

## 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 ...]

AMS subject classifications: 26A45, 49J20, 49M25, 65N15, 65N30.

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:

$$(P) \begin{cases} \min_{u \in BV(0,T)^m} \frac{1}{2} \|y_u - y_d\|_{L^2(\Omega_T)}^2 + \sum_{j=1}^m \alpha_j \|D_t u_j\|_{M(I)} =: J(y, u) \\ \text{subject to} \quad (\mathcal{W}) \begin{cases} \partial_t y - \Delta y = f = \sum_{j=1}^m u_j g_j & \text{in } I \times \Omega \\ y = 0 & \text{on } I \times \partial\Omega \\ (y, \partial_t y) = (y_0, y_1) & \text{in } \{0\} \times \Omega, \end{cases} \end{cases}$$

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  $y_d \in C^1(\bar{I}; L^2(\Omega))$ . The time depending controls  $u$  are given by  $u = (u_1, \dots, u_m) \in BV(0, T)^m$ , and  $BV(0, T)^m$  is endowed with the norm  $\|u\|_{BV(I)^m} = \sum_{j=1}^m (\|u_j\|_{L^1(I)^m} + \|D_t u_j\|_{M(I)})$ . 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

†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)

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

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

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 depending function

$$\bar{p}_{1,i} : t \mapsto - \int_t^T \int_{\Omega} \bar{p} g_i \, dx \, ds,$$

where  $\bar{p}$  is the optimal adjoint state and  $p_{1,i}$  is bounded by  $\pm\alpha_i$ , is equal to  $\alpha_i$  at finitely many points in  $I$  and that its derivatives do not vanish in these points (see (A1) and (A2)) 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  $\bar{p}_{1,i}$ , 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  $L^1(I)$ -norm is estimated in terms of the error of the state variable in the  $L^2(\Omega_T)$ -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  $H_0^1(\Omega)$ . Hence, let us introduce for  $\alpha \geq 0$  the Hilbert spaces

$$\mathbb{H}^\alpha = \left\{ w \in L^2(\Omega) \mid \|w\|_{\mathbb{H}^\alpha}^2 := \sum_{k \geq 1} \left( \lambda^{(k)} \right)^\alpha \langle w, \mu_k \rangle_{L^2(\Omega)}^2 < \infty \right\}.$$

For  $\alpha = 0, 1$  we get  $L^2(\Omega)$  respectively  $H_0^1(\Omega)$ . The convexity of  $\Omega$  implies that  $\mathbb{H}^2 = H^2(\Omega) \cap H_0^1(\Omega)$ . In general holds  $\mathbb{H}^\alpha \hookrightarrow \mathbb{H}^\beta$  for  $\alpha \geq \beta$ . We denote the dual space of  $\mathbb{H}^\alpha$  by  $\mathbb{H}^{-\alpha}$ . 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 (Ladyjenskaya [1973], Chap.IV, Sec.4)

Let  $(f, y_0, y_1) \in L^1(I; L^2(\Omega)) \times H_0^1(\Omega) \times L^2(\Omega)$ . We call a function  $y \in C(\bar{I}; H_0^1(\Omega))$  with  $\partial_t y \in C(\bar{I}; L^2(\Omega))$  a weak solution of  $(\mathscr{W})$ , if

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

for any  $\eta \in L^1(I; H_0^1(\Omega))$  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.]:

**THEOREM 2.2** For each  $(f, y_0, y_1) \in L^1(0, T; L^2(\Omega)) \times H_0^1(\Omega) \times L^2(\Omega)$  there exists a unique weak solution  $y$  of  $(\mathcal{W})$ . Moreover, there exists a constant  $c > 0$  such that the weak solution  $y$  satisfies

$$\|y\|_{C(\bar{I}; \mathbb{H}^{\alpha+1})} + \|\partial_t y\|_{C(\bar{I}; \mathbb{H}^\alpha)} + \|\partial_{tt} y\|_{L^\kappa(I; \mathbb{H}^{\alpha-1})} \leq c \left( \|y_0\|_{\mathbb{H}^{\alpha+1}} + \|y_1\|_{\mathbb{H}^\alpha} + \|f\|_{L^\kappa(I; \mathbb{H}^\alpha)} \right) \quad (2.2)$$

provided  $(f, y_0, y_1) \in L^\kappa(I; \mathbb{H}^\alpha) \times \mathbb{H}^{\alpha+1} \times \mathbb{H}^\alpha$  and

$$\|y\|_{C(\bar{I}; \mathbb{H}^{\alpha+2})} + \|\partial_t y\|_{C(\bar{I}; \mathbb{H}^{\alpha+1})} + \|\partial_{tt} y\|_{C(\bar{I}; \mathbb{H}^\alpha)} \leq c \left( \|y_0\|_{\mathbb{H}^{\alpha+2}} + \|y_1\|_{\mathbb{H}^{\alpha+1}} + \|f\|_{W^{1,1}(I; \mathbb{H}^\alpha)} \right). \quad (2.3)$$

provided  $(f, y_0, y_1) \in W^{1,1}(I; \mathbb{H}^\alpha) \times \mathbb{H}^{\alpha+2} \times \mathbb{H}^{\alpha+1}$  with  $0 \leq \alpha, 1 \leq \kappa < \infty$ .

*Proof.* The proof can be found in [Zlotnik, 1994, Proposition 1.3, Remark 1.2].  $\square$

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

$$L: L^2(\Omega_T) \rightarrow L^2(\Omega_T), f \mapsto y(f) \quad \text{and} \quad Q: H_0^1(\Omega) \times L^2(\Omega) \rightarrow L^2(\Omega_T), (y_0, y_1) \mapsto y(y_0, y_1)$$

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  $p(w) \in C(\bar{I}; H_0^1(\Omega)) \cap C(\bar{I}; L^2(\Omega))$  is the weak solution of the backwards in time equation

$$(\mathcal{W}^*) \begin{cases} \partial_{tt} p - \Delta p = w & \text{in } I \times \Omega \\ p = 0 & \text{on } I \times \partial\Omega \\ (p, \partial_t p) = (0, 0) & \text{in } \{T\} \times \Omega. \end{cases} \quad (2.4)$$

**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  $\tilde{y} = y - y_0$ .  $\square$

### 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  $S_h \subset H_0^1(\Omega) \cap C(\bar{\Omega})$  and its nodal basis by  $(\varphi_i)_{i=1}^N$ .

#### 3.1 Space-Time Finite Element Method

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

$$S_\tau := \left\{ w \in C(\bar{I}) \mid w|_{[t_{k-1}, t_k]} \text{ linear}, 1 \leq k \leq M \right\}.$$

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

**DEFINITION 3.1** Let  $\sigma \geq 0$ . We call  $y_\vartheta \in \hat{S}_\vartheta := \text{span}\{v_h \cdot v_\tau | v_h \in S_h, v_\tau \in S_\tau\}$  a discrete solution of (2.1) if  $y_\vartheta$  satisfies:

$$\begin{aligned} \int_0^T -(\partial_t y_\vartheta, \partial_t \eta)_{L^2(\Omega)} - \left(\sigma - \frac{1}{6}\right) \tau^2 (\nabla \partial_t y_\vartheta, \nabla \partial_t \eta)_{L^2(\Omega)} + (\nabla y_\vartheta, \nabla \eta)_{L^2(\Omega)} dt \\ = (y_1, \eta(0))_{L^2(\Omega)} + \int_0^T (f, \eta)_{L^2(\Omega)} dt \end{aligned} \quad (3.1)$$

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

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

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

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

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

### 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  $\|\nabla \varphi\|_{L^2(\Omega)} \leq c_1 h^{-1} \|\varphi\|_{L^2(\Omega)}$  for all  $\varphi_h \in S_h$ . 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  $\|u\|_{C_\tau(I; L^2(\Omega))} := \max_{t_i \in \bar{w}^\tau} \|u(t_i)\|_{L^2(\Omega)}$  for all  $u \in C(\bar{I}; L^2(\Omega))$ .

LEMMA 3.2 The solution  $y_\vartheta$  of (3.1) for  $(f, y_0, y_1) \in L^1(I, L^2(\Omega)) \times H_0^1(\Omega) \times L^2(\Omega)$  satisfies the following inequality

$$\|y_\vartheta\|_{C(\bar{I}; L^2(\Omega))} \leq c \left( \|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|f\|_{L^1(I; L^2(\Omega))} \right) \quad (3.2)$$

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].  $\square$

THEOREM 3.2 The following error estimate holds

$$\|y - y_\vartheta\|_{C(\bar{I}; L^2(\Omega))} \leq c(h^2 + \tau^2)^{\frac{\alpha}{3}} \left( \|y_0\|_{\mathbb{H}^\alpha} + \|y_1\|_{\mathbb{H}^{\alpha-1}} + \|f\|_{W^{\alpha_2, 1}(I; \mathbb{H}^{\alpha_1})} \right) \quad (3.3)$$

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

*Proof.* The result follows directly from [Zlotnik, 1994, Theorem 4.1., 4.3.].  $\square$

COROLLARY 3.1 The following error estimate holds

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

*Proof.* This is shown by using that

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

where  $i_\tau: C(\bar{I}; L^2(\Omega)) \rightarrow \mathcal{S}_\tau \otimes L^2(\Omega)$  is the nodal interpolant. According to [Zlotnik, 1994, Theorem 4.1] we have

$$\begin{aligned} \|i_\tau y - y_\vartheta\|_{L^2(\Omega_T)} &\leq c \|i_\tau y - y_\vartheta\|_{C(\bar{I}; L^2(\Omega))} \\ &\leq c \|y - y_\vartheta\|_{C_\tau(\bar{I}; L^2(\Omega))} \leq c(\tau^2 + h^2) \left( \|y_0\|_{\mathbb{H}^3} + \|y_1\|_{\mathbb{H}^2} + \|f\|_{L^1(I; \mathbb{H}^2)} \right). \end{aligned}$$

Moreover, we have according to Theorem 2.2

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

This proves the assertion.  $\square$

#### 4. Equivalent Problem ( $\tilde{P}$ )

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].  $\square$

Next we introduce several linear and continuous operators and discuss its properties. The operator  $B: M(I)^m \times \mathbb{R}^m \rightarrow L^2(\Omega_T)$  is given by

$$(v, c) \mapsto \sum_{j=1}^m \left( \int_0^t dv_j(s) - \frac{1}{T} \int_0^T \int_0^t dv_j(s) ds + c_j \right) g_j. \quad (4.1)$$

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  $B^*: L^2(\Omega_T) \rightarrow C_0(I)^m \times \mathbb{R}^m$

$$B^*: q \mapsto \left( w'_1, \dots, w'_m, \int_0^T \int_{\Omega} qg_1 dx dt, \dots, \int_0^T \int_{\Omega} qg_m dx dt \right)$$

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

$$\begin{cases} -w''_j = \int_{\Omega} q(\cdot, x)g_j(x) dx - \frac{1}{T} \int_0^T \int_{\Omega} q(t, x)g_j(x) dx dt & \text{in } (0, T) \\ w'_j(0) = w'_j(T) = 0 & \text{with } \int_0^T w_j(t) dt = 0 \text{ for } j = 1, \dots, m. \end{cases} \quad (4.2)$$

**PROPOSITION 4.2** The operator  $B^*: L^2(\Omega_T) \rightarrow C_0(I) \times \mathbb{R}^m$  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  $\int_{\Omega} q(\cdot, x)g_j dx - \frac{1}{T} \int_0^T \int_{\Omega} q(t, x) dx dt \in L^2(I)$  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

$$\begin{aligned} \langle (v, c), B^*(q) \rangle &= \sum_{j=1}^m \int_0^T w'_j dv_j + \sum_{j=1}^m c_j \int_0^T \int_{\Omega} qg_j dx dt \\ &= \sum_{j=1}^m \int_0^T -w''_j \int_0^t dv_j dt + \sum_{j=1}^m c_j \int_0^T \int_{\Omega} qg_j dx dt \\ &= \sum_{j=1}^m \int_0^T \left( \int_{\Omega} qg_j dx - \frac{1}{T} \int_0^T \int_{\Omega} qg_j dx dt \right) \int_0^t dv_j dt + \sum_{j=1}^m c_j \int_0^T \int_{\Omega} qg_j dx dt \\ &= \int_0^T \int_{\Omega} q \sum_{j=1}^m \left( \int_0^t dv_j - \frac{1}{T} \int_0^T \int_0^t dv_j dt + c_j \right) g_j dx dt = \int_{\Omega_T} B(v, c)q dx dt \end{aligned}$$

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  $H_0^1(I)$ .  $\square$

**PROPOSITION 4.3** Let  $w_j \in H^2(I)$ ,  $j = 1, \dots, m$  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  $B: M(I)^m \times \mathbb{R}^m \rightarrow BV(I)^m \otimes \{(g_j)_{j=1}^m\}$  is an isomorphism.

*Proof.* The inverse of  $B$  is given by

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

□

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

$$(\tilde{P}) \begin{cases} \min_{\substack{v \in M(I)^m \\ c \in \mathbb{R}^m}} \frac{1}{2} \|S(v, c) - y_d\|_{L^2(\Omega_T)}^2 + \sum_{j=1}^m \alpha_j \|v_j\|_{M(I)} =: \tilde{J}(v, c), \end{cases}$$

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

#### 4.1 First-Order optimality condition of $(\tilde{P})$

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

$$\bar{p}_{1,i} := - \int_t^T \int_{\Omega} \bar{p} g_i dx ds$$

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

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

$$\bar{p}_{1,i} \in \alpha_i \partial \|\bar{v}_i\|_{M(I)} \quad i = 1, \dots, m, \quad (4.3)$$

$$\bar{p}_1(0) = 0. \quad (4.4)$$

Equivalently it holds

$$\langle v - \bar{v}_i, \bar{p}_{1,i} \rangle_{M(I), C_0(I)} + \alpha_i \|\bar{v}_i\|_{M(I)} \leq \alpha_i \|v\|_{M(I)} \quad \forall v \in M(I) \text{ and } i = 1, \dots, m \quad (4.5)$$

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

*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  $(\tilde{P})$  we have, that  $(\bar{v}, \bar{c}) \in M(I)^m \times \mathbb{R}^m$  is an optimal control of  $(\tilde{P})$  if and only if

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

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

$$DF_{(v,c)}(v, c) = B^* L^*(S(v, c) - y_d) \in C_0(I)^m \times \mathbb{R}^m$$



According to the theory of convex analysis, e.g. Ekeland and Témam [1999], we have

$$0 \in DF_{(v,c)}(\bar{v}, \bar{c}) + \partial \left( \sum_{i=1}^m \alpha_i \|\bar{v}_i\|_{M(I)} \right) \subseteq (M(I)^m \times \mathbb{R}^m)^*. \quad (4.6)$$

Using

$$\partial \left( \sum_{i=1}^m \alpha_i \|P_i \mathbb{D}(\bar{v}, \bar{c})\|_{M(I)} \right) = \sum_{i=1}^m \alpha_i \mathbb{D}^* P_i^* \partial \|\bar{v}_i\|_{M(I)} = \left( \begin{array}{c} (\alpha_i \partial \|\bar{v}_i\|_{M(I)})_{i=1}^m \\ 0 \end{array} \right)$$

and (4.6) as well as Proposition 4.3 imply

$$\bar{p}_{1,i} \in \alpha_i \partial \|\bar{v}_i\|_{M(I)} \quad \forall i = 1, \dots, m, \quad \bar{p}_1(0) = 0. \quad (4.7)$$

□

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

**PROPOSITION 4.6** Let  $(\bar{v}, \bar{c}) \in M(I)^m \times \mathbb{R}^m$  be an optimal control of  $(\tilde{P})$ , then for all  $i = 1, \dots, m$  and  $\bar{p}_1 = (\bar{p}_{1,i})_{i=1}^m$  given in (4.3) holds

- a)  $\|\bar{p}_{1,i}\|_{C_0(I)} \leq \alpha_i$ ,
- b)  $\int_0^T -\frac{\bar{p}_{1,i}}{\alpha_i} d\bar{v}_i = \int_0^T d|\bar{v}_i| = \|\bar{v}_i\|_{M(I)}$ ,
- c)  $\text{supp}(\bar{v}_i^\pm) \subseteq \{t \in I \mid \bar{p}_{1,i}(t) = \pm \alpha_i\}$ , where  $\bar{v}_i = \bar{v}_i^+ - \bar{v}_i^-$  is the Jordan decomposition of  $\bar{v}_i$ .

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

## 5. The Variationally Discretized Problem

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

$$(\tilde{P}_\vartheta^{\text{semi}}) \left\{ \begin{array}{l} \min_{\substack{v \in M(I)^m \\ c \in \mathbb{R}^m}} \frac{1}{2} \|S_\vartheta(v, c) - y_d\|_{L^2(\Omega_T)}^2 + \sum_{j=1}^m \alpha_j \|v_j\|_{M(I)} =: J_\vartheta(v, c) \end{array} \right.$$

with  $S_\vartheta : M(I)^m \times \mathbb{R}^m \rightarrow L^2(\Omega_T)$  defined by  $(v, c) \mapsto L_\vartheta(B(v, c)) + Q_\vartheta(y_0, y_1)$ . Here  $L_\vartheta : L^2(\Omega_T) \rightarrow L^2(\Omega_T)$  is defined by  $f \mapsto y_\vartheta(f)$ , where  $y_\vartheta(f)$  solves (3.1) for a source  $f$  and  $(y_0, y_1) = (0, 0)$ . The operator  $Q_\vartheta : H_0^1(\Omega) \times L^2(\Omega) \rightarrow L^2(\Omega_T)$  is defined by  $(y_0, y_1) \mapsto y_\vartheta(y_0, y_1)$ , where  $y_\vartheta(y_0, y_1)$  solves (3.1) with  $(y_0, y_1)$  as initial datum and  $f = 0$ .

**REMARK 5.1** We can represent the adjoint of  $L_\vartheta$  in the form  $w \mapsto L_\vartheta^*(w)(t, x) = L_\vartheta(w \circ \tilde{\phi})(\tilde{\phi}(t, x))$  with  $\tilde{\phi}(t, x) = (T - t, x)$ , and  $w \in L^2(\Omega_T)$ . This is true since  $L_\vartheta(f)(0) = 0$  and  $L_\vartheta^*(w)(T) = 0$  and thus  $L_\vartheta(f)$  and  $L_\vartheta^*(w)$  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  $L_\vartheta^*(w)$  as well.

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

*Proof.* The existence of an optimal control for  $(\tilde{P}_\vartheta^{\text{semi}})$  can be similarly shown as in the proof of Theorem 4.1. □

Note, that a BV-representation of the solutions  $(\bar{v}, \bar{c})$ ,  $(\bar{v}_\vartheta, \bar{c}_\vartheta)$  of  $(\tilde{P})$ , respectively  $(\tilde{P}_\vartheta^{\text{semi}})$  are defined by

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

Next we define the quantities  $\bar{p}_\vartheta = L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d)$  and

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

which is continuously differentiable and piecewise quadratic in time.

**THEOREM 5.2** The pair  $(\bar{v}_\vartheta, \bar{c}_\vartheta) \in M(I)^m \times \mathbb{R}^m$  is a optimal control of  $(\tilde{P}_\vartheta^{\text{semi}})$  if and only if

$$\bar{p}_{1,\vartheta,i} \in \alpha_i \partial \|\bar{v}_{\vartheta,i}\|_{M(I)} \quad i = 1, \dots, m, \quad (5.2)$$

$$\bar{p}_{1,\vartheta}(0) = 0. \quad (5.3)$$

Equivalently it holds

$$\langle v - \bar{v}_i, \bar{p}_{1,\vartheta,i} \rangle_{M(I), C_0(I)} + \alpha_i \|\bar{v}_i\|_{M(I)} \leq \alpha_i \|v\|_{M(I)} \quad \forall v \in M(I), \quad i = 1, \dots, m, \quad (5.4)$$

and  $\bar{p}_{1,\vartheta}(0) = 0$ .

*Proof.* The proof is similar to Theorem 4.5.  $\square$

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

## 6. A Priori Error Estimates

In this section error estimates of problem  $(\tilde{P}_\vartheta^{\text{semi}})$  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  $(\bar{v}_\vartheta, \bar{c}_\vartheta) \in M(I)^m \times \mathbb{R}^m$  of  $(\tilde{P}_\vartheta^{\text{semi}})$  and the optimal control  $(\bar{v}, \bar{c}) \in M(I)^m \times \mathbb{R}^m$  of  $(\tilde{P})$  we introduce the corresponding optimal states by  $\bar{y}_\vartheta := S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta)$  and  $\bar{y} := S(\bar{v}, \bar{c})$ . Further, we define the mixed state by  $\hat{y}_\vartheta := L_\vartheta(B(\bar{v}, \bar{c}) + Q_\vartheta(y_0, y_1))$ . The mixed adjoint state is chosen as  $\hat{p}_\vartheta := L_\vartheta^*(\bar{y} - y_d)$ . 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

$$\langle \bar{p}_{1,\vartheta} - \bar{p}_1, \bar{v}_\vartheta - \bar{v} \rangle \leq 0 \quad (6.1)$$

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

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

**LEMMA 6.2** Consider optimal control  $(\bar{v}, \bar{c})$  of  $(\tilde{P})$ , and  $(\bar{v}_\vartheta, \bar{c}_\vartheta)$  of  $(\tilde{P}_\vartheta^{\text{semi}})$ , as well as their BV-representations  $\bar{u}$ , and  $\bar{u}_\vartheta$ . For the optimal states  $\bar{y}$  and  $\bar{y}_\vartheta$  of problem  $(\tilde{P})$ , respectively  $(\tilde{P}_\vartheta^{\text{semi}})$ , we have

$$\|\bar{y}_\vartheta - \bar{y}\|_{L^2(\Omega_T)} \leq c \|\bar{y} - \hat{y}_\vartheta\|_{L^2(\Omega_T)} + c \|\bar{u}_\vartheta - \bar{u}\|_{L^1(I)^m}^{\frac{1}{2}} \|\bar{p} - \hat{p}_\vartheta\|_{L^\infty(I; L^2(\Omega))}^{\frac{1}{2}} \quad (6.2)$$

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

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

$$\begin{aligned} 0 &\geq \langle p_{1,\vartheta} - p_1, \bar{v}_\vartheta - \bar{v} \rangle = \langle B^*(\bar{p}_\vartheta - \bar{p}), (\bar{v}_\vartheta - \bar{v}, \bar{c}_\vartheta - \bar{c}) \rangle = \langle \bar{p}_\vartheta - \hat{p}_\vartheta, (\bar{u}_\vartheta - \bar{u}) \cdot g \rangle_{L^2(\Omega_T)} \\ &\quad + \langle \hat{p}_\vartheta - \bar{p}, (\bar{u}_\vartheta - \bar{u}) \cdot g \rangle_{L^2(\Omega_T)} \\ &= (\bar{y}_\vartheta - \bar{y}, L_\vartheta((\bar{u}_\vartheta - \bar{u})g))_{L^2(\Omega_T)} + \langle \hat{p}_\vartheta - \bar{p}, (\bar{u}_\vartheta - \bar{u}) \cdot g \rangle_{L^2(\Omega_T)} \\ &= \|\bar{y}_\vartheta - \bar{y}\|_{L^2(\Omega_T)}^2 + \langle \bar{y}_\vartheta - \bar{y}, \bar{y} - \hat{y}_\vartheta \rangle_{L^2(\Omega_T)} + \langle \hat{p}_\vartheta - \bar{p}, (\bar{u}_\vartheta - \bar{u}) \cdot g \rangle_{L^2(\Omega_T)}. \end{aligned}$$

From these calculations we obtained (6.2) by

$$\begin{aligned} \|\bar{y}_\vartheta - \bar{y}\|_{L^2(\Omega_T)}^2 &\leq (\bar{y} - \bar{y}_\vartheta, \bar{y} - \hat{y}_\vartheta)_{L^2(\Omega_T)} + (\bar{p} - \hat{p}_\vartheta, (\bar{u}_\vartheta - \bar{u}) \cdot g)_{L^2(\Omega_T)} \\ &\leq \frac{1}{2} \|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)}^2 + \frac{1}{2} \|\bar{y} - \hat{y}_\vartheta\|_{L^2(\Omega_T)}^2 + c \|\bar{u}_\vartheta - \bar{u}\|_{L^1(I)^m} \|\bar{p} - \hat{p}_\vartheta\|_{L^\infty(I; L^2(\Omega))}. \end{aligned}$$

□

LEMMA 6.3 The sequence of the BV representatives  $(\bar{u}_\vartheta)_\vartheta$  of the optimal controls of  $(\tilde{P}_\vartheta^{\text{semi}})$  are bounded in  $BV(I)^m$  with respect to  $\vartheta \rightarrow 0$ .

*Proof.* At first, we show that

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

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

$$\|y_\vartheta\|_{C(\bar{I}; L^2(\Omega))} \leq c \left( \|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} \right).$$

Thus, the discrete states  $y_\vartheta$  are bounded in  $L^2(\Omega_T)$ . Hence  $\{J_\vartheta(0)\}_\vartheta$  is bounded in  $\mathbb{R}$ . This implies that  $J_\vartheta(\bar{u}_\vartheta)$  is bounded and thus,  $(\bar{y}_\vartheta)_\vartheta$  and  $(D_t \bar{u}_\vartheta = \bar{v}_\vartheta)_\vartheta$  are bounded in  $L^2(\Omega_T)$ , and  $M(I)^m$  respectively. Now it suffices to show that  $\bar{c}_\vartheta \in \mathbb{R}^m$  is bounded in order to get the boundedness of  $(\bar{u}_\vartheta)_\vartheta$  in  $BV(I)^m$ . Assume that  $\bar{c}_\vartheta \in \mathbb{R}^m$  is unbounded for  $\vartheta \rightarrow 0$ . It holds

$$\sum_{j=1}^m \alpha_j \|D_t \hat{u}_{\vartheta, j}\|_{M(I)} = \sum_{j=1}^m \alpha_j \|D_t \bar{u}_{\vartheta, j}\|_{M(I)} \leq J_\vartheta(\bar{u}_\vartheta) \leq J_\vartheta(0)$$

and with the Poincare inequality for  $BV(I)$  functions ([Ambrosio et al., 2000, p. 152]), we get that  $(\hat{u}_\vartheta)_\vartheta$  is bounded in  $BV(I)^m$ . Consider  $z_\vartheta = \bar{y}_\vartheta - \tilde{y}_\vartheta$  with  $\tilde{y}_\vartheta = L_\vartheta(\hat{u}_\vartheta \cdot g) + Q_\vartheta(y_0, y_1)$ . The BV boundedness of  $(\hat{u}_\vartheta)_\vartheta$ , and therefore the boundedness in  $L^2(I)^m$ , implies by Lemma 3.2 that  $(\tilde{y}_\vartheta)_\vartheta$  is bounded in  $L^2(\Omega_T)$ . The boundedness of  $(\tilde{y}_\vartheta)_\vartheta$  and  $(\bar{y}_\vartheta)_\vartheta$  lead to the boundedness of  $(z_\vartheta)_\vartheta$  in  $L^2(\Omega_T)$ . The linearity of  $L_\vartheta(B(\cdot, \cdot))$ , implies  $z_\vartheta = L_\vartheta(B(0, \bar{c}_\vartheta))$ . Consider now  $p_\vartheta := \max_{1 \leq j \leq m} |\bar{c}_{\vartheta, j}|$ , with  $\bar{c}_\vartheta \rightarrow \infty$ ,  $\xi_\vartheta := \frac{1}{p_\vartheta} z_\vartheta$ , and  $\bar{a}_\vartheta = \frac{\bar{c}_\vartheta}{p_\vartheta}$ . There exists a  $\vartheta_0 > 0$  such that for all  $\vartheta < \vartheta_0$  the sequence  $\bar{a}_\vartheta$  is bounded by definition in  $\mathbb{R}^m$ . Thus, let us now consider  $\vartheta \leq \vartheta_0$  for all sequences in this proof. Hence, there exists a subsequence of  $\bar{a}_\vartheta$ , which converges to some  $\bar{a}$ . Denote this converging subsequence again by  $\bar{a}_\vartheta$ . The linear structure of  $L_\vartheta(B(\cdot, \cdot))$  gives us  $\xi_\vartheta = L_\vartheta(B(0, \bar{a}_\vartheta))$ . Define by  $\bar{\xi}_\vartheta$  the solution  $L_\vartheta(B(0, \bar{a}))$ . Next we show that Lemma 3.2 leads to  $\|\xi_\vartheta - \bar{\xi}_\vartheta\|_{L^2(\Omega_T)} \rightarrow 0$ . Thus, we have

$$\|\xi_\vartheta - \bar{\xi}_\vartheta\|_{L^2(\Omega_T)} = \|L_\vartheta(B(0, \bar{a}_\vartheta - \bar{a}))\|_{L^2(\Omega_T)} \leq c |\bar{a}_\vartheta - \bar{a}|_{\mathbb{R}^m} \rightarrow 0.$$

Define  $\bar{\xi}$  by  $L(B(0, \bar{a}))$ . Then we have

$$\|\bar{\xi} - \bar{\xi}_\vartheta\|_{C(\bar{I}; L^2(\Omega))} = \|L(B(0, \bar{a})) - L_\vartheta(B(0, \bar{a}))\|_{C(\bar{I}; L^2(\Omega))} \rightarrow 0$$

according to Theorem 3.2. With the boundedness of  $z_\vartheta$  in  $L^2(\Omega_T)$ , the unboundedness of  $|p_\vartheta|$ , and the definition of  $\xi_\vartheta = \frac{z_\vartheta}{p_\vartheta}$ , we can deduce that  $\xi_\vartheta \rightarrow 0$  in  $L^2(\Omega_T)$ . Hence,  $\bar{\xi}_\vartheta = \xi_\vartheta + (\bar{\xi}_\vartheta - \xi_\vartheta) \rightarrow 0$  in  $L^2(\Omega_T)$ . Thus, we obtain that  $\bar{\xi} = 0$ , which implies  $\sum_{j=1}^m \bar{a}_j g_j = 0$ . Because  $g_j \in L^\infty(\Omega) \setminus \{0\}$  have pointwise disjoint supports, we get that  $\bar{a}_j = 0$ , which is a contradiction. Thus, it is shown that  $\bar{c}_\vartheta$  is bounded, and hence  $(\bar{u}_\vartheta)_\vartheta$  is bounded in  $BV(I)^m$ .  $\square$

The next theorem states an a priori error estimate for the optimal state. Under additional assumptions on the structure of 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  $y_d \in C^1(\bar{I}; L^2(\Omega))$  and  $(y_0, y_1) \in H_0^1(\Omega) \times L^2(\Omega)$ , the following a priori error estimate holds

$$\|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)} \leq c(h^{2/3} + \tau^{2/3}) \left( \|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|y_d\|_{C^1(\bar{I}; L^2(\Omega))} \right). \quad (6.3)$$

For  $y_d \in C^1(\bar{I}; H_0^1(\Omega))$ ,  $g \in (\mathbb{H}^2)^m$ , and  $(y_0, y_1) \in \mathbb{H}^3 \times \mathbb{H}^2$ , the following error rate hold

$$\|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)} \leq c(h + \tau) \left( \|y_0\|_{\mathbb{H}^3} + \|y_1\|_{\mathbb{H}^2} + \|y_d\|_{C^1(\bar{I}; H_0^1(\Omega))} \right). \quad (6.4)$$

*Proof.* First we consider  $\|\bar{p} - \hat{p}_\vartheta\|_{L^\infty(I; L^2(\Omega))}^{\frac{1}{2}}$  in (6.2). By the regularity of  $B(\bar{v}, \bar{c}) \in L^2(\Omega_T)$ , we get that  $\bar{y} - y_d \in C^1(\bar{I}; L^2(\Omega))$  from Theorem 2.2. The error estimate in (3.3) with  $\alpha_1 = 0$  and  $\alpha_2 = 1$  implies then

$$\|\bar{p} - \hat{p}_\vartheta\|_{L^\infty(I; L^2(\Omega))}^{\frac{1}{2}} \leq c(h^2 + \tau^2)^{\frac{1}{3}} \|\bar{y} - y_d\|_{W^{1,1}(I; L^2(\Omega))}^{\frac{1}{2}}. \quad (6.5)$$

Consider now the term  $\|\bar{y} - \hat{y}_\vartheta\|_{L^2(\Omega_T)}$  in (6.2). Using (3.3) with  $\alpha_1 = 0$  and  $\alpha_2 = 0$  implies

$$\|\bar{y} - \hat{y}_\vartheta\|_{L^2(\Omega_T)} \leq c(h^2 + \tau^2)^{\frac{1}{3}} \left( \|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|B(\bar{v}, \bar{c})\|_{L^1(I; L^2(\Omega))} \right). \quad (6.6)$$

Hence, Lemma 6.3, (6.2), (6.5) and (6.6) imply (6.3). Assume that  $y_d \in C^1(\bar{I}; H_0^1(\Omega))$ ,  $g \in (\mathbb{H}^2)^m$  and  $(y_0, y_1) \in \mathbb{H}^3 \times \mathbb{H}^2$  hold. By Theorem 2.2 we have  $\bar{y} - y_d \in C^1(\bar{I}; H_0^1(\Omega))$ . Using (3.3) with  $\alpha_1 = 1$  and  $\alpha_2 = 1$ , implies

$$\|\bar{p} - \hat{p}_\vartheta\|_{L^\infty(I; L^2(\Omega))}^{\frac{1}{2}} \leq c(h^2 + \tau^2)^{\frac{1}{2}} \|\bar{y} - y_d\|_{W^{1,1}(I; H_0^1(\Omega))}^{\frac{1}{2}}. \quad (6.7)$$

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

**LEMMA 6.4** Let  $p$  be the weak solution of  $\mathcal{W}^*$  for  $w \in C^1(\bar{I}; \mathbb{H}^\gamma)$  with  $\gamma = 0, 1$  and  $p_\vartheta$  its discrete counterpart. Then there holds

$$\left\| R_h \int_0^T p \, dt - \int_0^T p_\vartheta \, dt \right\|_{H_0^1(\Omega)} \leq c(h^2 + \tau^2)^{(\gamma+2)/3} \|w\|_{C^1(\bar{I}; \mathbb{H}^\gamma)}.$$

*Proof.* This follows directly from [Zlotnik, 1994, Theorem 4.1].  $\square$

To avoid repetitions and increase readability, we introduce the following constant  $\kappa$ , 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

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

**THEOREM 6.2** For the optimal control  $(\bar{v}, \bar{c})$  of  $(\tilde{P})$  and solutions  $(\bar{v}_\vartheta, \bar{c}_\vartheta)$  of  $(\tilde{P}_\vartheta^{\text{semi}})$  the following a priori error estimates hold:

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

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

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

$$J(\bar{v}, \bar{c}) \leq J(\bar{v}_\vartheta, \bar{c}_\vartheta) \quad \text{and} \quad J_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) \leq J_\vartheta(\bar{v}, \bar{c}).$$

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

$$\begin{aligned} & J(v, c) - J_\vartheta(v, c) \\ &= -\frac{1}{2} \|S(v, c) - S_\vartheta(v, c)\|_{L^2(\Omega_T)}^2 + (B(v, c), L^*(S(v, c) - y_d) - L_\vartheta^*(S(v, c) - y_d))_{L^2(\Omega_T)} \\ & \quad + (Q(y_0, y_1) - Q_\vartheta(y_0, y_1), S(v, c) - y_d)_{L^2(\Omega_T)} \end{aligned} \quad (6.10)$$

Lemmata 2.2 and 3.1 imply

$$\begin{aligned} & (Q(y_0, y_1) - Q_\vartheta(y_0, y_1), S(v, c) - y_d)_{L^2(\Omega_T)} \\ &= (y_1, p(0) - p_\vartheta(0))_{L^2(\Omega)} + \left( y_0 - R_h y_0, \int_0^T S(v, c) - y_d \, dt \right)_{L^2(\Omega)} \\ & \quad + \left( \nabla(R_h y_0 - y_0), \nabla \int_0^T p \, dt \right)_{L^2(\Omega)} + \left( \nabla R_h y_0, \nabla \int_0^T p_\vartheta - p \, dt \right)_{L^2(\Omega)} \end{aligned} \quad (6.11)$$

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

$$\begin{aligned} (y_1, p(0) - p_\vartheta(0))_{L^2(\Omega)} &\leq \|y_1\|_{L^2(\Omega)} \|p - p_\vartheta\|_{C(\bar{I}; L^2(\Omega))} \\ &\leq c(\tau^\alpha + h^\alpha) \|y_1\|_{L^2(\Omega)} \|S(v, c) - y_d\|_{C^1(\bar{I}; \mathbb{H}^{\alpha-1})}. \end{aligned}$$

Further we have

$$\begin{aligned} \left( y_0 - R_h y_0, \int_0^T S(v, c) - y_d dt \right)_{L^2(\Omega)} &\leq \|y_0 - R_h y_0\|_{L^2(\Omega)} \|S(v, c) - y_d\|_{L^2(\Omega_T)} \\ &\leq ch^\alpha \|y_0\|_{\mathbb{H}^\alpha} \|S(v, c) - y_d\|_{L^2(\Omega_T)}. \end{aligned}$$

Since  $S(v, c) - y_d \in C^1(\bar{I}; L^2(\Omega))$  we have  $p \in C(\bar{I}, \mathbb{H}^2)$  according to Theorem 2.2 and thus

$$\begin{aligned} \left( \nabla(R_h y_0 - y_0), \nabla \int_0^T p dt \right)_{L^2(\Omega)} &= \left( R_h y_0 - y_0, -\Delta \int_0^T p dt \right)_{L^2(\Omega)} \\ &\leq c \|y_0 - R_h y_0\|_{L^2(\Omega)} \|p\|_{L^2(I, \mathbb{H}^2)} \\ &\leq ch^\alpha \|y_0\|_{\mathbb{H}^\alpha} \|S(v, c) - y_d\|_{C^1(\bar{I}, L^2(\Omega))}. \end{aligned}$$

Finally, we have according to Lemma 6.4

$$\begin{aligned} \left( \nabla R_h y_0, \nabla \int_0^T p_\vartheta - p dt \right)_{L^2(\Omega)} &= \left( \nabla R_h y_0, \nabla \left( \int_0^T p_\vartheta - R_h \int_0^T p dt \right) \right)_{L^2(\Omega)} \\ &\leq c \|y_0\|_{L^2(\Omega)} \left\| \int_0^T p_\vartheta - R_h \int_0^T p dt \right\|_{H_0^1(\Omega)} \leq c \|y_0\|_{L^2(\Omega)} (\tau^\alpha + h^\alpha) \|S(v, c) - y_d\|_{C^1(\bar{I}, \mathbb{H}^{\alpha-1})}. \end{aligned}$$

For  $(B(v, c), y_0, y_1, y_d) \in L^2(\Omega_T) \times H_0^1(\Omega) \times L^2(\Omega) \times C^1(\bar{I}; L^2(\Omega))$ , the function  $S(v, c) - y_d$  is an element of  $C^1(\bar{I}; L^2(\Omega))$  according to Theorem 2.2. If  $(g, y_0, y_1, y_d) \in (\mathbb{H}^2)^m \times \mathbb{H}^3 \times \mathbb{H}^2 \times C^1(\bar{I}; H_0^1(\Omega))$  we get that  $S(v, c) - y_d \in C^1(\bar{I}; H_0^1(\Omega))$ . Hence, the a priori estimate (3.3) implies

$$|(B(v, c), L^*(S(v, c) - y_d) - L_\vartheta^*(S(v, c) - y_d))_{L^2(\Omega_T)}| = O(\tau^\alpha + h^\alpha). \quad (6.12)$$

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

### 6.1 Optimal Convergence Rates for the Optimal Controls of $(\tilde{P}_\vartheta^{semi})$

Under certain assumptions we show that the BV-representations  $\bar{u}_\vartheta$  of the optimal controls of  $(\tilde{P}_\vartheta^{semi})$  converge with a specific rate in the  $L^1$ -norm to the optimal control  $\bar{u}$  of  $(P)$ . Further, define the following functions:

$$z(t) := \partial_t p_1(t) = \int_\Omega L^*(S(\bar{v}, \bar{c}) - y_d)(t) g dx \quad (6.13)$$

$$z_\vartheta(t) := \partial_t p_{1, \vartheta}(t) = \int_\Omega L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d)(t) g dx \quad (6.14)$$

with  $(\bar{v}, \bar{c})$  as the optimal control of  $(\tilde{P})$  and  $(\bar{v}_\vartheta, \bar{c}_\vartheta)$  as optimal control of  $(\tilde{P}_\vartheta^{semi})$ . Due to Proposition 4.6 and Remark 5.2, it holds that  $\text{supp}(\bar{v}_i) \subseteq \{t | z_i(t) = 0\}$  and  $\text{supp}(\bar{v}_{\vartheta, i}) \subseteq \{t | z_{\vartheta, i}(t) = 0\}$ .

LEMMA 6.5 The matrix  $G := (\langle Lg_i, Lg_j \rangle_{L^2(\Omega_T)})_{i, j=1}^m \in \mathbb{R}^{m \times m}$  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  $\{g_i\}_{i=1}^m$  is a linear independent system.  $\square$

**THEOREM 6.3**  $\bar{u}_\vartheta$  converges weakly\* in  $BV(0, T)^m$  to the solution  $\bar{u}$  for  $\vartheta \rightarrow 0$ .

*Proof.* Let  $(\vartheta_n)_{n=1}^\infty = (\tau_n, h_n)$  be a null sequence such that  $(\tau_n)_{n=1}^\infty \subset \mathbb{R}^+$ ,  $(h_n)_{n=1}^\infty \subset \mathbb{R}^+$ . Lemma 6.3 implies that  $(\bar{u}_{\vartheta_n})_{n=1}^\infty$  is a bounded sequence in  $BV(I)^m$  where  $(\bar{v}_{\vartheta_n}, \bar{c}_{\vartheta_n})$  are optimal controls of  $(P_{\vartheta_n}^{semi})$ . The weak\* compactness of closed and bounded sets in  $BV(I)^m$  implies the existence of a subsequence  $(\bar{u}_{\vartheta_{n_k}})_k$  which converges weakly\* to some  $\tilde{u} \in BV(I)^m$ . Hence,  $(\bar{u}_{\vartheta_{n_k}})_k$  converges in  $L^2(I)^m$  to  $\tilde{u}$  and  $D_t \bar{u}_{\vartheta_{n_k}} = \bar{v}_{\vartheta_{n_k}}$  converges weakly\* in  $M(I)^m$  to  $D_t \tilde{u}$ . There exists a unique element  $(\tilde{v}, \tilde{c}) \in M(I)^m \times \mathbb{R}^m$  such that  $\tilde{u} = \int_0^T d\tilde{v}(s) - \frac{1}{T} \int_0^T \int_0^t d\tilde{v}(s) dt + \tilde{c}$ . Due to the weak\* l.s.c. of  $\|\cdot\|_{M(I)}$  in  $M(I)$ , we get

$$\liminf_{k \rightarrow \infty} \sum_{i=1}^m \alpha_i \|\bar{v}_{\vartheta_{n_k, i}}\|_{M(I)} \geq \sum_{i=1}^m \alpha_i \|D_t \tilde{u}_i\|_{M(I)}. \quad (6.15)$$

Let us show that

$$\lim_{k \rightarrow \infty} \|S_{\vartheta_{n_k}}(\bar{v}_{\vartheta_{n_k}}, \bar{c}_{\vartheta_{n_k}}) - y_d\|_{L^2(\Omega_T)}^2 = \|S(\tilde{v}, \tilde{c}) - y_d\|_{L^2(\Omega_T)}^2 \quad (6.16)$$

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

$$\begin{aligned} \|S_{\vartheta_{n_k}}(\bar{v}_{\vartheta_{n_k}}, \bar{c}_{\vartheta_{n_k}}) - S(\tilde{v}, \tilde{c})\|_{L^2(\Omega_T)} &\leq \|S_{\vartheta_{n_k}}(\bar{v}_{\vartheta_{n_k}}, \bar{c}_{\vartheta_{n_k}}) - S_{\vartheta_{n_k}}(\tilde{v}, \tilde{c})\|_{L^2(\Omega_T)} \\ &\quad + \|S_{\vartheta_{n_k}}(\tilde{v}, \tilde{c}) - S(\tilde{v}, \tilde{c})\|_{L^2(\Omega_T)} \leq c \|\bar{u}_{\vartheta_k} - \tilde{u}\|_{L^2(I)^m} \\ &\quad + c(h_{n_k}^2 + \tau_{n_k}^2)^{\frac{1}{3}} (\|y_0\|_{H_0^1(\Omega)} + \|y_1\|_{L^2(\Omega)} + \|B(\tilde{v}, \tilde{c})\|_{L^1(I; L^2(\Omega))}) \end{aligned}$$

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

$$J(\bar{v}, \bar{c}) = \liminf_{k \rightarrow \infty} J_{\vartheta_k}(\bar{v}_{\vartheta_k}, \bar{c}_{\vartheta_k}) \geq J(\tilde{v}, \tilde{c}).$$

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

**COROLLARY 6.1** There holds  $\bar{u}_\vartheta \rightarrow \bar{u}$  in  $L^2(I)$  for  $\vartheta \rightarrow 0$ .

Next we prove pointwise convergence of  $z_\vartheta$  and  $\partial_t z_\vartheta$ .

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

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

$$\begin{aligned} \|z_\vartheta - z\|_{L^\infty(I)^m} &\leq c \sup_{t \in \bar{I}} \|L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d)(t, \cdot) - L^*(S(\bar{v}, \bar{c}) - y_d)(t, \cdot)\|_{L^1(\Omega)} \\ &\leq c \left[ \|L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} + \|L_\vartheta^*(y_d) - L^*(y_d)\|_{C(\bar{I}; L^2(\Omega))} \right]. \quad (6.17) \end{aligned}$$

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

$$\|L_\vartheta^*(y_d) - L^*(y_d)\|_{C(\bar{I}; L^2(\Omega))} \leq c(h^2 + \tau^2)^{\frac{1}{3}} \|y_d\|_{L^1(I; L^2(\Omega))}. \quad (6.18)$$

Furthermore, we have

$$\begin{aligned} \|L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} &\leq \|L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d) - L_\vartheta^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} \\ &\quad + \|L_\vartheta^*(S(\bar{v}, \bar{c}) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} \quad (6.19) \end{aligned}$$

For the second term on the right hand side of (6.19), we have using Theorem 3.2:

$$\|L_{\vartheta}^*(S(\bar{v}, \bar{c})) - L^*(S(\bar{v}, \bar{c}))\|_{C(\bar{I}; L^2(\Omega))} \leq c(h^2 + \tau^2)^{\frac{1}{3}} \|S(\bar{v}, \bar{c})\|_{L^1(I; L^2(\Omega))}. \quad (6.20)$$

For the first term on the right hand side of (6.19), we use the stability of  $L_{\vartheta}$  and  $L_{\vartheta}^*$  from Lemma 3.2 to obtain

$$\begin{aligned} \|L_{\vartheta}^*(S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})) - L_{\vartheta}^*(S(\bar{v}, \bar{c}))\|_{C(\bar{I}; L^2(\Omega))} &\leq c \|S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - S(\bar{v}, \bar{c})\|_{C(\bar{I}; L^2(\Omega))} \\ &\leq c \left( \|S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - S_{\vartheta}(\bar{v}, \bar{c})\|_{C(\bar{I}; L^2(\Omega))} + \|S_{\vartheta}(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{C(\bar{I}; L^2(\Omega))} \right) \\ &\leq c \|\bar{u}_{\vartheta} - \bar{u}\|_{L^2(I)} + \|S_{\vartheta}(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{C(\bar{I}; L^2(\Omega))}. \end{aligned} \quad (6.21)$$

The strong convergence of  $\bar{u}_{\vartheta}$  to  $\bar{u}$  in  $L^2(I)$ , see Corollary 6.1, and Theorem 3.2 imply that  $\|L_{\vartheta}^*(S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})) - L_{\vartheta}^*(S(\bar{v}, \bar{c}))\|_{C(\bar{I}; L^2(\Omega))}$  converges to 0 for  $\vartheta \rightarrow 0$ . Hence,  $\|z_{\vartheta} - z\|_{L^{\infty}(I)^m}$  converges to 0 for  $\vartheta \rightarrow 0$ .  $\square$

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

$$\|\bar{\partial}_t L_{\vartheta}^*(f)\|_{C_{\tau}(I; L^2(\Omega))} \leq c \|f\|_{L^1(I; L^2(\Omega))}. \quad (6.22)$$

*Proof.* This follows directly from [Zlotnik, 1994, Theorem 2.1].  $\square$

LEMMA 6.8 The following a priori error estimate

$$\|\bar{\partial}_t(L^*(f) - L_{\vartheta}^*(f))\|_{C_{\tau}(I; L^2(\Omega))} \leq (h^2 + \tau^2)^{\frac{1}{3}} \|f\|_{W^{1,1}(I; L^2(\Omega))}$$

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

*Proof.* This follows directly from [Zlotnik, 1994, Theorem 4.2].  $\square$

LEMMA 6.9 We have

$$\|\bar{\partial}_t(z_{\vartheta} - z)\|_{C_{\tau}(I)^m} \rightarrow 0 \quad \text{for } \vartheta \rightarrow 0. \quad (6.23)$$

*Proof.* Lemma 6.7 implies

$$\begin{aligned} \|\bar{\partial}_t(z_{\vartheta} - z)\|_{C_{\tau}(I)^m} &= \sum_{\ell=1}^m \sup_{t_i \in \bar{w}^{\tau}} \left| \bar{\partial}_t \int_{\Omega} (L_{\vartheta}^*(S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - y_d)(t_i) - L^*(S(\bar{v}, \bar{c}) - y_d)(t_i)) g_{\ell} dx \right| \\ &\leq c \|\bar{\partial}_t(L_{\vartheta}^*(S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d))\|_{C_{\tau}(I; L^2(\Omega))} \\ &\leq c \|\bar{\partial}_t(L_{\vartheta}^*(S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - y_d) - L_{\vartheta}^*(S(\bar{v}, \bar{c}) - y_d))\|_{C_{\tau}(I; L^2(\Omega))} \\ &\quad + c \|\bar{\partial}_t(L_{\vartheta}^*(S(\bar{v}, \bar{c}) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d))\|_{C_{\tau}(I; L^2(\Omega))} \\ &\leq c \|S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - S(\bar{v}, \bar{c})\|_{L^1(I; L^2(\Omega))} \\ &\quad + c \|\bar{\partial}_t(L_{\vartheta}^*(S(\bar{v}, \bar{c}) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d))\|_{C_{\tau}(I; L^2(\Omega))} \end{aligned} \quad (6.24)$$

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

LEMMA 6.10 We have

$$\|\bar{\partial}_t(z_{\vartheta} - z)\|_{L^{\infty}(I)^m} \rightarrow 0 \quad \text{for } \vartheta \rightarrow 0. \quad (6.25)$$



*Proof.* At first we define a cell-wise discretization of the derivative of  $z$  as follows

$$\bar{\delta}_t z := \sum_{i=1}^M \frac{z(t_i) - z(t_{i-1})}{\tau} \mathbf{1}_{I_i}, \text{ with } I_i := (t_{i-1}, t_i), i = 1, \dots, M. \quad (6.26)$$

Then we proceed with

$$\|\partial_t(z_\vartheta - z)\|_{L^\infty(I)^m} \leq \|\partial_t z_\vartheta - \bar{\delta}_t z\|_{L^\infty(I)^m} + \|\bar{\delta}_t z - \partial_t z\|_{L^\infty(I)^m}.$$

Using the disjoint supports of the characteristic functions in the definition of  $\bar{\delta}_t z$  leads to

$$\|\partial_t z_\vartheta - \bar{\delta}_t z\|_{L^\infty(I)^m} = \sum_{j=1}^m \max_{i=1, \dots, M} \left| \frac{z_\vartheta^j(t_i) - z_\vartheta^j(t_{i-1})}{\tau} - \frac{z_j(t_i) - z_j(t_{i-1})}{\tau} \right| = \|\bar{\partial}_t(z_\vartheta - z)\|_{C_t(I)^m}$$

which converges to 0 under the consideration of (6.23). Further, calculations show that

$$\begin{aligned} \|\bar{\delta}_t z - \partial_t z\|_{L^\infty(I)^m} &= \left\| \sum_{i=1}^M \left( \frac{z(t_i) - z(t_{i-1})}{\tau} - \partial_t z(t) \right) \mathbf{1}_{I_i}(t) \right\|_{L^\infty(I)^m} \\ &\leq \left\| \sum_{i=1}^M \left( \frac{z(t_i) - z(t_{i-1})}{\tau} - \partial_t z(t_i) \right) \mathbf{1}_{I_i}(t) \right\|_{L^\infty(I)^m} + \left\| \sum_{i=1}^M (\partial_t z(t_i) - \partial_t z(t)) \mathbf{1}_{I_i}(t) \right\|_{L^\infty(I)^m} \\ &= \sum_{j=1}^m \left( \max_{i=1, \dots, M} \left| \frac{z_j(t_i) - z_j(t_{i-1})}{\tau} - \partial_t z_j(t_i) \right| + \max_{i=1, \dots, M} \sup_{t \in I_i} |\partial_t z_j(t_i) - \partial_t z_j(t)| \right). \end{aligned}$$

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  $\partial_t z(t)$  in  $\bar{I}$ . Hence, the result follows for  $\vartheta \rightarrow 0$ , which implies the claim.  $\square$

LEMMA 6.11 The convergence  $\|p_{1,\vartheta} - p_1\|_{L^\infty(I)^m} \rightarrow 0$  holds for  $\vartheta \rightarrow 0$ .

*Proof.* This follows directly from

$$\|p_{1,\vartheta} - p_1\|_{L^\infty(I)^m} \leq c \|L_\vartheta^*(S_\vartheta(\bar{v}_\vartheta, \bar{c}_\vartheta) - y_d) - L^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}, L^2(\Omega))},$$

which converges to 0 using the same steps as in Lemma 6.6.  $\square$

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)**  $\{t \in I \mid |\bar{p}_{1,i}(t)| = \alpha\} = \{t_{1,i}, \dots, t_{m_i,i}\}$  for  $m_i \in \mathbb{N}$ , with  $i = 1, \dots, m$ .

(A2)  $\partial_t z_i(t_{j,i}) \neq 0$ , for  $j = 1, \dots, m_i$  and  $i = 1, \dots, m$ .

REMARK 6.1 The assumption (A1) enforces finitely many jumps for the optimal control of  $(P)$ , i.e. it holds  $\text{supp}(D_t \bar{u}_i) \subseteq \{t_{1,i}, \dots, t_{m_i,i}\}$  for  $\bar{u} \in BV(I)^m$ .

LEMMA 6.12 Let  $(\bar{v}_\vartheta, \bar{c}_\vartheta)$  be an optimal control of  $(\tilde{P}_\vartheta^{\text{semi}})$ . Under the assumptions (A1) and (A2) above, there exists a  $\delta > 0$ , and  $\vartheta_0 > 0$  such that for all  $0 < \vartheta \leq \vartheta_0$  holds

$$\bar{v}_{\vartheta,i} = \sum_{l=1}^{m_i} c_{l,\vartheta}^i \delta_{t_{l,\vartheta}^i} \text{ with } c_{l,\vartheta}^i \in \mathbb{R} \text{ and } t_{l,\vartheta}^i \in B_\delta(t_{j,i}),$$

where  $B_\delta(t_{j,i})$  are pairwise disjoint for a fixed  $i = 1, \dots, m$  with respect to the index  $j = 1, \dots, m_i$  and with  $0 < \vartheta \rightarrow 0$ . The coefficients in front of the Dirac measures of  $\bar{v}_{\vartheta,i}$ , i.e.  $c_{l,\vartheta}^i$  for  $l = 1, \dots, m_i$ , are possibly 0.

*Proof.* Let us begin with the case  $m = 1, m_1 = 1$ , i.e.  $\{t \in I \mid |p_1(t)| = \alpha\} = \{\tilde{t}_1\}$ . First of all we know that  $|p_1(t)| \leq \alpha$  for all  $t \in \bar{I}$  holds and since  $\bar{p}_1 \in C^1(\bar{I})$  as well as that  $\tilde{t}_1$  is an interior point follows  $z(\tilde{t}_1) = -\partial_t p_1(\tilde{t}_1) = 0$ . Moreover, due to (A2) there exists a  $\delta > 0$  and  $c_1 > 0$  such that  $|\partial_t z(t)| > c_1$  for all  $t \in B_\delta(\tilde{t}_1) \subset I$ . Since  $\partial_t z$  is continuous,  $\partial_t z$  does not change its sign on  $B_\delta(\tilde{t}_1)$  and hence  $z$  is strictly monotone in  $B_\delta(\tilde{t}_1)$ . Therefore  $\tilde{t}_1$  is the only root of  $z$  in  $B_\delta(\tilde{t}_1)$ . Moreover, there exist  $t_-, t_+ \in B_\delta(\tilde{t}_1)$  with  $z(t_-) < 0 < z(t_+)$ . By Lemma 6.6 there exists a  $\vartheta_0 = (\tau_0, h_0)$  such that  $z_\vartheta(t_-) < 0 < z_\vartheta(t_+)$  for all  $\vartheta < \vartheta_0$ . Since  $z_\vartheta$  is continuous there exists a  $t_\vartheta \in (t_-, t_+)$  such that  $z_\vartheta(t_\vartheta) = 0$  for all  $\vartheta < \vartheta_0$ . Next we show that there exists a  $\tilde{\vartheta}_0 \leq \vartheta_0$  such that  $t_\vartheta$  is the only root of  $z_\vartheta$  in  $B_\delta(\tilde{t}_1)$  for all  $\vartheta < \tilde{\vartheta}_0$ . Lemma 6.10 implies existence of a  $\tilde{\vartheta}_0 < \vartheta_0$  that  $\partial_t z_\vartheta$  is either strictly positive or strictly negative on  $B_\delta(\tilde{t}_1)$ . Now let  $\hat{t}_\vartheta$  be a second root of  $z_\vartheta$  in  $B_\delta(\tilde{t}_1)$ . Then it holds

$$0 = z_\vartheta(t_\vartheta) - z_\vartheta(\hat{t}_\vartheta) = \int_{\hat{t}_\vartheta}^{t_\vartheta} \partial_t z_\vartheta(t) dt \neq 0.$$

Hence, there is no second root of  $z_\vartheta$  in  $B_\delta(\tilde{t}_1)$ . Next we show that  $t \neq t_\vartheta$  and  $z_\vartheta(t) = 0$  imply the existence of  $\hat{\vartheta}_0 < \tilde{\vartheta}_0$  such that  $|p_{1,\vartheta}(t)| < \alpha$  for all  $\vartheta < \hat{\vartheta}_0$  and thus  $\bar{v}_\vartheta = c_{1,\vartheta} \delta_{t_\vartheta}$  with  $c_{1,\vartheta}$  possibly zero. Such a  $t$  can only exist in  $\bar{I} \setminus B_\delta(\tilde{t}_1)$ . Due to Assumption 6.1 and the condition  $|p_{1,\vartheta}(t)| \leq \alpha$  for all  $t \in \bar{I}$  there exists a  $\varepsilon > 0$  such that  $|p_1(t)| < \alpha - \varepsilon$  for all  $t \in \bar{I} \setminus B_\delta(\tilde{t}_1)$ . Lemma 6.11 implies the existence of a  $\hat{\vartheta}_0 < \tilde{\vartheta}_0$  with  $|p_{1,\vartheta}(t)| < \alpha - \varepsilon/2$  for all  $\vartheta < \hat{\vartheta}_0$  and  $t \in \bar{I} \setminus B_\delta(\tilde{t}_1)$ . In the case of  $m = 1$  and  $\{t \in I \mid |p_1(t)| = \alpha\} = \{\tilde{t}_1, \dots, \tilde{t}_{m_1}\}$  with  $m_1 > 1$ , we can find for each  $\tilde{t}_i$  a  $\delta_i > 0$  with  $\bigcap_{i=1}^{m_1} B_{\delta_i}(\tilde{t}_i) = \emptyset$  and a  $\vartheta_i$  such that there exists a  $t_\vartheta^i \in B_{\delta_i}(\tilde{t}_i)$  with  $\bar{v}_\vartheta|_{B_{\delta_i}(\tilde{t}_i)} = c_{i,\vartheta} \delta_{t_\vartheta^i}$  and  $\bar{v}_\vartheta|_{\bar{I} \setminus \bigcup_{i=1}^{m_1} B_{\delta_i}(\tilde{t}_i)} = 0$ . Then choose  $\vartheta_0 < \min_{i=1, \dots, m_1} \vartheta_i$ . In the case of  $m > 1$ , one has to consider the same proof as above with respect to an additional subindex  $i = 1, \dots, m$ , and the smallest  $\vartheta_0$  and  $\delta$  used in the proofs of each component  $i = 1, \dots, m$ .  $\square$

From now on, we will assume that  $\vartheta \leq \vartheta_0$  holds with  $\vartheta_0$  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  $\tilde{\delta} > 0$  for which  $\tilde{\delta} < \text{dist}(B_\delta(t_{j_1,i}), \{0, T\})$ , and  $\tilde{\delta} < \text{dist}(B_\delta(t_{j_1,i}), B_\delta(t_{j_2,i}))$ ,  $j_1, j_2 = 1, \dots, m_i$ ,  $j_1 \neq j_2$  for  $i = 1, \dots, m$ . Let us note that Remark 4.1 and Lemma 6.12 guarantee that such a  $\tilde{\delta} > 0$  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  $(P_\vartheta^{\text{semi}})$  in a more explicit form

$$\bar{u}_i = c_i + \sum_{j=1}^{m_i} c_j^i \left( \mathbf{1}_{(t_{j,i}, T)}(t) - \frac{T - t_{j,i}}{T} \right), \quad \bar{u}_{i,\vartheta} = c_{i,\vartheta} + \sum_{j=1}^{m_i} c_{j,\vartheta}^i \left( \mathbf{1}_{(t_{j,\vartheta}^i, T)}(t) - \frac{T - t_{j,\vartheta}^i}{T} \right)$$

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

**LEMMA 6.13** The following inequality holds

$$\|\bar{u} - \bar{u}_\vartheta\|_{L^1(I)^m} \leq c \left( \sum_{i=1}^m (|c_i - c_{i,\vartheta}| + \sum_{j=1}^{m_i} |c_j^i| \cdot |t_{j,i} - t_{j,\vartheta}^i| + |c_j^i - c_{j,\vartheta}^i|) \right) \quad (6.27)$$

for some constant  $c$  which depends only on  $T$ .

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

**LEMMA 6.14** For each  $t_{j,i}$  there is a function  $f_{j,i} \in C_c^\infty(\Omega_T)$ , with  $j = 1, \dots, m_i$  and  $i = 1, \dots, m$ , such that the function

$$g_i^j(t, x) := \int_t^T L^*(f_{j,i})(s, x) ds \in C_0(I; L^2(\Omega)) \quad (6.28)$$

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  $f_{j,i} = f_i \partial_t h_{i,j} - h_{i,j} \Delta f_i$ ,
- b)  $0 \leq \int_t^T h_{i,j}(s) ds < 1$  in  $[t_{j,i} - \delta - \frac{\tilde{\delta}}{2}, t_{j,i} + \delta + \frac{\tilde{\delta}}{2}] \setminus B_\delta(t_{j,i})$ ,
- c)  $\int_t^T h_{i,j}(s) ds = 1$  in  $B_\delta(t_{j,i})$ ,
- d)  $\text{supp} \left( \int_t^T h_{i,j}(s) ds \right) \subseteq [t_{j,i} - \delta - \frac{\tilde{\delta}}{2}, t_{j,i} + \delta + \frac{\tilde{\delta}}{2}] \subset\subset I$ , i.e.  $\int_\Omega g_i^j dx \in C_0(\bar{I})$ ,
- e)  $\langle f_i, g_l \rangle_{L^2(\Omega)} = \delta_{i,l}$ ,

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

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

$$\tilde{p}_{j,i} = 0 \text{ in } I \setminus I_{j,i}, \quad \tilde{p}_{j,i} = 1 \text{ in } [t_{j,i} - \delta, t_{j,i} + \delta]. \quad (6.29)$$

Let us define  $h_{i,j} = -\partial_t \tilde{p}_{j,i}$ . For each  $g_i$ ,  $i = 1, \dots, m$ , 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  $L^*(f_{j,i}) = h_{i,j} f_i$  with  $f_{j,i} := f_i \partial_t h_{i,j} - h_{i,j} \Delta f_i$  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  $\vartheta$  such that

$$|c_j^i - c_{j,\vartheta}^i| \leq c \left( \tau^K + h^K + \|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)} \right) \text{ with } j = 1, \dots, m_i \text{ and } i = 1, \dots, m. \quad (6.30)$$

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

$$\begin{aligned} c_j^i - c_{j,\vartheta}^i &= \int_0^T \int_\Omega h_{i,j}(s) ds dD_t(\bar{u}_i - \bar{u}_{\vartheta,i})(t) = \int_0^T h_{i,j}(\bar{u}_i - \bar{u}_{\vartheta,i}) dt \\ &= \sum_{l=1}^m \int_0^T h_{i,j}(\bar{u}_l - \bar{u}_{\vartheta,l}) dt \int_\Omega f_i g_l dx = \int_0^T \int_\Omega -\partial_t g_i^j \left( \sum_{l=1}^m (\bar{u}_l - \bar{u}_{\vartheta,l}) g_l \right) dt dx \end{aligned} \quad (6.31)$$

Thus, it follows

$$\begin{aligned} c_j^i - c_{j,\vartheta}^i &= \int_0^T \int_\Omega -\partial_t g_i^j \left( \sum_{l=1}^m (\bar{u}_l - \bar{u}_{\vartheta,l}) g_l \right) dt dx = \int_0^T \int_\Omega L^*(f_{j,i})(B(\bar{v}, \bar{c}) - B(\bar{v}_\vartheta, \bar{c}_\vartheta)) dt dx \\ &= (L^*(f_{j,i}) - L_\vartheta^*(f_{j,i}), B(\bar{v}, \bar{c}) - B(\bar{v}_\vartheta, \bar{c}_\vartheta))_{L^2(\Omega_T)} + (L_\vartheta^*(f_{j,i}), B(\bar{v}, \bar{c}) - B(\bar{v}_\vartheta, \bar{c}_\vartheta))_{L^2(\Omega_T)}. \end{aligned}$$

By Theorem 3.2 and the boundedness of  $(\bar{v}_\vartheta, \bar{c}_\vartheta)$  we obtain

$$\begin{aligned} &(L^*(f_{j,i}) - L_\vartheta^*(f_{j,i}), B(\bar{v}, \bar{c}) - B(\bar{v}_\vartheta, \bar{c}_\vartheta))_{L^2(\Omega_T)} \\ &\leq \|L^*(f_{j,i}) - L_\vartheta^*(f_{j,i})\|_{L^2(\Omega_T)} \|B(\bar{v}, \bar{c}) - B(\bar{v}_\vartheta, \bar{c}_\vartheta)\|_{L^2(\Omega_T)} = O(\tau^2 + h^2). \end{aligned}$$

Moreover, there holds

$$\begin{aligned}
(L_{\vartheta}^*(f_{j,i}), B(\bar{v}, \bar{c}) - B(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}))_{L^2(\Omega_T)} &= (f_{j,i}, L_{\vartheta}(B(\bar{v}, \bar{c})) - L_{\vartheta}(B(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})))_{L^2(\Omega_T)} \\
&\leq \|f_{j,i}\|_{L^2(\Omega_T)} \|S_{\vartheta}(\bar{v}, \bar{c}) - S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})\|_{L^2(\Omega_T)} \\
&\leq c \|S_{\vartheta}(\bar{v}, \bar{c}) - S(\bar{v}, \bar{c})\|_{L^2(\Omega_T)} + c \|S(\bar{v}, \bar{c}) - S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})\|_{L^2(\Omega_T)} \\
&\leq c(\tau^{\mathbf{K}} + h^{\mathbf{K}} + \|\bar{y} - \bar{y}_{\vartheta}\|_{L^2(\Omega_T)})
\end{aligned}$$

according to Theorem 3.2, and Corollary 3.1.  $\square$

LEMMA 6.16 There holds that

$$|t_{j,1}^i - t_{j,\vartheta}^i| \leq c \left( \tau^{\mathbf{K}} + h^{\mathbf{K}} + \|\bar{y} - \bar{y}_{\vartheta}\|_{L^2(\Omega_T)} \right) \text{ with } j = 1, \dots, m_i \text{ and } i = 1, \dots, m. \quad (6.32)$$

*Proof.* Using that  $z_i(t_{j,i}) = z_{\vartheta,i}(t_{j,\vartheta}^i) = 0$  and  $z_i = \partial_t p_{1,i} \in C^1(\bar{I})$  gives us  $z_i(t_{j,\vartheta}^i) - z_{\vartheta,i}(t_{j,\vartheta}^i) = z_i(t_{j,\vartheta}^i) = z_i(t_{j,i}) + \partial_t z_i(\xi)(t_{j,i} - t_{j,\vartheta}^i)$  for some  $\xi \in B_{\delta}(t_{j,i})$ . In the proof of Lemma 6.12 we have shown that  $|\partial_t z_i(\xi)| > 0$  for all  $\xi \in B_{\delta}(t_{j,i})$  and therefore we have  $\partial_t z_i(\xi) \neq 0$ . Then Lemma 3.2 and Theorem 3.2 imply

$$\begin{aligned}
|t_{j,i} - t_{j,\vartheta}^i| &\leq c \|z_i - z_{\vartheta,i}\|_{L^{\infty}(I)} \leq c \|z - z_{\vartheta}\|_{L^{\infty}(I)^m} \\
&\leq c \|L^*(S(\bar{v}, \bar{c}) - y_d) - L_{\vartheta}^*(S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} \\
&\leq c \|L^*(S(\bar{v}, \bar{c}) - y_d) - L_{\vartheta}^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} + c \|L_{\vartheta}^*(S(\bar{v}, \bar{c}) - S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta}))\|_{C(\bar{I}; L^2(\Omega))} \\
&\leq c \|L^*(S(\bar{v}, \bar{c}) - y_d) - L_{\vartheta}^*(S(\bar{v}, \bar{c}) - y_d)\|_{C(\bar{I}; L^2(\Omega))} + c \|S(\bar{v}, \bar{c}) - S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})\|_{L^1(I; L^2(\Omega))} \\
&\leq c(\tau^{\mathbf{K}} + h^{\mathbf{K}} + \|S(\bar{v}, \bar{c}) - S_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})\|_{L^2(\Omega_T)}). \quad (6.33)
\end{aligned}$$

$\square$

LEMMA 6.17 Let  $\vartheta_0$  be small enough such that

$$\left( (L_{\vartheta_0}(g_i), L_{\vartheta_0}(g_j))_{L^2(\Omega_T)} \right)_{i,j=\{1, \dots, m\}} > 0 \text{ holds.}$$

Then we obtain

$$|c_i - c_{i,\vartheta}| \leq c(\tau^{\mathbf{K}} + h^{\mathbf{K}} + \|\bar{y} - \bar{y}_{\vartheta}\|_{L^2(\Omega_T)}). \quad (6.34)$$

*Proof.* First, define the function  $h(t) = \mathbf{1}_{(t,T]} - \frac{T-t}{T}$ . The optimality conditions of the continuous and discrete problem lead to

$$\begin{aligned}
0 = p_{1,i}(0) &= \sum_{l=1}^m c_l \int_0^T \int_{\Omega} L^*(L(g_l)) g_l \, dx \, dt \\
&\quad + \sum_{l=1}^m \sum_{j=1}^{m_l} c_j^l \int_0^T \int_{\Omega} L^*(L(h(t_{j,l})g_l)) g_l \, dx \, dt + \int_0^T \int_{\Omega} L^*(Q(y_0, y_1) - y_d) g_l \, dx \, dt
\end{aligned}$$

as well as in the discrete case to

$$\begin{aligned}
0 = p_{1,\vartheta,i}(0) &= \sum_{l=1}^m c_{l,\vartheta} \int_0^T \int_{\Omega} L_{\vartheta}^*(L_{\vartheta}(g_l)) g_l \, dx \, dt \\
&\quad + \sum_{l=1}^m \sum_{j=1}^{m_l} c_{j,\vartheta}^l \int_0^T \int_{\Omega} L_{\vartheta}^*(L_{\vartheta}(h(t_{j,\vartheta}^i)g_l)) g_l \, dx \, dt + \int_0^T \int_{\Omega} L_{\vartheta}^*(Q_{\vartheta}(y_0, y_1) - y_d) g_l \, dx \, dt.
\end{aligned}$$

By taking the difference of the last two terms we get

$$\begin{aligned}
\sum_{l=1}^m (c_{l,\vartheta} - c_l) \int_0^T \int_{\Omega} L^*(L(g_l)) g_l \, dx \, dt &= \sum_{l=1}^m c_{l,\vartheta} \int_0^T \int_{\Omega} [L^*(L(g_l)) - L_{\vartheta}^*(L_{\vartheta}(g_l))] g_l \, dx \, dt \\
&\quad + \sum_{l=1}^m \sum_{j=1}^{m_l} (c_j^l - c_{j,\vartheta}^l) \int_0^T \int_{\Omega} L^*(L(h(t_{j,l})g_l)) g_l \, dx \, dt \\
&\quad + \sum_{l=1}^m \sum_{j=1}^{m_l} c_{j,\vartheta}^l \left( \int_0^T \int_{\Omega} [L^*(L(h(t_{j,l})g_l)) - L_{\vartheta}^*(L_{\vartheta}(h(t_{j,\vartheta}^l)g_l))] g_l \, dx \, dt \right) \\
&\quad + \int_0^T \int_{\Omega} [L^*(Q(y_0, y_1) - y_d) - L_{\vartheta}^*(Q_{\vartheta}(y_0, y_1) - y_d)] g_l \, dx \, dt. \quad (6.35)
\end{aligned}$$

For the following we remark that  $\bar{u}_{\vartheta}$  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  $L(g_l) \in C^1(\bar{I}; \mathbb{H}^{\alpha})$  with  $\alpha = 0, 1$  depending on  $\kappa$  according to Theorem 2.2. Thus, with (3.3), Lemma 3.2 and Corollary 3.1 it follows,

$$\begin{aligned}
\sum_{l=1}^m c_{l,\vartheta} \int_0^T \int_{\Omega} [L^*(L(g_l)) - L_{\vartheta}^*(L_{\vartheta}(g_l))] g_l \, dx \, dt &\leq \sum_{l=1}^m c \|L^*(L(g_l)) - L_{\vartheta}^*(L_{\vartheta}(g_l))\|_{L^2(\Omega_T)} \\
&\leq c \sum_{l=1}^m \left( \|L^*(L(g_l)) - L_{\vartheta}^*(L(g_l))\|_{L^2(\Omega_T)} + \|L(g_l) - L_{\vartheta}(g_l)\|_{L^2(\Omega_T)} \right) = O(\tau^{\kappa} + h^{\kappa}).
\end{aligned}$$

By Lemma 6.15 we obtain:

$$\begin{aligned}
\sum_{l=1}^m \sum_{j=1}^{m_l} (c_j^l - c_{j,\vartheta}^l) \int_0^T \int_{\Omega} L^*(L(h(t_{j,l})g_l)) g_l \, dx \, dt \\
\leq c \sum_{l=1}^m \sum_{j=1}^{m_l} |c_j^l - c_{j,\vartheta}^l| \leq c \left( \tau^{\kappa} + h^{\kappa} + \|\bar{y} - \bar{y}_{\vartheta}\|_{L^2(\Omega_T)} \right) \quad (6.36)
\end{aligned}$$

Now we consider the third term on the righthand side of (6.35). The stability of  $L_{\vartheta}^*$ , see Lemma 3.2, imply

$$\begin{aligned}
\sum_{l=1}^m \sum_{j=1}^{m_l} c_{j,\vartheta}^l \left( \int_0^T \int_{\Omega} [L^*(L(h(t_{j,l})g_l)) - L_{\vartheta}^*(L_{\vartheta}(h(t_{j,\vartheta}^l)g_l))] g_l \, dx \, dt \right) \\
\leq c \sum_{l=1}^m \sum_{j=1}^{m_l} \left\| L^*(L(h(t_{j,l})g_l)) - L_{\vartheta}^*(L_{\vartheta}(h(t_{j,\vartheta}^l)g_l)) \right\|_{L^1(\Omega_T)} \\
\leq c \sum_{l=1}^m \sum_{j=1}^{m_l} \left\| L^*(L(h(t_{j,l})g_l)) - L_{\vartheta}^*(L(h(t_{j,l})g_l)) \right\|_{L^1(\Omega_T)} + \left\| L(h(t_{j,l})g_l) - L_{\vartheta}(h(t_{j,\vartheta}^l)g_l) \right\|_{L^1(\Omega_T)} \quad (6.37)
\end{aligned}$$

Again, by Theorem 2.2 we have  $L(h(\tilde{t})g_l) \in C^1(\bar{I}; \mathbb{H}^{\alpha})$  with  $\alpha = 0, 1$  and any  $\tilde{t} \in I$ . Hence, with (3.3) we get

$$\sum_{l=1}^m \sum_{j=1}^{m_l} \|L^*(L(h(t_{j,l})g_l)) - L_{\vartheta}^*(L(h(t_{j,l})g_l))\|_{L^1(\Omega_T)} = O(\tau^{\kappa} + h^{\kappa}).$$

Next we consider the following inequality

$$\begin{aligned}
\|L(h(t_{j,l})g_l) - L_{\vartheta}(h(t_{j,\vartheta}^l)g_l)\|_{L^1(\Omega_T)} \\
\leq \|L(h(t_{j,l})g_l) - L_{\vartheta}(h(t_{j,\vartheta}^l)g_l)\|_{L^1(\Omega_T)} + \|L_{\vartheta}(h(t_{j,l})g_l) - L_{\vartheta}(h(t_{j,\vartheta}^l)g_l)\|_{L^1(\Omega_T)}. \quad (6.38)
\end{aligned}$$

Due to (3.3) and Corollary 3.1, the first term on the right hand side of (6.38) possess the asymptotic rate  $O(\tau^\kappa + h^\kappa)$ . By Lemma 3.2, and Lemma 6.16 we obtain for the second term an estimate in terms of  $c(\tau^\kappa + h^\kappa + \|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)})$ . Finally, we consider the last term in (6.35). We have

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

The first term converges in (6.39) with a rate  $(\tau^\kappa + h^\kappa)$  according to Theorem 3.2 since  $Q(y_0, y_1) - y_d \in C^1(\bar{I}, \mathbb{H}^\alpha)$  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  $(\tau^\kappa + h^\kappa)$  of the last term in (6.39). Thus, we have

$$\sum_{l=1}^m (c_l - c_{l,\vartheta}) \int_0^T \int_\Omega L^*(L(g_l)) g_i \, dx \, dt \leq c(\tau^\kappa + h^\kappa + \|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)}). \quad (6.40)$$

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

$$G(\bar{c} - \bar{c}_\vartheta) = \left( \sum_{l=1}^m (c_l - c_{l,\vartheta}) \int_0^T \int_\Omega L^*(L(g_l)) g_i \, dx \, dt \right)_{i=1}^m.$$

Furthermore, we have

$$\|G(\bar{c} - \bar{c}_\vartheta)\|_{\mathbb{R}^m} \geq \lambda_{\min} \|\bar{c} - \bar{c}_\vartheta\|_{\mathbb{R}^m} \geq c \lambda_{\min} \|\bar{c} - \bar{c}_\vartheta\|_\infty \geq c \lambda_{\min} |c_i - c_{i,\vartheta}| \quad (6.41)$$

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

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

**COROLLARY 6.2** It holds that

$$\|\bar{u} - \bar{u}_\vartheta\|_{L^1(I)} \leq c(\tau^\kappa + h^\kappa + \|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)}).$$

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.

$$\|\bar{u} - \bar{u}_\vartheta\|_{L^1(I)^m} = O(\tau^\kappa + h^\kappa), \quad |c_i - c_{i,\vartheta}| = O(\tau^\kappa + h^\kappa), \quad (6.42)$$

$$|t_{j,i} - t_{j,\vartheta}^i| = O(\tau^\kappa + h^\kappa), \quad |c_j^i - c_{j,\vartheta}^i| = O(\tau^\kappa + h^\kappa) \quad (6.43)$$

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

$$\|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)} = O(\tau^\kappa + h^\kappa). \quad (6.44)$$

*Proof.* In case of  $\kappa = \frac{2}{3}$ , we obtain by Theorem 6.1 that  $\|\bar{y} - \bar{y}_\vartheta\|_{L^2(\Omega_T)} = O(\tau^{\frac{2}{3}} + h^{\frac{2}{3}})$ . 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 assume 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  $\varepsilon > 0$

$$\begin{aligned} \|\bar{y}_\vartheta - \bar{y}\|_{L^2(\Omega_T)} &\leq c\|\bar{y} - \hat{y}_\vartheta\|_{L^2(\Omega_T)} + c\|\bar{u}_\vartheta - \bar{u}\|_{L^1(I)^m}^{\frac{1}{2}} \|\bar{p} - \hat{p}_\vartheta\|_{C(I;L^2(\Omega))}^{\frac{1}{2}} \\ &\leq c(\tau^\kappa + h^\kappa) + c\|\bar{u}_\vartheta - \bar{u}\|_{L^1(I)^m}^{\frac{1}{2}} \|\bar{p} - \hat{p}_\vartheta\|_{C(I;L^2(\Omega))}^{\frac{1}{2}} \\ &\leq c(\tau^\kappa + h^\kappa) + c\varepsilon\|\bar{u}_\vartheta - \bar{u}\|_{L^1(I)^m} + \frac{c}{4\varepsilon} \|\bar{p} - \hat{p}_\vartheta\|_{C(I;L^2(\Omega))} \\ &\leq c(\tau^\kappa + h^\kappa) + c\varepsilon\|\bar{y}_\vartheta - \bar{y}\|_{L^2(\Omega_T)} + \frac{c}{4\varepsilon} \|\bar{p} - \hat{p}_\vartheta\|_{C(I;L^2(\Omega))}. \end{aligned}$$

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

$$\|\bar{y}_\vartheta - \bar{y}\|_{L^2(\Omega_T)} \leq c(\tau^\kappa + h^\kappa) + \|\bar{p} - \hat{p}_\vartheta\|_{C(I;L^2(\Omega))}.$$

Then the a priori estimate (3.3) implies

$$\|\bar{p} - \hat{p}_\vartheta\|_{C(I;L^2(\Omega))} \leq c(h^2 + \tau^2) \|\bar{y} - y_d\|_{C^1(I;H_0^1(\Omega))}.$$

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

**COROLLARY 6.3** For the BV representations of the optimal controls of  $(\tilde{P})$  and  $(\tilde{P}_\vartheta^{\text{semi}})$  hold

$$\|D_t \bar{u}\|_{M(I)} - \|D_t \bar{u}_\vartheta\|_{M(I)} = O(\tau^\kappa + h^\kappa).$$

Furthermore,  $\bar{u}_\vartheta$  converges strictly in  $BV(0, T)$  to  $\bar{u}$  for  $\vartheta \rightarrow 0$  with the convergence rate  $O(\tau^\kappa + h^\kappa)$ .

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

## 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  $(P)$ . 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  $m$  different controls. For this purpose define the map  $\mathcal{U}_\mathcal{A}(\lambda) := \sum_{\ell \in \mathcal{A}} \lambda_\ell \delta_\ell$  for any finite set  $\mathcal{A} \subset I$  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

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

A straightforward proof can show that  $\Psi(\tilde{v}, \tilde{c}) = 0$  implies (5.2) with  $(\tilde{v}, \tilde{c})$  as optimal control of  $(\tilde{P}_\vartheta^{\text{semi}})$ . Furthermore, we have  $J_\vartheta(w_k, c_k) - J_\vartheta(\tilde{v}_\vartheta, \tilde{c}_\vartheta) \leq \Psi(w_k, c_k)$ , where the proof is similar to [Walter, 2016,

**Algorithm 1:** BV-PDAP Algorithm

Input: For  $i = 1, \dots, m$  define  $A_{0,i} \subset I$  with  $|A_{0,i}| < \infty$ ,  $\lambda_{0,i} \in \mathbb{R}^{|A_{0,i}|}$ ,

$w_0 := (w_{0,i})_{i=1}^m := \left( \mathcal{U}_{A_{0,i}}(\lambda_{0,i}) \right)_{i=1}^m \in M(0, T)^m$ ,  $c_0 \in \mathbb{R}^m$ , and  $k = 0$ :

Calculate:

1.  $\hat{t}_i = \arg \max_{t \in \bar{I}} \left| \int_t^T \int_{\Omega} L_{\vartheta}^*(S_{\vartheta}(w_k, c_k) - y_d) g_i \, dx \, ds \right|$   
for  $i = 1, \dots, m$ .

2. Set  $A_{k,i} := \text{supp}(w_{k,i}) \cup \{\hat{t}_i\}$ , for  $i = 1, \dots, m$ ,  $A_k = \bigcup_{i=1}^m A_{k,i}$  and compute

$$(\bar{\lambda}, \bar{c}) = \arg \min_{\lambda \in \mathbb{R}^{|A_k|}, c \in \mathbb{R}^m} \frac{1}{2} \left\| S_{\vartheta} \left( \left( \mathcal{U}_{A_{k,i}}(\lambda_i) \right)_{i=1}^m, c \right) - y_d \right\|_{L^2(\Omega_T)}^2 + \sum_{i=1}^m \alpha_i \|\mathcal{U}_{A_{k,i}}(\lambda_i)\|_{M(0,T)}.$$

3. Define  $w_{k+1} = \left( \mathcal{U}_{A_{k,i}}(\bar{\lambda}_i) \right)_{i=1}^m$ ,  $c_{k+1} = \bar{c}$ ; set  $k = k + 1$  and return to 1.

Lemma 12]. In the numerical experiment, that is considered below, we replace  $J_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})$  with  $J(\bar{v}, \bar{c})$  and use this for the abort criterion for the while-loop. In this experiment we will see that  $J(\bar{v}, \bar{c})$  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  $J(\bar{v}, \bar{c})$ , 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  $m = 1$ :

$$\Omega := (-1, 1)^d, \quad I := (0, T), \quad g(x) := \prod_{i=1}^d \cos\left(\frac{\pi}{2} x_i\right) \in C_0^\infty(\Omega), \text{ and}$$

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

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

$$\bar{u} := \int_0^t dD_t \bar{u} - \frac{1}{T} \int_0^T \int_{[0,s]} dD_t \bar{u} ds + \bar{c} \text{ with } \bar{c} \in \mathbb{R}.$$

Note that

$$D_t \bar{u} = \sum_{i=0}^{\ell-1} \text{sign} \left( \sin \left( \pi \frac{2i+1}{2} \right) \right) \delta_{\frac{T(1+2i)}{2\ell}} \quad (7.2)$$

$$\text{with } \frac{1}{T} \int_0^T \int_0^s dD_t \bar{u} ds = \sum_{i=0}^{\ell-1} \text{sign} \left( \sin \left( \pi \frac{2i+1}{2} \right) \right) \left( \frac{2(\ell-i)-1}{2\ell} \right) \quad (7.3)$$

holds. With this we can now define the following desired state as  $y_d = S(\bar{v}, \bar{c}) - (\partial_t - \Delta)\bar{\varphi}$ . 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  $\bar{u}$  where  $p_1(t)$  is of the form  $-\int_t^T \int_\Omega \bar{\varphi}(x, s) \cdot g(x) dx ds = \beta \frac{2T}{3\pi\ell} \left( \sin\left(\frac{\pi\ell t}{T}\right) \right)^3$ . By several calculation steps can be shown

$$\tilde{J}(D_i \bar{u}, \bar{c}) = \frac{T}{8} (\beta^2 a_1^2 + a_2^2) + \alpha\ell, \text{ with } a_1 := \frac{d\pi^2}{4} - \frac{5\ell^2\pi^2}{T^2}, a_2 := 4\frac{\ell^2\pi^2}{T^2}\beta. \quad (7.4)$$

**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  $i = 1, \dots, m$ , considered in step 1, can lie somewhere between two time nodes in  $\bar{I}$  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:

$$B(w_k, c_k) = \left( \sum_{\bar{i} \in \text{supp}(w_k)} \bar{\lambda}_{\bar{i}} \mathbf{1}_{[\bar{i}, T]}(t) - \bar{\lambda}_{\bar{i}} (1 - \frac{\bar{i}}{T}) + c_k \right) g(x) =: f_k(t)g(x). \quad (7.5)$$

$B(w_k, c_k)$  acts as forcing function in  $L_\vartheta$  and thus has to be projected on  $S_\tau \otimes S_h$  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  $S_h$  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  $S_\vartheta(w_k, c_k)$  in turns out, that we only need an additional discretization step for the projection of  $g$  on the finite element space  $S_h$ .

Another difficulty is the discretization of the desired state  $y_d = S(\bar{v}, \bar{c}) - (\partial_t - \Delta)\varphi$ . At first, let us approximate  $S(\bar{v}, \bar{c})$  by  $S_\vartheta(\bar{v}, \bar{c})$ , which we denote by  $S_\vartheta^{(c,v)}$ , where  $B(\bar{v}, \bar{c})$  is discretized as we discussed before. For  $(\partial_t - \Delta)\varphi$  we are able to obtain an explicit formula which we projected on  $S_\tau \otimes S_h$  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  $\varphi_\vartheta^D$ . Hence, we can use directly the full discretization scheme in [Zlotnik, 1994, p.165] for  $L_\vartheta^*(S_\vartheta^{(c,v)} - \varphi_\vartheta^D)$ , where  $(y_0, y_1)$  is projected on  $S_h$  as zero function. Using the projected  $g$  function on  $S_h$ , defined by  $g^h$ , and the space-time function  $L_\vartheta^*(S_\vartheta(w_k, c_k) - y_d)$ , we can explicitly calculate  $\int_\Omega L_\vartheta^*(S_\vartheta(w_k, c_k) - y_d)g^h dx$ , which is a element of  $S_\tau$  with  $S_\vartheta(w_k, c_k) - y_d$  discretized as above. The last time integral in step 1 of Algorithm 1 can now be calculated explicitly, for  $t \in \bar{I}$ , by using the following formula:

$$\int_0^t \sum_{\ell=0}^M v_\ell e_\ell(s) ds = \sum_{\ell=1}^{\iota_\tau(t)} \tau \left( \frac{v_\ell + v_{\ell-1}}{2} \right) + v_{\iota_\tau(t)} \frac{t_{\iota_\tau(t)}^2 - t^2 + 2(t_{\iota_\tau(t)+1}t - t_{\iota_\tau(t)+1}t_{\iota_\tau(t)})}{2\tau} \\ + v_{\iota_\tau(t)+1} \frac{t^2 - t_{\iota_\tau(t)}^2 + 2(t_{\iota_\tau(t)}^2 - t_{\iota_\tau(t)}t)}{2\tau}$$

where  $v_\ell \in \mathbb{R}$  and  $\sum_{\ell=0}^M v_\ell e_\ell(t)$  is an arbitrary element of  $S_\tau$ . If  $t > 0$ ,  $\iota_\tau(t)$  is returning the highest index of all strictly smaller time nodes in  $\bar{w}^\tau$  else it returns the index 0. Thus, we discretized  $\int_t^T \int_\Omega L_\vartheta^*(S_\vartheta(w_k, c_k) - y_d)g dx ds$  into a  $C^1(0, T)$  function, which we define by  $f_\vartheta(t)$ .

Step 1 is then executed as follows: First we determine the time derivative of  $f_\vartheta$ . Since the derivative is an element of  $S_\tau$ , we check where coefficients of the hat functions change their sign. At these

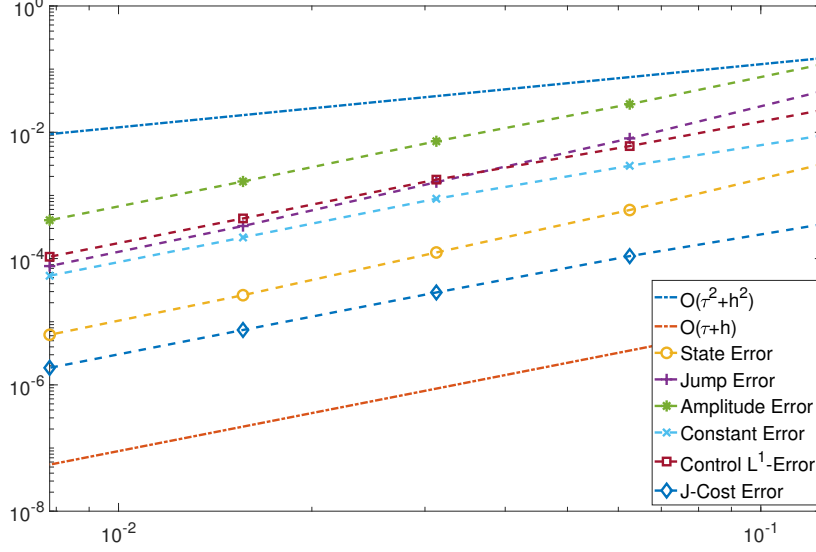


FIG. 1

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  $\partial_t f_{\vartheta}(t) = \sum_{\ell=0}^M f_{\ell} e_{\ell}(t)$  for some  $f_{\ell} \in \mathbb{R}$ . Let us observe a sign change in  $i \in \{1, \dots, M\}$ . Hence,  $\partial_t f_{\vartheta}(t)$  has a root in  $\hat{t} = \frac{f_i t_{i-1} - f_{i-1} t_i}{f_i - f_{i-1}}$ , 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  $\hat{t}$  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  $\gamma/2 \|(\lambda_{\ell})_{\ell \in A_k}\|_{\mathbb{R}^{|A_k|}}^2$  in the minimization problem of step 2 and use a prox-operator approach to obtain an 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:

$$d = 2, \quad I = (0, 2), \quad \ell = 2, \quad \alpha = 2.3 \cdot 10^{-4}, \quad \bar{c} = 0. \quad (7.6)$$

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  $|\tilde{J}(\bar{v}, \bar{c}) - J_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})|$  which are defined as "J-Cost Error". In particular we define in Figure 1 the "State Error" by  $\|\bar{y} - \bar{y}_{\vartheta}\|_{L^2(\Omega_T)}$ , the "Jump Error" by  $|t_{j,1} - t_{j,\vartheta}^1|$ , the "Amplitude Error" by  $|c_j^1 - c_{j,\vartheta}^1|$ , the "Constant Error" by  $|c_1 - c_{1,\vartheta}|$ , and the "Control  $L^1$ -Error" by  $\|\bar{u} - \bar{u}_{\vartheta}\|_{L^1(I)}$ , for  $j = 1, 2$  with respect to (7.2) and (7.6). We used (7.4) for  $J(\bar{v}, \bar{c})$  and considered  $(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})$  as solution of the full discretized BV-PDAP algorithm. Furthermore,  $J_{\vartheta}(\bar{v}, \bar{c})$  is calculated by all approximation steps used for  $S_{\vartheta}(\bar{v}, \bar{c})$ ,  $y_d$ , 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  $\vartheta$ , i.e.  $\tau = 2^{-k}$  and  $h = 2 \cdot \sqrt{2} \cdot 2^{-k}$  with  $k = 3, 4, \dots, 7$ . For each discretization level  $\vartheta$  we initialized the BV-PDAP algorithm with  $(w_0, c_0) = (2 \cdot \delta_{0,36}, 0)$ . The semi-smooth Newton algorithm is initialized with the data of the current iterate  $(w_k, c_k)$ , where the initial  $\lambda_f$  is defined by 1. The state error is calculated with respect to a reference state  $S_{\vartheta}(\bar{v}, \bar{c})$  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  $B(\bar{v}, \bar{c})$  in the reference state is discretized as we discussed above. Let us note that we have used the cost difference  $|\tilde{J}(\bar{v}, \bar{c}) - J_{\vartheta}(\bar{v}_{\vartheta}, \bar{c}_{\vartheta})|$  as a termination criterion, in each discretization level  $\vartheta$ , for the BV-PDAP algorithm. In detail, for each finer  $\vartheta$ , we stopped Algorithm 1 if the cost difference  $|\tilde{J}(\bar{v}, \bar{c}) - J_{\vartheta}(w_k, c_k)|$  is quadratically decreasing compared to the cost difference of the last coarser  $\vartheta$ .

In Algorithm 1 step 1, we additionally considered in the numerical experiment all stationary points of  $\int_t^T \int_{\Omega} L_{\vartheta}^*(S_{\vartheta}(w_k, c_k) - y_d) g_i dx ds$  and thus obtained in some iteration steps more than one new point  $\hat{t}_i$ , for  $i = 1, \dots, m$ . It turned out that this additional step accelerated the algorithm in finding an approximation of our solution. Numerically, we observed that Diracs were clustering nearby. Therefore, in an additional step to Algorithm 1 step 3, we merged all Diracs which were 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  $\delta_t$ , in  $\mathcal{U}_{A_{k,i}}(\bar{\lambda})$ , is added to the corresponding amplitude of  $\delta_{\hat{t}_i}$ , in  $\mathcal{U}_{A_{k,i}}(\bar{\lambda})$ , for  $i = 1, \dots, m$ , where  $\delta_t$  is erased from  $\mathcal{U}_{A_{k,i}}(\bar{\lambda})$ . Furthermore, we observed that it is reasonable, after step 2, to remove all Diracs  $\delta_{t_i}$ ,  $t \in A_{k,i}$ , for  $i = 1, \dots, m$ , with small amplitudes  $\lambda_{t_i}$  related to  $\delta_{t_i}$ , before defining the next iterated  $w_{k+1}$ . For example, we erased all Diracs with amplitude less than  $10^{-5}$ .

#### REFERENCES

- L. Ambrosio, N. Fusco, and D. Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs, The Clarendon Press Oxford University Press, New York, 2000.
- G. A. Anastassiou. *Intelligent Computations: Abstract Fractional Calculus, Inequalities, Approximations*. Studies in Computational Intelligence. Springer International Publishing, 2017. ISBN 9783319669366. URL <https://books.google.de/books?id=xsEzDwAAQBAJ>.
- E. Casas and K. Kunisch. Optimal control of semilinear elliptic equations in measure spaces. *SIAM Journal on Control and Optimization*, 52(1):339–364, 2014.
- E. Casas and K. Kunisch. Analysis of optimal control problems of semilinear elliptic equations by bv-functions. *Set-Valued and Variational Analysis*, 1-25, 2017.
- E. Casas and F. Tröltzsch. A general theorem on error estimates with application to a quasilinear elliptic optimal control problem. *Computational Optimization and Applications*, 53(1):173–206, 2012.
- E. Casas, K. Kunisch, and C. Pola. Some applications of BV functions in optimal control and calculus of variations. *ESAIM: Proceedings*, 4:181–198, 1998.
- E. Casas, K. Kunisch, and C. Pola. Regularization by functions of bounded variation and applications to image enhancement. *Appl. Math. and Optimization*, 40:229–258, 1999.
- E. Casas, F. Kruse, and K. Kunisch. Optimal control of semilinear parabolic equations by bv-functions. *SIAM J. Control Optim.*, 55(3), 1752–1788, 2017.

- C. Clason and K. Kunisch. A duality-based approach to elliptic control problems in non-reflexive Banach spaces. *ESAIM Control Optim. Calc. Var.*, 17:243–266, 2011.
- I. Ekeland and R. Témam. *Convex Analysis and Variational Problems, Classics in Appl. Math.* SIAM, Philadelphia, PA, english ed., 1999.
- S. Engel and K. Kunisch. Optimal control of the linear wave equation by time-depending bv-controls: A semi-smooth newton approach. *ArXiv manuscript e-prints*, 2018.
- D. Hafemeyer, F. Mannel, I. Neitzel, and B. Vexler. Finite element error estimates for one-dimensional elliptic optimal control by bv functions. *ArXiv e-prints 1902.05893v2*, and *Math. Control Relat. Fields*, accepted, 2019.
- O.A. Ladyjenskaya. Boundary value problems of mathematical physics. *Nauka, Moscow*, 1973.
- K. Pieper and B. Vexler. A priori error analysis for discretization of sparse elliptic optimal control problems in measure space. *SIAM Journal on Control and Optimization* 51(4),pp. 2788- 2808, 2013.
- K. Pieper and D. Walter. Linear convergence of accelerated conditional gradient algorithms in spaces of measures. *ArXiv e-prints 1904.09218v1*, 2019.
- P. Trautmann, B. Vexler, and A. Zlotnik. Finite element error analysis for measure-valued optimal control problems governed by a 1d wave equation with variable coefficients. *Mathematical Control and Related Fields*, 8:411, 2018a. ISSN 2156-8472.
- P. Trautmann, B. Vexler, and A. A. Zlotnik. Finite element error analysis for measure-valued optimal control problems governed by a 1d wave equation with variable coefficients. *Mathematical Control and Related Fields*, 8(2), pp. 411-449, 2018b.
- D. Walter. Gradient based optimization algorithms for sensor placement problems with sparsity. Master's thesis, Technical University of Munich, 2016.
- A. A. Zlotnik. Convergence rate estimates of finite-element methods for second-order hyperbolic equations. numerical methods and applications, p.153 et.seq. *Guri I. Marchuk, CRC Press*, 1994.