^m. \quad (6.41)
$$

Furthermore, we have

$$
\|G(\tilde{\tau} - \tilde{\tau}_\theta)\|_{\mathbb{R}^m} \geq \lambda_{\min} \|\tilde{\tau} - \tilde{\tau}_\theta\|_{\mathbb{R}^m} \geq c\lambda_{\min} \|\tilde{\tau} - \tilde{\tau}_\theta\|_{\infty} \geq c\lambda_{\min} |c_i - c_i, \theta| \quad (6.42)
$$

for $i = 1, \cdots, m$ where $\lambda_{\min} > 0$ is the smallest eigenvalue of $G$. Using (6.41) and the convergence rates in (6.40) gives us (6.34).

From now on we assume that all assumptions in Lemma 6.17 hold.

**Corollary 6.2** It holds that

$$
\|\tilde{\pi} - \pi_\theta\|_{L^1(I)^m} \leq c(\tau^k + h^k + \|y - \tilde{y}_\theta\|_{L^2(\Omega_T)}).
$$

This corollary is a consequence of Lemma 6.13 6.15 6.16 6.17. Next we state the main result of this work.

**Theorem 6.5** The following convergence rates hold.

$$
\|\tilde{\pi} - \pi_\theta\|_{L^1(I)^m} = O(\tau^k + h^k), \quad |c_i - c_i, \theta| = O(\tau^k + h^k), \quad (6.42)
$$

$$
|t_{j,i} - t_{j,i,\theta}| = O(\tau^k + h^k), \quad |c'_j - c'_j, \theta| = O(\tau^k + h^k) \quad (6.43)
$$

with $j = 1, \cdots, m$, $i = 1, \cdots, m$. Furthermore, we have for the optimal states of $(\tilde{P})$ and $(\tilde{P}_\theta^{semi})$

$$
\|y - \tilde{y}_\theta\|_{L^2(\Omega_T)} = O(\tau^k + h^k). \quad (6.44)
$$

**Proof** In case of $\kappa = \frac{1}{2}$, we obtain by Theorem 6.1 that $\|y - \tilde{y}_\theta\|_{L^2(\Omega_T)} = O(\tilde{\tau}^{\frac{1}{2}} + h^{\frac{1}{2}})$. By Corollary 6.2 and Lemma 6.15 6.16 6.17 the claimed result holds for this case. Consider now the case $\kappa = 2$ and...
with this assumption that the needed regularity holds for our data. Using the inequality in (6.2), Corollary 3.1 and Corollary 6.2, we obtain for some $\varepsilon > 0$

$$
\|\mathbf{y}_\theta - \mathbf{y}\|_{L^2(\Omega_T)} \leq c\|\mathbf{y} - \mathbf{y}_\theta\|_{L^2(\Omega_T)} + c\|\pi_\theta - \pi\|_{L^1(I)^n}^\frac{1}{2} \|\mathbf{p} - \mathbf{p}_\theta\|_{C(I;L^2(\Omega))}^\frac{1}{2} 
$$

$$
\leq c(\varepsilon^k + h^k) + c\|\pi_\theta - \pi\|_{L^1(I)^n} + \frac{c}{4\varepsilon} \|\mathbf{p} - \mathbf{p}_\theta\|_{C(I;L^2(\Omega))} 
$$

$$
\leq c(\varepsilon^k + h^k) + c\varepsilon \|\pi_\theta - \mathbf{y}\|_{L^2(\Omega_T)} + \frac{c}{4\varepsilon} \|\mathbf{p} - \mathbf{p}_\theta\|_{C(I;L^2(\Omega))}. 
$$

Consider a $\varepsilon > 0$ such that $c\varepsilon = \frac{1}{2}$, then we have

$$
\|\mathbf{y}_\theta - \mathbf{y}\|_{L^2(\Omega_T)} \leq c(\varepsilon^k + h^k) + \|\mathbf{p} - \mathbf{p}_\theta\|_{C(I;L^2(\Omega))}. 
$$

Then the a priori estimate 3.3 implies

$$
\|\mathbf{p} - \mathbf{p}_\theta\|_{C(I;L^2(\Omega))} \leq c(h^2 + \varepsilon^2) \|\mathbf{y} - \mathbf{y}_\theta\|_{C(I;L^2(\Omega))}. 
$$

So we have $\|\mathbf{y}_\theta - \mathbf{y}\|_{L^2(\Omega_T)} = O(\varepsilon^2 + h^2)$ and thus the same rate for the control in the $L^1(I)$-norm. Using the optimal rates of $\|\mathbf{y}_\theta - \mathbf{y}\|_{L^2(\Omega_T)}$ in (6.30), (6.32), and (6.34) implies the optimal quadratic convergence rates for $c_{i,\theta}, t_{i,\theta}$ and $c_{i,j,\theta}$. \hfill \square

**Corollary 6.3** For the BV representations of the optimal controls of $(\hat{P})$ and $(P_{\text{semi}}^0)$ hold

$$
\|D_i\pi_\theta\|_{M(I)} - \|D_i\pi_\theta\|_{M(I)} = O(\varepsilon^k + h^k). 
$$

Furthermore, $\pi_\theta$ converges strictly in $BV(0, T)$ to $\pi$ for $\theta \rightarrow 0$ with the convergence rate $O(\varepsilon^k + h^k)$.

**Proof.** The statements are a consequence of (6.8), (6.44), and Theorem 6.5 \hfill \square

**7. Numerical Experiments**

In order to numerically verify the previously presented optimal error rates, an appropriate algorithm is of particular importance due to the variational discretization of problem $(\hat{P})$. Similarly as in [Hafemeyer et al. 2019], we introduce Algorithm 1 as modified version of the primal dual active point (PDAP) algorithm, which has its origin in [Pieper and Walter 2019, Algorithm 2]. The main difference, apart from using a time-dependent hyperbolic equation, to the algorithm in [Hafemeyer et al. 2019] is that we consider different controls. For this purpose define the map $\mathcal{W}_{\mathcal{A}}(\lambda) := \sum_{\mathcal{A} \subseteq I} \lambda_{\mathcal{A}} \partial_{i}^\mathcal{A}$ for any finite set $\mathcal{A} \subseteq I$ and $\lambda \in \mathbb{R}^{|\mathcal{A}|}$. If a stopping criterion for Algorithm 1 is of interest, one can define an optimal certificate $\Psi(w_k, c_k)$. The optimal certificate is of the form

$$
\Psi(v, c) = \langle B^*L^0_\theta(S_\theta(v, c) - y_d), (v, c) \rangle + \sum_{i=1}^m \alpha_i \|v\|_{M(0,T)} - \min_{(w,q)} \langle B^*L^0_\theta(S_\theta(v, c) - y_d), (w, q) \rangle + \sum_{i=1}^m \alpha_i \|w_i\|_{M(0,T)}. \quad (7.1) 
$$

A straightforward proof can show that $\Psi(\hat{v}, \hat{c}) = 0$ implies (5.2) with $(\hat{v}, \hat{c})$ as optimal control of $(P_{\text{semi}}^0)$. Furthermore, we have $J_\theta(w_k, c_k) - J_\theta(\pi_\theta, c_\theta) \leq \Psi(w_k, c_k)$, where the proof is similar to [Walter 2016]
Algorithm 1: BV-PDAP Algorithm

Input: For \( i = 1, \ldots, m \) define \( A_{0,i} \subset I \) with \( |A_{0,i}| < \infty, \lambda_{0,i} \in \mathbb{R}^{\mathbb{R}} \).
\[
\begin{aligned}
\quad w_0 := (w_0)_{i=1}^m := \left( \mathcal{W}_{A_{0,i}}(\lambda_{0,i}) \right)_{i=1}^m \in M(0,T)^m, c_0 \in \mathbb{R}^m, \text{ and } k = 0:
\end{aligned}
\]

Calculate:
1. \( \hat{t}_i = \arg \max_{t \in I} \left| \int_t^T \int_\Omega L_{\theta}^k (S_{\theta}(w_k, c_k) - y_d) g_i \, dx \, ds \right| \)
   for \( i = 1, \ldots, m \).
2. Set \( A_{k,i} := \text{supp}(w_{k,i}) \cup \{ \hat{t}_i \} \), for \( i = 1, \ldots, m \), \( A_k = \bigcup_{i=1}^m A_{k,i} \) and compute
\[
(\overline{\lambda}, \overline{\tau}) = \arg \min_{\lambda \in \mathbb{R}^{A_{k,i}}, c \in \mathbb{R}^m} \frac{1}{2} \left\| S_{\theta} \left( \left( \mathcal{W}_{A_{k,i}}(\lambda_i) \right)_{i=1}^m, c \right) - y_d \right\|_{L^2(\Omega_T)}^2 + \sum_{i=1}^m \alpha_i \left\| \mathcal{W}_{A_{k,i}}(\lambda_i) \right\|_{M(0,T)}.
\]
3. Define \( w_{k+1} = \left( \mathcal{W}_{A_{k,i}}(\lambda_{k+1}) \right)_{i=1}^m, c_{k+1} = \overline{\tau}; \text{ set } k = k + 1 \) and return to 1.

Lemma 12]. In the numerical experiment, that is considered below, we replace \( J_\theta(\overline{\tau}, \overline{\tau}_\theta) \) with \( J(\overline{\tau}, \overline{\tau}) \) and use this for the abort criterion for the while-loop. In this experiment we will see that \( J(\overline{\tau}, \overline{\tau}) \) can be calculated due to an constructed example that is using specific data for \( (P) \) which lead to an optimal control that can be expressed explicitly and thus imply calculable costs \( J(\overline{\tau}, \overline{\tau}) \), see (7.4).

Since we want experimentally verify the optimality of our error rates, let us consider specific configurations of our data with respect to \( (P) \), such that an explicit solution is a piecewise constant \( BV \)-function with finitely many jumps. This example will fulfill all assumptions we need for the optimal error rates. We use a construction procedure that can be found in [Engel and Kunisch] [2018]. Let us fix the following data for \( d = 1, 2, 3 \) and \( m = 1 \):

\[
\Omega := (-1,1)^d, \quad I := (0,T), \quad g(x) := \prod_{i=1}^d \cos \left( \frac{\pi}{2} x_i \right) \in C^0_0(\Omega), \quad \text{and}
\]

\[
(y_0, y_1) = (0, 0), \quad \overline{\varphi}(t, x) := \beta \sin \left( \frac{2 \pi t}{T} \right) \sin \left( \frac{\ell \pi}{T} \right) \prod_{i=1}^d \cos \left( \frac{\pi}{2} x_i \right)
\]

with \( \beta = \alpha^2 \frac{3m}{2T} \). Define the following control in \( BV(0,T) \) by

\[
\overline{u} := \int_0^t dD_t \overline{u} - \frac{1}{T} \int_0^T \int_{[0,x]} dD_t \overline{u} ds + \overline{\tau} \text{ with } \overline{\tau} \in \mathbb{R}.
\]

Note that

\[
D_t \overline{u} = \sum_{i=0}^{\ell-1} \text{sign} \left( \sin \left( \frac{2i+1}{2} \right) \right) \frac{\delta_{y_d}^{(1,2)}}{2^\ell} \tag{7.2}
\]

\[
\text{with } \frac{1}{T} \int_0^T \int_{[0,x]} dD_t \overline{u} ds = \sum_{i=0}^{\ell-1} \text{sign} \left( \sin \left( \frac{2i+1}{2} \right) \right) \left( \frac{2(\ell - i) - 1}{2^\ell} \right) \tag{7.3}
\]

holds. With this we can now define the following desired state as \( y_d = S(\overline{\tau}, \overline{\tau}) - (\overline{\partial}_t - \triangle) \overline{\varphi} \). Finally, all necessary data is set for \( (P) \). It can be shown that the optimal solution, with respect to the fixed data
for \((P)\), is \(\pi\) where \(p_1(t)\) is of the form \(-\int_0^T \int_\Omega \varphi(x,s) \cdot g(x) \, dx \, ds = \beta \frac{T^2}{3\tau^3} \left( \sin \left( \frac{\pi t}{\tau} \right) \right)^3\). By several calculation steps can be shown

\[
\hat{J}(D_t \bar{u}, \bar{\tau}) = \frac{T}{8} (\beta^3 a_1^2 + a_2^2) + \alpha t, \quad \text{with } a_1 := \frac{d\pi^2}{4} - \frac{5\ell^2 \pi^2}{T^2}, \quad a_2 := \frac{4\ell^2 \pi^2}{T^2} \beta. \tag{7.4}
\]

**Complete numerical discretization:** In the following let us discuss the most important steps for the complete discretization of Algorithm 1 related to the example above. For the full discretization it should be ensured that the new jump position \(\hat{t}_i\), for \(i = 1, \ldots, m\), considered in step 1, can lie somewhere between two time nodes in \(T\) and thus the support of the new iterated \(w_{k+1}\) has not to be restricted or projected to the time nodes \(\bar{w}^\tau\). If we ignore this detail, the jump position is depending on the grid discretization, which can imply linear rates instead. Due to the finite support of each iterate \(w_k\), we can write \(B(w_k, c_k)\) as follows:

\[
B(w_k, c_k) = \left( \sum_{\ell \in \text{supp}(w_k)} \bar{\tau}_i \hat{1}_{[\tau]}(t) - \bar{\tau}_i (1 - \frac{t}{\tau}) + c_k \right) g(x) =: f_k(t) g(x). \tag{7.5}
\]

\(B(w_k, c_k)\) acts as forcing function in \(L_\theta\) and thus has to be projected on \(S_\tau \otimes S_\theta\) for the discretized wave solution we introduced before. Considering the full discretization scheme for the discrete wave solution in [Zlotnik, 1994], p.165, we need to project \(g\) on \(S_\theta\) which involves approximations of the space integral by using a Gaussian quadrature rule for example. In particular, we used a Gaussian quadrature of order 3. The projection formula for the time function \(f_k(t)\) involves in [Zlotnik, 1994, p.165, p.167] the integration of hat-functions with \(f_k(t)\). This can be done without further discretization of the involved time integrals. In particular, to obtaining \(S_\theta(w_k, c_k)\) in turns out, that we only need an additional discretization step for the projection of \(g\) on the finite element space \(S_\theta\).

Another difficulty is the discretization of the desired state \(\chi_d = S(\varphi, \bar{T}) - (\partial_t - \Delta) \varphi\). At first, let us approximate \(S(\varphi, \bar{T})\) by \(S_\theta(\varphi, \bar{T})\), which we denote by \(S^{(c,v)}_\theta\), where \(B(\varphi, \bar{T})\) is discretized as we discussed before. For \((\partial_t - \Delta) \varphi\) we are able to obtain an explicit formula which we projected on \(S_\tau \otimes S_\theta\) by using Gaussian quadrature of order 5 for all time integrals and order 3 for all spatial integrals in the projection steps of [Zlotnik, 1994, p.165, p.167]. Let us define the discretized function \((\partial_t - \Delta) \varphi\) by \(\varphi^{(c,v)}_\theta\). Hence, we can use directly the full discretization scheme in [Zlotnik, 1994, p.165] for \(L^*_{\partial} S^{(c,v)}_\theta - \varphi^{(c,v)}_\theta\), where \((\chi_0, \chi_1)\) is projected on \(S_\theta\) as zero function. Using the projected \(g\) function on \(S_\theta\), defined by \(g^{(c)}\), and the space-time function \(L^*_{\partial} S_\theta(w_k, c_k) - \chi_d\), we can explicitly calculated \(\int_0^T L^*_{\partial} S_\theta(w_k, c_k) - \chi_d g^{(c)} \, dx\), which is a element of \(S_\tau\) with \(S_\theta(w_k, c_k) - \chi_d\) discretized as above. The last time integral in step 1 of Algorithm 1 can now be calculated explicitly, for \(t \in \bar{T}\), by using the following formula:

\[
\int_0^T \sum_{\ell=0}^M v_\ell e_\ell(t) \, ds = \sum_{\ell=1}^{t_{\ell}(t)} \left( \frac{v_\ell + v_{\ell-1}}{2} \right) + v_{t_{\ell}(t)} \frac{t_{\ell}(t) - t^2 - 2(t_{\ell(t)} + t_{\ell(t)} - t_{\ell(t)})}{2\tau} + \frac{t^2 - 2t_{\ell(t)} + 2(t^2 - t_{\ell(t)})}{2\tau}.
\]

where \(v_\ell \in \mathbb{R}\) and \(\sum_{\ell=0}^M v_\ell e_\ell(t)\) is an arbitrary element of \(S_\tau\). If \(t > 0\), \(t_{\ell}(t)\) is returning the highest index of all strictly smaller time nodes in \(\bar{w}^\tau\) else it returns the index 0. Thus, we discretized \(\int_0^T L^*_{\partial} S_\theta(w_k, c_k) - \chi_d g \, dx \, ds\) into a \(C^1(0, T)\) function, which we define by \(f_d(t)\).

Step 1 is then executed as follows: First we determine the time derivative of \(f_d\). Since the derivative is an element of \(S_\tau\), we check where coefficients of the hat functions change their sign. At these
node positions we can assume a local extremum and use it to determine a candidate for $\hat{t}$. The exact position of the extremum can be calculated as follows: The time derivative of $f_\vartheta$ has the form $\frac{\partial t}{\partial t} f_\vartheta(t) = \sum_{t=0}^{M} f_{\ell} c_{\ell}(t)$ for some $f_{\ell} \in \mathbb{R}$. Let us observe a sign change in $i \in \{1, \ldots, M\}$. Hence, $\frac{\partial t}{\partial t} f_\vartheta(t)$ has a root in $\tilde{t} = \frac{f_{i+1} - f_{i-1}^{-1}}{f_{i} - f_{i-1}^{-1}}$, where $f_i \neq f_{i-1}$ due to the sign change assumption. If there is no sign change of all coefficients observed, then one can set a point $\tilde{t}$ from $I$ randomly. Since there are only finitely many coefficients to choose from, the corresponding number of extrema is finite. Step 1 is now finished, if we take the candidate, which maximizes $|f_\vartheta|$. In our numerical implementation, we consider all possible extrema of $f_\vartheta$. In practice, we have seen that adding several new jump points makes the algorithm produce meaningful results faster.

In case of step 2 of Algorithm 1 we used a similar prox operator approach as described in [Engel and Kunisch 2018]. In particular, for fixed $\gamma > 0$ we consider the additional cost term $\gamma/2 \| (\mathcal{A}_l)_{l \in A_k} \|_{\mathcal{A}_k}^2$ in the minimization problem of step 2 and use a prox-operator approach to obtain a equation that is equivalent to the optimality condition of the problem in step 2. This equation is solved by a semi-smooth Newton method. For more details consider [Engel and Kunisch 2018], where the finite dimensionality of the problem in step 2 has to be taken into account. It is assumed that for a small $\gamma$ we obtain a solution that approximates well the solution of the unregularized problem in step 2. In our numerical experiment below, we will use for example $\gamma = 10^{-10}$.

**Experimental results:** In our numerical experiment, we consider the following fixed data in the example introduced before:

$$d = 2, \quad I = (0, 2), \quad \ell = 2, \quad \alpha = 2.3 \cdot 10^{-4}, \quad \tau = 0.$$  

(7.6)

The spatial mesh is discretized quasi-uniformly. In Figure 1 we present the numerically verified errors.
that are presented in Theorem [6,5] as well as the cost rates \( J(v, \bar{c}) - J_{\partial}(\bar{v}, \bar{c}) \) which are defined as "J-Cost Error". In particular we define in Figure [1] the "State Error" by \( \| y - \bar{y} \|_{L^2(\Omega_t)} \), the "Amplitude Error" by \( |c_i^j - c_{i,j}^\partial| \), the "Constant Error" by \( |c_i - c_{i}^\partial| \), and the "Control L1-Error" by \( \| u - \bar{u} \|_{L^1(T)} \), for \( j = 1,2 \) with respect to (7.2) and (7.6). We used (7.4) for \( J(v, c) \) and considered \((v_{\partial}, \bar{c}_{\partial})\) as solution of the full discretized BV-PDAP algorithm. Furthermore, \( J_{\partial}(v, c) \) is calculated by all approximation steps used for \( S_0([\bar{v}, c], y_d) \), where the least squares part is calculated under the consideration of the mass matrix in space and time, where we used Gaussian quadrature of order 3 for the approximation of the mass matrix in space. On the x-axis in Figure [1] we see the discretization level considered for \( \partial \), i.e. \( \tau = 2^{-k} \) and \( h = 2 \cdot \sqrt{2} \cdot 2^{-k} \) with \( k = 3, 4, \ldots, 7 \). For each discretization level \( \partial \) we initialized the BV-PDAP algorithm with \((w_0, c_0) = (2 \cdot \delta_{0,36}, 0)\). The semismooth Newton algorithm is initialized with the data of the current iterate \((w_k, c_k)\), where the initial \( \lambda_k \) is defined by 1. The state error is calculated with respect to a reference state \( \bar{S}_\partial([\bar{v}, c], \bar{y}_d) \) with \( \tau = 2^{-10} \), \( h = 2 \cdot \sqrt{2} \cdot 2^{-8} \), where we used a nodal-wise projection on a considered coarser mesh. The forcing function \( B([\bar{v}, c]) \) in the reference state is discretized as we discussed above. Let us note that we have used the cost difference \( J([\bar{v}, c]) - J_{\partial}([\bar{v}, c]) \) as a termination criterion, in each discretization level \( \partial \), for the BV-PDAP algorithm. In detail, for each finer \( \partial \), we stopped Algorithm [1] if the cost difference \( J([\bar{v}, c]) - J_{\partial}(w_k, c_k) \) is quadratically decreasing compared to the cost difference of the last coarser \( \partial \).

In Algorithm [1] step 1, we additionally considered in the numerical experiment all stationary points of \( \int_0^T \int_{\Omega} L_\partial^*(\bar{S}_\partial(w_k, c_k) - y_d) g_t \ dx \ ds \) and thus obtained in some iteration steps more than one new point \( \hat{t}_i \), for \( i = 1, \ldots, m \). It turned out that this additional step accelerated the algorithm in finding an approximation of our solution. Numerically, we observed that Diracs where clustering nearby. Therefore, in an additional step to Algorithm [1] step 3, we merged all Diracs which where near to \( \hat{t}_i \), i.e. if for \( t \in A_{k,i} \) holds \( |t - \hat{t}_i| < \varepsilon \), with e.g. \( \varepsilon = 0.09 \), then the corresponding amplitude of \( \delta_i \) in \( \mathcal{W}_{A_{k,i}}(\bar{\lambda}) \), is added to the corresponding amplitude of \( \delta_{\hat{t}_i} \) in \( \mathcal{W}_{A_{k,i}}(\bar{\lambda}) \), for \( i = 1, \ldots, m \), where \( \delta_i \) is erased from \( \mathcal{W}_{A_{k,i}}(\bar{\lambda}) \). Furthermore, we observed that it is reasonable, after step 2, to remove all Diracs \( \delta_{\hat{t}_i}, t \in A_{k,i}, \) for \( i = 1, \ldots, m \), with small amplitudes \( \lambda_i \) related to \( \delta_{\hat{t}_i} \) before defining the next iterated \( w_{k+1} \). For example, we erased all Diracs with amplitude less than \( 10^{-5} \).

\textbf{REFERENCES}


