OPTIMAL CONTROL OF SEMILINEAR PARABOLIC EQUATIONS BY BV-FUNCTIONS *

EDUARDO CASAS[†], FLORIAN KRUSE[‡], AND KARL KUNISCH[‡]

Abstract. Optimal control problems for semilinear parabolic equations with control costs involving the total bounded variation seminorm are analyzed. This choice of control cost favors optimal controls which are piecewise constant and it penalizes the number of jumps. It is an appropriate choice if a simple structure of the optimal controls is desired, which, however, is still sufficiently flexible so that good tracking properties can be maintained. Well-posedness of the optimal controls, necessary and sufficient optimality conditions, and sparsity properties of the derivatives are obtained. Convergence of a finite element approximation is analyzed and numerical examples illustrating structural properties of the optimal controls are provided.

AMS subject classifications. 35K58, 49J20, 49J52, 49K20, 49M25

Key words. Bounded variation functions, optimal control, sparsity, semilinear parabolic equations, first and second order optimality conditions, discontinuous Galerkin approximations

1. Introduction.

This paper is dedicated to the analysis of the optimal control problem

$$\text{(P)} \qquad \min_{u\in BV(0,T)^m} J(u) = \frac{1}{2}\|y_u - y_d\|^2_{L^2(Q)} + \sum_{j=1}^m \Big(\alpha_j \|u_j'\|_{\mathcal{M}(0,T)} + \frac{\beta_j}{2}\big(\int_0^T u_j(t)\,dt\big)^2\Big),$$

where $u = (u_j)_{j=1}^m$ and y_u is the solution to the parabolic state equation

$$\left\{\begin{array}{rcll} \dfrac{\partial y}{\partial t}(x,t) - \Delta y(x,t) + f(x,t,y(x,t)) &=& \displaystyle\sum_{j=1}^m u_j g_j & \text{in } Q = \Omega\times(0,T),\\ y(x,t) &=& 0 & \text{on } \Sigma = \Gamma\times(0,T),\\ y(x,0) &=& y_0(x) & \text{in } \Omega. \end{array}\right. \tag{1.1}$$

Here, we assume that Ω is a bounded domain in $\mathbb{R}^n$, $1 \le n \le 3$, with a Lipschitz boundary Γ, and $y_0 \in L^\infty(\Omega)$. $BV(0,T)$ denotes the space of bounded variation functions defined in $(0,T)$, with $0 < T < \infty$ given. The controllers in (P) are supposed to be separable functions with respect to fixed spatial shape functions g_j and free temporal amplitudes u_j. The specificity in (P) is given by the choice of the control norm as the BV-seminorm $\|u_j'\|_{\mathcal{M}(0,T)}$. It enhances that the optimal controls are piecewise constant in time and that the number of jumps is penalized. The weights in (P) are assumed to satisfy $\alpha_j > 0$ and $\beta_j \ge 0$. Thus the goal of the optimal control problem (P) is to achieve a simple control strategy while simultaneously being as close to the target y_d as possible. The appearance of the mean $\int_0^T u_j(t)\,dt$ in the cost is related to the kernel of the BV-seminorm. For linear and certain classes of nonlinear functions f the choice $\beta_j = 0$ is admissible, while for more severe nonlinearities we have chosen the option $\beta_j > 0$ to guarantee existence of a solution to (P).

*The first author was supported by Spanish Ministerio de Economía y Competitividad under project MTM2014-57531-P. The third was supported by the Austrian Science Fund (FWF) under grant SFB F32 (SFB "Mathematical Optimization and Applications in Biomedical Sciences")

[†]Departmento de Matemática Aplicada y Ciencias de la Computación, E.T.S.I. Industriales y de Telecomunicación, Universidad de Cantabria, 39005 Santander, Spain (eduardo.casas@unican.es).

[‡]Institute for Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, A-8010 Graz, Austria (florian.kruse@uni-graz.at, karl.kunisch@uni-graz.at).

The choice of the control costs related to BV-norms or BV-seminorms has not received much attention in the literature. However, let us mention [9] where the effect of L^2-, H^1-, measure-valued and BV-valued control costs on the qualitative behavior of the optimal control was pointed out and compared. In [12] the use of BV-costs was investigated further for the case of linear elliptic equations. BV-seminorm control costs are also employed in [5], where the control appears as coefficient in the p-Laplace equation.

Let us also compare (P) with the efforts that have been made for studying optimal control problems with sparsity constraints. These formulations involve either measure-valued norms of the control or L^1-functionals combined with pointwise constraints on the control. We cite [4, 13] from among the many results which are now already available. Thus the use of the BV-seminorm can also be understood as a sparsity constraint for the first derivative, which in our case is the temporal derivative.

Let us briefly outline the following sections. Section 2 contains a precise problem statement, the analysis of the state equation, and the differentiability properties of the cost functional. The analysis of the optimal control problem, sparsity properties of the optimal controls as well as second order necessary and sufficient optimality conditions are contained in Section 3. Section 4 is devoted to a finite element approximation of the control problem and its well-posedness. A convergence analysis of this approximation scheme is provided in Section 5. Numerical results illustrating that the desired behavior of the optimal controls can actually be observed numerically are presented in Section 6. To obtain these results convexity properties of the cost functional are exploited.

2. Assumptions and First Consequences. We recall that a function $u \in L^1(0,T)$ is a function of bounded variation if its distributional derivative u' belongs to the Banach space of real and regular Borel measures $\mathcal{M}(0,T)$. Given a measure $\mu \in \mathcal{M}(0,T)$, its norm is given by

$$\|\mu\|_{\mathcal{M}(0,T)} = \sup\{\int_0^T z\,d\mu : z \in C_0(0,T) \text{ and } \|z\|_{C_0(0,T)} \le 1\} = |\mu|(0,T),$$

where $C_0(0,T)$ denotes the Banach space of continuous functions $z : [0,T] \longrightarrow \mathbb{R}$ such that $z(0) = z(T) = 0$, and $|\mu|$ is the total variation measure associated with μ. On $BV(0,T)$ we consider the usual norm

$$\|u\|_{BV(0,T)} = \|u\|_{L^1(0,T)} + \|u'\|_{\mathcal{M}(0,T)},$$

that makes $BV(0,T)$ a Banach space; see [1, Chapter 3] or [11, Chapter 1] for details. In the sequel we will denote

$$a_u = \frac{1}{T}\int_0^T u(t)\,dt \quad \text{and} \quad \hat{u} = u - a_u \quad \text{for every } u \in BV(0,T).$$

By using [1, Theorem 3.44] it is easy to deduce that there exists a constant C_T such that

$$\|u\| := |a_u| + \|u'\|_{\mathcal{M}(0,T)} \le \max(1,T)\|u\|_{BV(0,T)} \le C_T\|u\|. \tag{2.1}$$

In addition, we mention that $BV(0,T)$ is the dual space of a separable Banach space. Therefore every bounded sequence $\{u_k\}_{k=1}^\infty$ in $BV(0,T)$ has a subsequence converging weakly* to some $u \in BV(0,T)$. The weak* convergence $u_k \stackrel{*}{\rightharpoonup} u$ implies that $u_k \to u$

strongly in $L^1(0,T)$ and $u'_k \stackrel{*}{\rightharpoonup} u'$ in $\mathcal{M}(0,T)$; see [1, pages 124-125]. We will also use that $BV(0,T)$ is continuously embedded in $L^\infty(0,T)$ and compactly embedded in $L^p(0,T)$ for every $p < +\infty$; see [1, Corollary 3.49]. From this property we deduce that the convergence $u_k \stackrel{*}{\rightharpoonup} u$ in $BV(0,T)$ implies that $u_k \to u$ strongly in every $L^p(0,T)$ for all $p < +\infty$.

In the functional J, y_d is given in $L^{\hat p}(Q)$, where $\hat p > 1 + \frac{n}{2}$ if $n > 1$, and $\hat p \geq 2$ if $n = 1$, $\alpha_j > 0$ and $\beta_j \geq 0$ for $1 \leq j \leq m$. Further, the functions $\{g_j\}_{j=1}^m \subset L^\infty(\Omega)\setminus\{0\}$ have pairwise disjoint supports $\omega_j = \operatorname{supp} g_j$. Finally, we assume that $f : Q \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Borel function, of class C^2 with respect to the last variable, and satisfies for almost all $(x,t) \in Q$

$$f(\cdot,\cdot,0) \in L^{\hat p}(Q), \tag{2.2}$$

$$\frac{\partial f}{\partial y}(x,t,y) \geq 0 \quad \forall y \in \mathbb{R}, \tag{2.3}$$

$$\forall M > 0\, \exists C_M : \left|\frac{\partial f}{\partial y}(x,t,y)\right| + \left|\frac{\partial^2 f}{\partial y^2}(x,t,y)\right| \leq C_M \quad \forall |y| \leq M, \tag{2.4}$$

$$\left\{\begin{array}{l} \forall M > 0 \text{ and } \forall \rho > 0\, \exists \varepsilon > 0 \text{ such that} \\ \left|\dfrac{\partial^2 f}{\partial y^2}(x,t,y_2) - \dfrac{\partial^2 f}{\partial y^2}(x,t,y_1)\right| \leq \rho \text{ if } |y_2 - y_1| < \varepsilon \text{ and } |y_1|, |y_2| \leq M. \end{array}\right. \tag{2.5}$$

Let us observe that if f is an affine function, $f(x,t,y) = c_0(x,t)y + d_0(x,t)$, then (2.2)-(2.5) hold if $c_0 \geq 0$ in Q, $c_0 \in L^\infty(Q)$, and $d_0 \in L^{\hat p}(Q)$.

By using these assumptions, the following theorem can be proved in a standard way; see, for instance, [2] or [22, Theorem 5.5].

Proposition 2.1. *For every $u \in L^p(0,T)^m$, with $p > 1$, the state equation* (1.1) *has a unique solution $y_u \in L^\infty(Q) \cap L^2(0,T;H_0^1(\Omega))$. In addition, for every $M > 0$ there exists a constant K_M such that*

$$\|y_u\|_{L^\infty(Q)} + \|y_u\|_{L^2(0,T;H_0^1(\Omega))} \leq K_M \quad \forall u \in L^p(0,T)^m : \|u\|_{L^p(0,T)^m} \leq M. \tag{2.6}$$

In the sequel we will denote $Y = L^\infty(Q) \cap L^2(0,T;H_0^1(\Omega))$ and $S : L^p(0,T)^m \longrightarrow Y$ the mapping associating to each control u the corresponding state $S(u) = y_u$, with $p > 1$. By the implicit function theorem, we deduce in the classical way the following result, [7, Theorem 5.1].

Proposition 2.2. *The mapping $S : L^p(Q)^m \longrightarrow Y$ is of class C^2. For all elements u, v and w of $L^p(0,T)^m$, the functions $z_v = S'(u)v$ and $z_{vw} = S''(u)(v,w)$ are the solutions of the problems*

$$\left\{\begin{array}{rl} \dfrac{\partial z}{\partial t} - \Delta z + \dfrac{\partial f}{\partial y}(x,t,y_u)z = \displaystyle\sum_{j=1}^m \alpha_j v_j g_j & \text{in } Q, \\ z = 0 & \text{on } \Sigma, \\ z(x,0) = 0 & \text{in } \Omega, \end{array}\right. \tag{2.7}$$

and

$$\left\{\begin{array}{rl} \dfrac{\partial z}{\partial t} - \Delta z + \dfrac{\partial f}{\partial y}(x,t,y_u)z + \dfrac{\partial^2 f}{\partial y^2}(x,t,y_u)z_v z_w = 0 & \text{in } Q, \\ z = 0 & \text{on } \Sigma, \\ z(x,0) = 0 & \text{in } \Omega, \end{array}\right. \tag{2.8}$$

respectively.

Next we analyze the differentiability of the cost functional. In J we separate the smooth and the convex parts $J(u) = F(u) + G(u)$ with

$$F(u) = \frac{1}{2}\|y_u - y_d\|^2_{L^2(Q)} + \sum_{j=1}^m \frac{\beta_j}{2}\Big(\int_0^T u_j(t)\,dt\Big)^2 \quad\text{and}\quad G(u) = \sum_{j=1}^m \alpha_j g(u_j'),$$

where $g : \mathcal{M}(0,T) \longrightarrow \mathbb{R}$ is given by $g(\mu) = \|\mu\|_{\mathcal{M}(0,T)}$. From Proposition 2.2 and the chain rule the following proposition can be obtained.

PROPOSITION 2.3. *The functional $F : L^p(0,T)^m \longrightarrow \mathbb{R}$, with $p > 1$, is of class C^2. The derivatives of F are given by*

$$F'(u)v = \sum_{j=1}^m \int_0^T \Big(\int_{\omega_j} \varphi_u(x,t) g_j(x)\,dx + \beta_j \int_0^T u_j(t)\,dt\Big) v_j(t)\,dt, \tag{2.9}$$

and

$$F''(u)(v,w) = \int_Q \big(1 - \varphi_u \frac{\partial^2 f}{\partial y^2}(x,t,y_u)\big) z_v z_w \,dx\,dt + \sum_{j=1}^m \beta_j \int_0^T v_j\,dt \int_0^T w_j\,dt \tag{2.10}$$

with $z_v = S'(u)v$, $z_w = S'(u)w$, and $\varphi_u \in Y \cap C(\bar{Q})$ is the adjoint state which satisfies

$$\begin{cases} -\dfrac{\partial \varphi_u}{\partial t} - \Delta\varphi_u + \dfrac{\partial f}{\partial y}(x,t,y_u)\varphi_u = y_u - y_d & \text{in } Q,\\ \varphi_u = 0 & \text{on } \Sigma,\\ \varphi_u(T) = 0 & \text{in } \Omega. \end{cases} \tag{2.11}$$

The $L^\infty(Q)$ regularity of φ_u follows from the assumptions on y_d and the fact that $y_u \in L^\infty(Q)$. For the continuity of φ_u in $\bar{Q}$ it is enough to use that the terminal and boundary conditions are zero.

Since $BV(0,T)^m$ is continuously embedded in $L^\infty(0,T)^m$, the mapping F is well defined on $BV(0,T)^m$ and it is of class C^2.

Concerning the functional $g : \mathcal{M}(0,T) \longrightarrow \mathbb{R}$, $g(\mu) = \|\mu\|_{\mathcal{M}(0,T)}$, we note that it is Lipschitz continuous and convex. Hence, it has a subdifferential and a directional derivative, which are denoted by $\partial g(\mu)$ and $g'(\mu;\nu)$, respectively. The following propositions give some properties of $\partial g(\mu)$ and provide an expression for $g'(\mu;\nu)$.

PROPOSITION 2.4 ([6, Proposition 3.2]). *If $\lambda \in \partial g(\mu)$ and $\lambda \in C_0(0,T)$, then we have $\|\lambda\|_{C_0(0,T)} \le 1$. Moreover, if $\mu \ne 0$, the following properties hold*

1. $\|\lambda\|_{C_0(0,T)} = 1$ *and* $\int_0^T \lambda\,d\mu = \|\mu\|_{\mathcal{M}(0,T)}$.
2. *Taking the Jordan decomposition $\mu = \mu^+ - \mu^-$, we have*

$$\text{supp}(\mu^+) \subset \{t \in (0,T) : \lambda(t) = +1\},$$
$$\text{supp}(\mu^-) \subset \{t \in (0,T) : \lambda(t) = -1\}.$$

Before considering the directional derivative $g'(\mu;\nu)$, let us introduce some notation. Given two measures $\mu, \nu \in \mathcal{M}(0,T)$, we consider the Lebesgue decomposition of $\nu = \nu_a + \nu_s$ with respect to $|\mu|$, where ν_a is the absolutely continuous part of ν with respect to $|\mu|$, and ν_s is the singular part. Now, we take the Radon-Nikodym derivative of ν_a with respect to $|\mu|$, $d\nu_a = h_\nu d|\mu|$. Then we have

$$\|\nu\|_{\mathcal{M}(0,T)} = \|\nu_a\|_{\mathcal{M}(0,T)} + \|\nu_s\|_{\mathcal{M}(0,T)} = \int_0^T |h_\nu|\,d|\mu| + \|\nu_s\|_{\mathcal{M}(0,T)}. \tag{2.12}$$

In particular, it is obvious that μ is absolutely continuous with respect to $|\mu|$. Consequently we can express $d\mu = hd|\mu|$, where h is measurable with respect to $|\mu|$ and $|h(x)| = 1$ for all $x \in (0,T)$, $d\mu^+ = h^+ d|\mu|$ and $d\mu^- = h^- d|\mu|$, where $\mu = \mu^+ - \mu^-$ is the Jordan decomposition of μ. See, for instance, [19, Chapter 6] for details.

PROPOSITION 2.5 ([6, Proposition 3.3]). *Let $\mu, \nu \in \mathcal{M}(0,T)$, then*

$$g'(\mu;\nu) = \int_0^T h_\nu \, d\mu + \|\nu_s\|_{\mathcal{M}(0,T)}. \tag{2.13}$$

Now, we analyze the mapping G. To this end, let us introduce the operator $D_t : BV(0,T) \longrightarrow \mathcal{M}(0,T)$ by $D_t u = u'$. Its adjoint operator is defined by

$$D_t^* : \mathcal{M}(0,T)^* \longrightarrow BV(0,T)^*, \quad \langle D_t^* \lambda, u\rangle_{BV(0,T)^*,BV(0,T)} = \langle \lambda, u'\rangle_{\mathcal{M}(0,T)^*,\mathcal{M}(0,T)}.$$

PROPOSITION 2.6. *The following identities hold $\forall u \in BV(0,T)$*

$$\partial g(u) = D_t^* \partial g(u_j'), \tag{2.14}$$

$$(g \circ D_t)'(u;v) = \int_0^T h_{v'} \, du' + \|v_s'\|_{\mathcal{M}(0,T)}, \tag{2.15}$$

where $dv' = h_{v'} d|u'| + dv_s'$ is the Lebesgue decomposition of v' with respect to $|\mu'|$.

Proof. Since $g : \mathcal{M}(0,T) \longrightarrow \mathbb{R}$ is convex and continuous and $D_t : BV(0,T) \longrightarrow \mathcal{M}(0,T)$ is a linear and continuous mapping, we can apply the chain rule [10, Chapter I, Proposition 5.7] to deduce that $\partial(g \circ D_t)(u_j) = D_t^* \partial g(u_j')$, which immediately leads to (2.14).

To verify (2.15) it is enough to observe that

$$(g \circ D_t)'(u;v) = g'(u';v')$$

and to apply (2.13). ∎

3. Analysis of the Optimal Control Problem (P)**.** This section is devoted to the proof of the existence of at least one solution of (P) and to the optimality conditions and their consequences.

THEOREM 3.1. *Let us assume that one of the following assumptions hold.*

1. *$\beta_j > 0$ for every $1 \le j \le m$.*
2. *There exist $q \in [1,2)$ and $C > 0$ such that*

$$\frac{\partial f}{\partial y}(x,t,y) \le C(1 + |y|^q) \quad \text{for a.a. } (x,t) \in Q. \tag{3.1}$$

Then, problem (P) *has at least one solution. Moreover, if f is affine with respect to y, the solution is unique.*

Let us observe that condition (3.1) is satisfied in the case of affine functions with respect to y.

Proof. Let $\{u_k\}_{k=1}^\infty \subset BV(0,T)^m$ be a minimizing sequence. We prove that this sequence is bounded in $BV(0,T)^m$. As introduced in §2, we consider the decomposition $u_k = a_k + \hat{u}_k$, where $a_k = (a_{k,1}, \dots, a_{k,m})$, $\hat{u}_k = (\hat{u}_{k,1}, \dots, \hat{u}_{k,m})$ and

$$a_k = \frac{1}{T}\int_0^T u_k(t)\, dt \quad \text{and} \quad \hat{u}_k = u_k - a_k.$$

Since

$$\sum_{j=1}^m \big(\alpha_j\|\hat u'_{k,j}\|_{\mathcal{M}(0,T)} + \frac{\beta_j}{2}a_{kj}^2\big)$$
$$= \sum_{j=1}^m \big(\alpha_j\|u'_{kj}\|_{\mathcal{M}(0,T)} + \frac{\beta_j}{2}\big(\int_0^T u_{kj}(t)\,dt\big)^2\big) \le J(u_k) \le J(0) < +\infty,$$

taking into account (2.1), we deduce that $\{\hat u_k\}_{k=1}^\infty$ is bounded in $BV(0,T)^m$. Now we prove the boundedness of $\{a_k\}_{k=1}^\infty$ in $\mathbb{R}^m$. This boundedness is obvious on the first assumption. Otherwise, let us denote by y_k and $\hat y_k$ the solutions (1.1) associated to the controls u_k and $\hat u_k$, respectively. From the inequalities

$$\frac{1}{2}\|y_k - y_d\|^2_{L^2(Q)} \le J(u_k) \le J(0) < +\infty,$$

we get the boundedness of $\{y_k\}_{k=1}^\infty$ in $L^2(Q)$. Moreover, the boundedness of $\{\hat u_k\}_{k=1}^\infty$ in $BV(0,T)^m$ and (2.6) we obtain that $\{\hat y_k\}_{k=1}^\infty$ is also bounded in $L^2(Q)$. Now, we define $z_k = y_k - \hat y_k$, which produces a bounded sequence in $L^2(Q)$ as well. Subtracting the equations satisfied by y_k and $\hat y_k$ and using the mean value theorem we infer that

$$\begin{cases} \dfrac{\partial z_k}{\partial t} - \Delta z_k + \dfrac{\partial f}{\partial y}(x,t,\xi_k)z_k = \displaystyle\sum_{j=1}^m a_{k,j}g_j & \text{in } Q,\\ z_k = 0 & \text{on } \Sigma,\\ z_k(x,0) = 0 & \text{in } \Omega, \end{cases} \tag{3.2}$$

where $\xi_k(x,t) = \hat y_k(x,t) + \theta_k(x,t)(y_k(x,t) - \hat y_k(x,t)) = \hat y_k(x,t) + \theta_k(x,t)z_k(x,t)$ with $0 \le \theta_k(x,t) \le 1$. We argue by contradiction and we assume that

$$\rho_k = \max_{1\le j\le m} |a_{k,j}| \to +\infty \quad \text{as} \quad k \to \infty.$$

Then, introducing $\zeta_k = \frac{1}{\rho_k}z_k$, we deduce from (3.2)

$$\begin{cases} \dfrac{\partial \zeta_k}{\partial t} - \Delta \zeta_k + \dfrac{\partial f}{\partial y}(x,t,\xi_k)\zeta_k = \dfrac{1}{\rho_k}\displaystyle\sum_{j=1}^m a_{k,j}g_j & \text{in } Q,\\ \zeta_k = 0 & \text{on } \Sigma,\\ \zeta_k(x,0) = 0 & \text{in } \Omega, \end{cases} \tag{3.3}$$

From this equation, using (2.3), (2.4) and the boundedness of the right hand side in $L^\infty(Q)$ we have that $\|\zeta_k\|_{L^\infty(Q)} \le M$ for some $M > 0$ and $\forall k$. Moreover, the boundedness of $\{z_k\}_{k=1}^\infty$ in $L^2(Q)$ implies that $\|\zeta_k\|_{L^2(Q)} \to 0$. Now, (3.1) and Hölder's inequality with $\frac{2}{q}$ and $\frac{2}{2-q}$ lead to

$$\int_Q \left|\frac{\partial f}{\partial y}(x,t,\xi_k)\zeta_k\right| dx\,dt \le C\left(\int_Q (1+|\xi_k|^q)^{\frac{2}{q}}\,dx\,dt\right)^{\frac{q}{2}}\left(\int_Q |\zeta_k|^{\frac{2}{2-q}}\,dx\,dt\right)^{\frac{2-q}{q}}$$
$$\le C\left(\int_Q (1+[|\hat y_k| + |z_k|]^q)^{\frac{2}{q}}\,dx\,dt\right)^{\frac{q}{2}}\left(\int_Q |\zeta_k|^2\,dx\,dt\right)^{\frac{2-q}{q}}\|\zeta_k\|_{L^\infty(Q)}^{\frac{2q-2}{q}} \to 0.$$

From this and the properties of $\{\zeta_k\}_{k=1}^\infty$ is obvious that the left hand side of the partial differential equation in (3.3) converges to zero in the distribution sense. However, by the definition of ρ_k we have that the right hand side does not converge to zero, which is a contradiction. Consequently $\{a_k\}_{k=1}^\infty$ is a bounded sequence in $\mathbb{R}^m$, hence the minimizing sequence $\{u_k\}_{k=1}^\infty$ is bounded in $BV(0,T)^m$ because of (2.1). Therefore, we can take a subsequence, denoted in the same way, such that $u_k \stackrel{*}{\rightharpoonup} \bar{u}$ in $BV(0,T)^m$, therefore $u_k \to \bar{u}$ strongly in every $L^p(0,T)^m$ for $p < +\infty$. As a consequence of Proposition 2.2 we have that $y_k \to \bar{y}$ strongly in Y, where $\bar{y}$ is the state associated to $\bar{u}$, and thus $F(u_k) \to F(\bar{u})$. Furthermore, the convergence $u'_{k,j} \stackrel{*}{\rightharpoonup} \bar{u}'_{k,j}$ in $\mathcal{M}(0,T)$ for every $1 \le j \le m$ implies that

$$G(\bar{u}) = \sum_{j=1}^m \alpha_j \|\bar{u}'_j\|_{\mathcal{M}(0,T)} \le \liminf_{k\to\infty} \sum_{j=1}^m \alpha_j \|u'_{k,j}\|_{\mathcal{M}(0,T)} = \liminf_{k\to\infty} G(u_k).$$

Hence, $J(\bar{u}) \le \liminf_{k\to\infty} J(u_k) = \inf\,(\mathrm{P})$ and $\bar{u}$ is a solution of (P).

The uniqueness of a solution when f is affine with respect to y is an immediate consequence of the strict convexity of F and the convexity of G. ◻

Next we analyze the first order optimality conditions. Since (P) is not a convex problem it is convenient to deal with local solutions.

Definition 3.2. *We shall call $\bar{u}$ a local solution of* (P) *if there exists $\varepsilon > 0$ so that*

$$J(\bar{u}) \le J(u) \quad \forall u : \|u - \bar{u}\|_{BV(0,T)^m} \le \varepsilon.$$

We say that $\bar{u}$ is an $L^p(0,T)^m$-local solution ($1 \le p \le \infty$) if the above inequality holds in a $L^p(0,T)^m$-ball around $\bar{u}$. Finally, $\bar{u}$ is called a strong local solution if

$$J(\bar{u}) \le J(u) \quad \forall u : \|y_u - \bar{y}\|_{L^\infty(Q)} \le \varepsilon$$

for some $\varepsilon > 0$, where $\bar{y}$ and y_u denote the states associated to $\bar{u}$ and u, respectively. The solution is said strict in any of the previous senses if the inequality $J(\bar{u}) < J(u)$ holds in the give neighborhoods whenever $\bar{u} \ne u$.

We have the following relationships among these concepts. Since $BV(0,T)$ is continuously embedded into $L^p(0,T)$ for any $p \in [1,+\infty]$, we deduce that if $\bar{u}$ is an $L^p(0,T)^m$-local solution of (P), then it is a local solution. On the other hand, from Propositions 2.1 and 2.2 we infer that any strong local solution is an $L^p(0,T)^m$-local solution for $1 < p \le +\infty$.

Given $\bar{u} \in BV(0,T)^m$ with associated state and adjoint state $\bar{y}$ and $\bar{\varphi}$, respectively, we define

$$\bar{\Phi}_j(t) = \int_0^t \int_{\omega_j} \bar{\varphi}(x,s) g_j(x)\,dx\,ds + \beta_j t \int_0^T u_j(s)\,ds, \quad 1 \le j \le m. \tag{3.4}$$

Theorem 3.3. *If $\bar{u}$ is a local solution of* (P)*, then $\bar{\Phi}_j \in C^1[0,T] \cap C_0(0,T)$ for $1 \le j \le m$ and they satisfy*

$$\|\bar{\Phi}_j\|_{C_0(0,T)} \begin{cases} = \alpha_j & \text{if } \bar{u}'_j \ne 0, \\ \le \alpha_j & \text{if } \bar{u}'_j = 0, \end{cases} \tag{3.5}$$

$$\int_0^T \bar{\Phi}_j\,d\bar{u}'_j = \|\bar{\Phi}_j\|_{C_0(0,T)} \|\bar{u}'_j\|_{\mathcal{M}(0,T)}. \tag{3.6}$$

Proof. From Proposition 2.3 we know that $\bar\varphi \in C(\bar Q)$, hence $\bar\Phi_j \in C^1[0,T]$ follows for every j. Let us fix one component j and denote by e_j the j-th unit vector of the canonical basis in $\mathbb{R}^m$. Given $u \in BV(0,T)$, from the local optimality of $\bar u$ and the convexity of G we deduce for every $0 < \rho < 1$ small enough

$$\begin{aligned}
0 \le \frac{J(\bar u + \rho u e_j) - J(\bar u)}{\rho} &= \frac{F(\bar u + \rho u e_j) - F(\bar u)}{\rho} + \frac{G(\bar u + \rho u e_j) - G(\bar u)}{\rho}\\
\le \frac{F(\bar u + \rho u e_j) - F(\bar u)}{\rho} &+ [G(\bar u + u e_j) - G(\bar u)]\\
= \frac{F(\bar u + \rho u e_j) - F(\bar u)}{\rho} &+ \alpha_j[(g\circ D_t)(\bar u_j + u) - g(\bar u_j')].
\end{aligned}$$

Passing to the limit as $\rho \to 0$ in the above inequality and using (2.9) we get for every $u \in BV(0,T)$

$$0 \le \int_0^T \Big(\int_{\omega_j} \bar\varphi(x,t) g_j(x)\,dx + \beta_j \int_0^T \bar u_j(s)\,ds\Big) u_j(t)\,dt + \alpha_j[(g\circ D_t)(\bar u_j + u) - g(\bar u_j')].$$

Using (3.4), the above inequality can be written as

$$-\frac{1}{\alpha_j}\int_0^T \bar\Phi_j'(t) u(t)\,dt + g(\bar u_j') \le (g\circ D_t)(u) \quad \forall u \in BV(0,T).$$

From the above inequality, the definition of the subdifferential of a convex function and using (2.14) it follows

$$-\frac{1}{\alpha_j}\bar\Phi_j' \in \partial(g\circ D_t)(\bar u_j) = D_t^*\partial g(\bar u_j'). \tag{3.7}$$

Therefore, there exists $\bar\lambda_j \in \partial g(\bar u_j') \subset \mathcal{M}(0,T)^*$ such that

$$-\frac{1}{\alpha_j}\int_0^T \bar\Phi_j'(t) u(t)\,dt = \langle \bar\lambda_j, u'\rangle \quad \forall u \in BV(0,T). \tag{3.8}$$

As a first consequence of this identity is that $\Phi_j(T) = 0$. Indeed, it is enough to take $u \equiv 1$ and use that $\bar\Phi_j(0) = 0$, which follows obviously from the definition.

Given $u \in BV(0,T)$, we can select a sequence $\{u_k\}_{k=1}^\infty \subset C^\infty[0,T]$ converging weakly* to u in $BV(0,T)$; see [1, Remark 3.22]. Using this fact and the property $\bar\Phi_j(T) = \bar\Phi_j(0) = 0$ we get

$$-\int_0^T \bar\Phi_j'(t) u(t)\,dt = -\lim_{k\to\infty}\int_0^T \bar\Phi_j'(t) u_k(t)\,dt = \lim_{k\to\infty}\int_0^T \bar\Phi_j(t) u_k'(t)\,dt = \langle u', \bar\Phi_j\rangle.$$

Since this identity holds for all $u \in BV(0,T)$, and any measure in $\mathcal{M}(0,T)$ is the derivative of a function of $BV(0,T)$, we infer from (3.8) that $\bar\lambda_j = \frac{1}{\alpha_j}\bar\Phi_j \in C_0(0,T)$. Thus we have that $\frac{1}{\alpha_j}\bar\Phi_j \in \partial g(\bar u_j')$, which means

$$\langle \mu - \bar u_j', \frac{1}{\alpha_j}\bar\Phi_j\rangle + \|\bar u_j'\|_{\mathcal{M}(0,T)} \le \|\mu\|_{\mathcal{M}(0,T)} \quad \forall \mu \in \mathcal{M}(0,T).$$

Taking $\mu = 2\bar{u}'_j$ and $\mu = \frac{1}{2}\bar{u}'_j$, respectively, we deduce that

$$\langle \bar{u}'_j, \frac{1}{\alpha_j}\bar{\Phi}_j\rangle = \|\bar{u}'_j\|_{\mathcal{M}(0,T)},$$

and consequently

$$\langle \mu, \frac{1}{\alpha_j}\bar{\Phi}_j\rangle \leq \|\mu\|_{\mathcal{M}(0,T)} \quad \forall \mu \in \mathcal{M}(0,T).$$

The last two relationships are equivalent to (3.5) and (3.6). □

COROLLARY 3.4. *Under the assumptions of Theorem 3.3, for each* $j \in \{1,\dots,m\}$ *such that* $\bar{u}_j$ *is not a constant function on* $[0,T]$*, then we have*

$$\left\{ \begin{array}{l} \operatorname{supp}(\bar{u}'^{+}_j) \subset \{t \in [0,T] : \bar{\Phi}_j(t) = +\alpha_j\}, \\ \operatorname{supp}(\bar{u}'^{-}_j) \subset \{t \in [0,T] : \bar{\Phi}_j(t) = -\alpha_j\}, \end{array}\right. \tag{3.9}$$

where $\bar{u}'_j = \bar{u}'^{+}_j - \bar{u}'^{-}_j$ *is the Jordan decomposition of the measure* $\bar{u}'_j$*.*

This corollary is straightforward consequence of (3.5), (3.6), Proposition 2.4 with $\lambda = -\frac{1}{\alpha_j}\bar{\Phi}_j$, and the fact that $\bar{u}'_j \not\equiv 0$ if $\bar{u}_j$ is not a constant function in $[0,T]$.

Remark 3.5. 1. Let us observe that if the set of points where $\bar{\Phi}_j(t) \in \{-\alpha_j, +\alpha_j\}$ is finite then $\bar{u}'_j$ is a combination of Dirac measures centered at those points. In particular, we obtain that $\bar{u}_j$ is piecewise constant in $[0,T]$. This will be illustrated in the numerical examples, cf. Section 7.1 and Section 7.2.

2. Given $\alpha = (\alpha_j)_{j=1}^m$, let us denote by $\bar{u}_\alpha = (\bar{u}_{\alpha,j})_{j=1}^m$ a solution of (P) and $(\bar{y}_\alpha, \bar{\varphi}_\alpha)$ the associated state and adjoint state. We note that if α_j is decreased, then the $BV(0,T)$ seminorm of $\bar{u}_{\alpha,j}$ is increasing. On the contrary, if α_j is increased, then the $BV(0,T)$ seminorm of $\bar{u}_{\alpha,j}$ is decreasing. In fact, there is a threshold $M_j < +\infty$ such that if $\alpha_j > M_j$, then $\bar{u}'_{\alpha,j} = 0$, i.e., $\bar{u}_{\alpha,j}$ is constant in $[0,T]$. Moreover, there exists a vector $\bar{\xi} \in \mathbb{R}^m$ such that for any α with $\alpha_j > M_j$ for all $1 \leq j \leq m$, the constant function $\bar{\xi}$ is a solution of (P). Let us give an upper bound for these values M_j.

Let y^0 be the solution of the state equation associated to the control $u \equiv 0$. From the optimality of $\bar{u}_\alpha$ we get

$$\frac{1}{2}\|\bar{y}_\alpha - y_d\|^2_{L^2(Q)} + \sum_{j=1}^m \frac{\beta_j}{2}\Big(\int_0^T \bar{u}_{\alpha,j}(t)\,dt\Big)^2 \leq J(\bar{u}_\alpha) \leq J(0) = \frac{1}{2}\|y^0 - y_d\|^2_{L^2(Q)}.$$

From these inequalities we get

$$\|\bar{y}_\alpha - y_d\|_{L^2(Q)} \leq \|y^0 - y_d\|_{L^2(Q)} \text{ and } \beta_j\Big|\int_0^T \bar{u}_{\alpha,j}(t)\,dt\Big| \leq \sqrt{\beta_j}\|y^0 - y_d\|_{L^2(Q)}.$$

From the adjoint state equation we obtain

$$\|\bar{\varphi}_\alpha\|_{L^\infty(0,T;L^2(\Omega))} \leq C_\Omega\|\bar{y}_\alpha - y_d\|_{L^2(Q)} \leq C_\Omega\|y^0 - y_d\|_{L^2(Q)},$$

where C_Ω is the constant satisfying $\|z\|_{L^2(\Omega)} \leq C_\Omega\|\nabla z\|_{L^2(\Omega)}$ for any $z \in H^1_0(\Omega)$. From the definition of $\bar{\Phi}_j$ and the above estimates we get for every $t \in [0,T]$

$$\begin{aligned} |\bar{\Phi}_j(t)| &\leq T\|\bar{\varphi}_\alpha\|_{L^\infty(0,T;L^2(\Omega))}\|g_j\|_{L^2(\omega_j)} + \beta_j\Big|\int_0^T \bar{u}_{\alpha,j}(t)\,dt\Big| \\ &\leq (TC_\Omega\|g_j\|_{L^2(\omega_j)} + \sqrt{\beta_j})\|y^0 - y_d\|_{L^2(Q)} = M_j. \end{aligned}$$

Relations (3.9) imply that $\bar{u}'_{\alpha,j} \equiv 0$ if $\alpha_j > M_j$.

To prepare for the second order necessary conditions we introduce the critical cone as follows

$$C_{\bar{u}} = \{v \in BV(0,T)^m : F'(\bar{u})v + G'(\bar{u}; v) = 0\}. \tag{3.10}$$

It seems natural that the second order optimality conditions must be imposed only on those directions where the directional derivatives vanish. Let us point out some properties of this critical cone.

PROPOSITION 3.6. *$C_{\bar{u}}$ is a closed convex cone that can equivalently be expressed in the form*

$$C_{\bar{u}} = \left\{ v \in BV(0,T)^m : \int_0^T \bar{\Phi}_j(t)\, dv'_{js}(t) = \alpha_j \|v'_{js}\|_{\mathcal{M}(0,T)},\ 1 \le j \le m \right\}, \tag{3.11}$$

where v'_{js} is the singular part of the measure v'_j with respect to $|\bar{u}'_j|$.

The identity (3.11) shows that the criterion for v to be in $C_{\bar{u}}$ can be expressed in terms of the singular part of v'_j with respect to $|\bar{u}'_j|$ for $1 \le j \le m$. In particular, any function $v \in B(0,T)^m$ such that v'_j is absolutely continuous with respect to $|\bar{u}'_j|$ for every j is an element of the critical cone.

Proof. The cone property and closedness of $C_{\bar{u}}$ are a straightforward consequence of the continuity and positive homogeneity of the mapping $v \to F'(\bar{u})v + G'(\bar{u}; v)$. Let us prove the convexity property. First we observe that (2.9) and the definition of $\bar{\Phi}_j$ implies that

$$F'(\bar{u})v = \sum_{j=1}^m \int_0^T \bar{\Phi}'_j(t) v_j(t)\, dt \quad \forall v \in BV(0,T)^m. \tag{3.12}$$

Taking into account (3.7), using the definition of the subdifferential and passing to the limit as $\rho \searrow 0$ we infer for $1 \le j \le m$

$$-\frac{1}{\alpha_j} \int_0^T \bar{\Phi}'_j(t) v_j(t)\, dt \le \frac{g(\bar{u}'_j + \rho v'_j) - g(\bar{u}'_j)}{\rho} \to g'(\bar{u}'_j; v'_j).$$

Multiplying this inequality by α_j and summing in j we get with (3.12)

$$F'(\bar{u})v + G'(\bar{u}; v) \ge 0 \ \ \forall v \in BV(0,T)^m. \tag{3.13}$$

Therefore, $v \in C_{\bar{u}}$ if and only if $F'(\bar{u})v + G'(\bar{u}; v) \le 0$. Since the mapping $v \in BV(0,T)^m \to F'(\bar{u})v + G'(\bar{u}; v)$ is convex, we conclude the convexity of $C_{\bar{u}}$.

From (3.12), making an integration by parts as in the proof of Theorem 3.3, and using the Lebesgue decomposition $dv'_j = h_{v'_j} d|\bar{u}'_j| + dv'_{js}$ we get

$$F'(\bar{u})v = -\sum_{j=1}^m \int_0^T \bar{\Phi}_j\, dv'_j = -\left\{ \sum_{j=1}^m \int_0^T \bar{\Phi}_j h_{v'_j}\, d|\bar{u}'_j| + \int_0^T \bar{\Phi}_j\, dv'_{js} \right\}.$$

From (3.9) we deduce that $d|\bar{u}'_j| = \frac{1}{\alpha_j} \bar{\Phi}_j d\bar{u}'_j$ for $1 \le j \le m$. Inserting this identity in the above equality we infer

$$F'(\bar{u})v = -\left\{ \sum_{j=1}^m \alpha_j \int_0^T h_{v'_j}\, d\bar{u}'_j + \int_0^T \bar{\Phi}_j\, dv'_{js} \right\}. \tag{3.14}$$

Now, using (2.15) it follows

$$G'(\bar u; v) = \sum_{j=1}^m \alpha_j \left\{ \int_0^T h_{v'_j}\, d\bar u'_j + \|v'_{js}\|_{\mathcal{M}(0,T)} \right\}.$$

This equality and (3.14) lead to

$$F'(\bar u)v + G'(\bar u; v) = \sum_{j=1}^m \left\{ -\int_0^T \bar\Phi_j\, dv'_{js} + \alpha_j \|v'_{js}\|_{\mathcal{M}(0,T)} \right\},$$

which is equivalent to the expressions given in (3.11) for $1 \le j \le m$. □

Now, we formulate the second order necessary optimality conditions.

THEOREM 3.7. *If $\bar u$ is a local minimum of* (P), *then $F''(\bar u)v^2 \ge 0$ for all $v \in C_{\bar u}$.*

Proof. Let v be an element in $C_{\bar u}$ and consider the Lebesgue decomposition $dv'_j = h_{v'_j} d|\bar u'_j| + dv'_{js}$, $1 \le j \le m$. For every integer $k \ge 1$ we set

$$h_{j,k}(t) = \mathrm{proj}_{[-k,+k]}(h_{v'_j}(t)) \quad \text{and} \quad dv'_{j,k} = h_{j,k} d|\bar u'_j| + dv'_{js}.$$

Let us take $v_{j,k} \in L^1(0,T)$ as the primitive of $v'_{j,k}$ with $\int_0^T (v_j - v_{j,k})\, dt = 0$, and set $v_k = (v_{1,k}, \ldots, v_{m,k})$. Then, we have $\|v'_j - v'_{j,k}\|_{\mathcal{M}(0,T)} = \|h_{v'_j} - h_{j,k}\|_{L^1(|\bar u'_j|)} \to 0$ by Lebesgue's dominated convergence theorem. Hence $v_k \to v$ in $BV(0,T)^m$. Moreover, since the singular parts of $v'_{j,k}$ and v'_j with respect to $|\bar u'_j|$ coincide and $v \in C_{\bar u}$, then (3.11) implies that $v_k \in C_{\bar u}$ for every k.

For any $0 < \rho < \frac{1}{k}$, using (2.12) and (2.13), we find

$$\begin{aligned}
&\frac{G(\bar u + \rho v_k) - G(\bar u)}{\rho} = \sum_{j=1}^m \alpha_j \frac{g(\bar u'_j + \rho v'_{j,k}) - g(\bar u'_j)}{\rho} \\
&= \sum_{j=1}^m \alpha_j \left\{ \int_0^T \frac{|1 + \rho h_{v'_{j,k}}| - 1}{\rho}\, d|\bar u'_j| + \|v'_{js}\|_{\mathcal{M}(0,T)} \right\} \\
&= \sum_{j=1}^m \alpha_j \left\{ \int_0^T \frac{|1 + \rho h_{v'_{j,k}}| - 1}{\rho}\, d{\bar u'_j}^{+} + \int_0^T \frac{|-1 + \rho h_{v'_{j,k}}| - 1}{\rho}\, d{\bar u'_j}^{-} + \|v'_{js}\|_{\mathcal{M}(0,T)} \right\} \\
&= \sum_{j=1}^m \alpha_j \left\{ \int_0^T h_{v'_{j,k}}\, d\bar u'_j + \|v'_{js}\|_{\mathcal{M}(0,T)} \right\} = G'(\bar u; v_k).
\end{aligned}$$

Now, using that $\bar u$ is a local minimum of J and making a Taylor expansion we get for every k and $0 < \rho < \frac{1}{k}$ the existence of $\theta = \theta(k,\rho)$, with $0 < \theta < 1$, such that

$$0 \le \frac{J(\bar u + \rho v_k) - J(\bar u)}{\rho} = F'(\bar u)v_k + \frac{\rho}{2} F''(\bar u + \theta\rho v_k)v_k^2 + G'(\bar u; v_k) = \frac{\rho}{2} F''(\bar u + \theta\rho v_k)v_k^2,$$

since $v_k \in C_{\bar u}$. Finally, dividing the last term by $\rho/2$ and taking the limit when $\rho \to 0$ and later when $k \to \infty$, we get that $F''(\bar u)v^2 \ge 0$. □

As usual, we have to consider an extended cone of critical directions to formulate a sufficient second order condition for optimality. For every $\tau > 0$, we denote

$$C^\tau_{\bar u} = \{v \in BV(0,T)^m : F'(\bar u)v + G'(\bar u; v) \le \tau\big(\|z_v\|_{L^2(Q)} + \sum_{j=1}^m \beta_j \big|\int_0^T v_j(t)\, dt\big|\big)\},$$

where $z_v = S'(\bar{u})v$, with S defined just above Proposition 2.2. The second order condition involves this cone as follows:

(SSOC) There exist positive constants κ and τ such that

$$F''(\bar{u})v^2 \geq \kappa \|z_v\|^2_{L^2(Q)} \quad \forall v \in C^\tau_{\bar{u}}. \tag{3.15}$$

THEOREM 3.8. *Let $\bar{u} \in BV(0,T)^m$ satisfy the first order optimality conditions* (3.5)-(3.6) *and (SSOC). Then, there exist positive constants $\varepsilon > 0$ and $\nu > 0$ such that*

$$J(\bar{u}) + \frac{\nu}{2}\|z_{u-\bar{u}}\|^2_{L^2(Q} \leq J(u) \quad \textit{for all } u \in BV(0,T)^m : \|y_u - \bar{y}\|_{L^\infty(Q)} < \varepsilon. \tag{3.16}$$

The proof of this theorem can be done along the lines of [8, Theorem 9]. Let us point out some small differences. First, the parameter γ in [8] must be taken zero. Second, we have a non-differentiable part in the cost functional and a slightly different cone of critical directions. To deal with the non-differentiable term G we use (3.13) and its convexity and Lipschitz continuity: for every $u \in BV(0,T)^m$

$$\begin{aligned} J(u) - J(\bar{u}) &= F'(\bar{u})(u-\bar{u}) + \frac{1}{2}F''(\bar{u} + \theta(u-\bar{u}))(u-\bar{u})^2 + G(u) - G(\bar{u}) \\ &\geq \frac{1}{2}F''(\bar{u} + \theta(u-\bar{u}))(u-\bar{u})^2 + F'(\bar{u})(u-\bar{u}) + G'(\bar{u}; u-\bar{u}) \\ &\geq \frac{1}{2}F''(\bar{u} + \theta(u-\bar{u}))(u-\bar{u})^2. \end{aligned}$$

In this way we eliminate the non-differentiable part of the cost functional. The rest is the same.

COROLLARY 3.9. *Under the assumptions of Theorem 3.8 there exist two constants $\varepsilon > 0$ and $\delta > 0$ such that*

$$J(\bar{u}) + \frac{\delta}{2}\|y_u - \bar{y}\|^2_{L^2(Q)} \leq J(u) \quad \textit{for all } u \in BV(0,T)^m : \|y_u - \bar{y}\|_{L^\infty(Q)} < \varepsilon. \tag{3.17}$$

This is an immediate consequence of (3.16) and the estimate

$$\|y_u - \bar{y}\|_{L^2(Q)} \leq M\|z_{u-\bar{u}}\|_{L^2(Q)} \quad \forall u \in BV(0,T)^m : \|y_u - \bar{y}\|_{L^\infty(Q)} < \varepsilon;$$

see [8, Corollary 3] for the proof.

We observe that the sufficient second order optimality condition (3.15) along with the first order optimality condition imply that $\bar{u}$ is a strong local solution of (P).

4. Approximation of the control problem. In this section, we assume that Ω is a convex set and $y_0 \in L^\infty(\Omega) \cap H^1_0(\Omega)$. Then, it is well known that the solutions y_u of (1.1) belong to $C([0,T], H^1_0(\Omega)) \cap H^{2,1}(Q)$; see, for instance [20, Proposition 2.4].

We consider a dG(0)cG(1) discontinuous Galerkin approximation of the state equation (1.1) (i.e., piecewise constant in time and linear nodal basis finite elements in space; see, e.g., [21]). Associated with a parameter h we consider a family of triangulations $\{\mathcal{K}_h\}_{h>0}$ of $\bar{\Omega}$. To every element $K \in \mathcal{K}_h$ we assign two parameters $\rho(K)$ and $\vartheta(K)$, where $\rho(K)$ denotes the diameter of K and $\vartheta(K)$ is the diameter of the biggest ball contained in K. The size of the grid is given by $h = \max_{K\in\mathcal{K}_h} \rho(K)$. We will denote by $\{x_j\}_{j=1}^{N_h}$ the interior nodes of the triangulation $\mathcal{K}_h$. In this section Ω will be assumed to be convex. In addition, the following usual regularity assumptions on the triangulation are assumed.

(i) There exist two positive constants ρ_Ω and ϑ_Ω such that

$$\frac{h}{\rho(K)} \le \rho_\Omega \quad \text{and} \quad \frac{\rho(K)}{\vartheta(K)} \le \vartheta_\Omega$$

hold for every $K \in \mathcal{K}_h$ and all $h > 0$.

(ii) Let us set $\overline{\Omega}_h = \cup_{K\in\mathcal{K}_h} K$ with Ω_h and Γ_h being its interior and boundary, respectively. In the case of polygonal or polyhedral domains, it is reasonable to assume that the triangulation satisfies that $\Gamma_h = \Gamma$. We assume that the vertices of $\mathcal{K}_h$ placed on the boundary Γ_h are also points of Γ and there exists a constant $C_\Gamma > 0$ such that $\mathrm{dist}(x,\Gamma) \le C_\Gamma h^2$ for every $x \in \Gamma_h$. This always holds if Γ is a C^2 boundary. From this assumption we know [18, Section 5.2] that

$$|\Omega \setminus \Omega_h| \le Ch^2, \tag{4.1}$$

where $|\cdot|$ denotes the Lebesgue measure.

We also introduce a temporal grid $0 = t_0 < t_1 < \ldots < t_{N_\tau} = T$ with $\tau_k = t_k - t_{k-1}$ and set $\tau = \max_{1\le k\le N_\tau} \tau_k$. We denote $I_k = (t_{k-1}, t_k)$. We assume that there exist $\rho_T > 0$, $C_{\Omega,T} > 0$ and $c_{\Omega,T} > 0$ independent of h and τ such that $\tau \le \rho_T \tau_k$ for $1 \le k \le N_\tau$. We will use the notation $\sigma = (h,\tau)$ and $Q_h = \Omega_h \times (0,T)$.

4.1. Discretization of the controls. Associated with the grid $\{t_k\}_{k=0}^{N_\tau}$ we define the subspace

$$U_\tau = \{u_\tau \in BV(0,T) : u_\tau = \sum_{k=1}^{N_\tau} u_k\chi_k, \text{ with } \{u_k\}_{k=1}^{N_\tau} \subset \mathbb{R}\},$$

where χ_k denotes the characteristic function of the interval I_k. Let us observe that the elements $u_\tau \in U_\tau$ are piecewise constant functions whose distributional derivative is given by

$$u'_\tau = D_t u_\tau = \sum_{k=2}^{N_\tau} (u_k - u_{k-1})\delta_{t_{k-1}} \quad \text{and} \quad \|u'_\tau\|_{\mathcal{M}(0,T)} = \sum_{k=2}^{N_\tau} |u_k - u_{k-1}|, \tag{4.2}$$

where δ_t denotes the Dirac measure concentrated at the point t. We further define the projection operator

$$\Lambda_\tau : BV(0,T) \longrightarrow U_\tau, \quad \Lambda_\tau u = \sum_{k=1}^{N_\tau} \left(\frac{1}{\tau_k}\int_{I_k} u(t)\,dt\right)\chi_k.$$

Proposition 4.1. *For any $u \in BV(0,T)$ the following properties hold:*

$$\|u - \Lambda_\tau u\|_{L^1(0,T)} \le \tau \|D_t u\|_{\mathcal{M}(0,T)}, \tag{4.3}$$

$$\|D_t \Lambda_\tau u\|_{\mathcal{M}(0,T)} \le \|D_t u\|_{\mathcal{M}(0,T)}, \tag{4.4}$$

$$\lim_{\tau\to 0} \|D_t \Lambda_\tau u\|_{\mathcal{M}(0,T)} = \|D_t u\|_{\mathcal{M}(0,T)}. \tag{4.5}$$

Proof. The inequality (4.3) is simple to establish for $u \in C^1[0,T]$. Henceforth let $u \in BV(0,T)$. Then there exists a sequence $\{u_j\}_{j=1}^\infty \subset C^\infty[0,T]$ such that

$$\|u - u_j\|_{L^1(0,T)} + \ |\|D_t u\|_{\mathcal{M}(0,T)} - \|D_t u_j\|_{\mathcal{M}(0,T)}| \le \frac{1}{j} \ \ \forall j \ge 1; \tag{4.6}$$

see [1, Remark 3.22]. Now we estimate as follows

$$\|u - \Lambda_\tau u\|_{L^1(0,T)} \le \|u - u_j\|_{L^1(0,T)} + \|u_j - \Lambda_\tau u_j\|_{L^1(0,T)} + \|\Lambda_\tau u_j - \Lambda_\tau u\|_{L^1(0,T)}$$

$$\le \|u - u_j\|_{L^1(0,T)} + \tau\|D_t u_j\|_{\mathcal{M}(0,T)} + \|u_j - u\|_{L^1(0,T)} \le \frac{2}{j} + \tau\|D_t u_j\|_{\mathcal{M}(0,T)}.$$

Using (4.6) we can pass to the limit in the above inequality as $j \to \infty$ to deduce (4.3).

Let us prove (4.4). First, we assume again that $u \in C^\infty[0,T]$. From the continuity of u and the mean value theorem for integrals we deduce the existence of points $\xi_k \in I_k$, $1 \le k \le N_\tau$, such that

$$\Lambda_\tau u = \sum_{k=1}^{N_\tau} u(\xi_k)\chi_k.$$

Then we have with (4.2)

$$\|D_t \Lambda_\tau u\|_{\mathcal{M}(0,T)} = \sum_{k=2}^{N_\tau} |u(\xi_k) - u(\xi_{k-1})|$$
$$\le \sum_{k=2}^{N_\tau} \int_{\xi_{k-1}}^{\xi_k} |u'(t)|\,dt \le \int_0^T |u'(t)|\,dt = \|D_t u\|_{\mathcal{M}(0,T)}.$$

For the case $u \in BV(0,T)$, we take again a sequence $\{u_j\}_{j=1}^\infty \subset C^\infty[0,T]$ satisfying (4.6). The convergence $u_j \to u$ in $L^1(0,T)$ obviously implies that $\Lambda_\tau u_j \to \Lambda_\tau u$ in $L^1(0,T)$. Then, using [1, Proposition 3.6], inequality (4.4) for every u_j, and (4.6) we conclude

$$\|D_t \Lambda_\tau u\|_{\mathcal{M}(0,T)} \le \liminf_{j\to\infty} \|D_t \Lambda_\tau u_j\|_{\mathcal{M}(0,T)} \le \liminf_{j\to\infty} \|D_t u_j\|_{\mathcal{M}(0,T)} = \|D_t u\|_{\mathcal{M}(0,T)},$$

which implies (4.4).

Finally, to prove (4.5) we use (4.3), [1, Proposition 3.6] and (4.4) to obtain

$$\|D_t u\|_{\mathcal{M}(0,T)} \le \liminf_{\tau\to 0} \|D_t \Lambda_\tau u\|_{\mathcal{M}(0,T)} \le \limsup_{\tau\to 0} \|D_t \Lambda_\tau u\|_{\mathcal{M}(0,T)} \le \|D_t u\|_{\mathcal{M}(0,T)}.$$

□

4.2. Discrete state equation. Associated with the interior nodes of the triangulation $\{x_j\}_{j=1}^{N_h}$ we consider the space

$$Y_h = \{y_h \in C_0(\Omega) : y_h = \sum_{j=1}^{N_h} y_j e_j \text{ with } \{y_j\}_{j=1}^{N_h} \subset \mathbb{R}\}$$

where $\{e_j\}_{j=1}^{N_h}$ is the nodal basis formed by the continuous piecewise linear functions such that $e_j(x_i) = \delta_{ij}$ for every $1 \le i, j \le N_h$. For every σ we define the space of discrete states by

$$\mathcal{Y}_\sigma = \{y_\sigma \in L^2(I, Y_h) : y_\sigma|_{I_k} \in Y_h,\ 1 \le k \le N_\tau\},$$

The elements $y_\sigma \in \mathcal{Y}_\sigma$ can be represented in the form

$$y_\sigma = \sum_{k=1}^{N_\tau} y_{k,h}\chi_k = \sum_{k=1}^{N_\tau}\sum_{j=1}^{N_h} y_{kj}\chi_k e_j \text{ with } \{y_{k,h}\}_{k=1}^{N_\tau} \subset Y_h \text{ and } \{y_{kj}\}_{\substack{1\le k\le N_\tau\\ 1\le j\le N_h}} \subset \mathbb{R}. \tag{4.7}$$

We approximate the state equation (1.1) as follows. For any control $u \in BV(0,T)^m$ we define the associated discrete state $y_\sigma \in \mathcal{Y}_\sigma$ as the solution of the system

$$\begin{cases} \big(\dfrac{y_{k,h}-y_{k-1,h}}{\tau_k}, z_h\big) + a(y_{k,h}, z_h) + \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (f(\cdot,t,y_{k,h}), z_h)\,dt \\ \qquad\qquad = \dfrac{1}{\tau_k}\displaystyle\sum_{j=1}^m (g_j, z_h)\int_{I_k} u_j(t)\,dt, \quad \forall z_h \in Y_h,\ 1\le k\le N_\tau \\ y_{0,h} = y_{0h}, \end{cases} \tag{4.8}$$

where $(\cdot,\cdot)$ denotes the scalar product in $L^2(\Omega)$, a is the bilinear form associated to the operator $-\Delta$, i.e.,

$$a(y,z) = \int_\Omega \nabla y \nabla z\,dx,$$

and y_{0h} is the projection $P_h y_0$ of y_0 on Y_h given by the variational equation

$$(P_h y_0, z_h) = (y_0, z_h) \quad \forall z_h \in Y_h.$$

It is well known that $y_{0h} \to y_0$ in $H_0^1(\Omega)$.

Proposition 4.2. *For every $u \in BV(0,T)^m$ the system (4.8) has a unique solution $y_\sigma \in \mathcal{Y}_\sigma$. In addition, if either f is affine with respect to the state or if $n < 3$, then the following estimate holds*

$$\|y_u - y_\sigma\|_{L^2(Q)} \le C(\tau + h^2), \tag{4.9}$$

where C is independent of σ.

These results are proved in [15] and [16] for f affine and nonlinear respectively. The constant C there depends on the norms of the state in $H^{2,1}(Q)$, and also on the $L^\infty(Q)$ norm in the semilinear case. These quantities can be estimated in our case by the $L^2(0,T)^m$ norm of u.

Remark 4.3. *Given $\{u_j\}_{j=1}^m \subset BV(0,T)$, we observe that*

$$\int_{I_k} u_j(t)\,dt = \int_{I_k} \Lambda_\tau u_j(t)\,dt \quad \textit{for all } 1\le j\le m \textit{ and } 1\le k\le N_\tau.$$

Utilizing this in (4.8), we deduce that the discrete states associated to $\{u_j\}_{j=1}^m$ and $\{\Lambda_\tau u_j\}_{j=1}^m$ coincide.

4.3. Discrete optimal control problem. The discrete control problem is defined as

$$(\mathrm{P}_\sigma) \quad \min_{u\in BV(0,T)^m} J_\sigma(u) = \frac{1}{2}\|y_\sigma - y_d\|^2_{L^2(Q_h)} + \sum_{j=1}^m \Big(\alpha_j\|u_j'\|_{\mathcal{M}(0,T)} + \frac{\beta_j}{2}\big(\int_0^T u_j(t)\,dt\big)^2\Big),$$

where y_σ is the discrete state associated to $u = (u_j)_{j=1}^m$.

The following assumption will be used to analyze the existence and uniqueness of a solution of (P_σ):

(A) The mapping $z_h \in Y_h \longrightarrow ((g_j, z_h))_{j=1}^m \in \mathbb{R}^m$ is surjective.

LEMMA 4.4. *There exists $h_0 > 0$ such that (A) holds for every $h < h_0$.*

Proof. Let us recall that $\{e_k\}_{k=1}^{N_h}$ denotes the nodal basis of Y_h. Since the supports ω_j of the functions g_j are compact and disjoint, we deduce the existence of $\hat{h} > 0$ such that for every $h < \hat{h}$, if for some e_k and some $1 \le j \le m$ we have that $\mathrm{supp}(e_k) \cap \omega_j \neq \emptyset$, then $\mathrm{supp}(e_k) \cap \omega_i = \emptyset$ for every $i \neq j$.

Moreover, there exists $\tilde{h}$ with the following property: $\forall h < \tilde{h}$ and $\forall j$ there exists some k such that $(g_j, e_k) \neq 0$. Indeed, if this is not the case, we infer the existence of sequence $\{h_i\}_{i=1}^\infty$ decreasing to 0 such that $(g_j, z_{h_i}) = 0$ for every $z_{h_i} \in Y_{h_i}$. In particular, taking z_{h_i} equal to the $L^2(\Omega)$-projection of g_j on Y_{h_i} we obtain

$$\|g_j\|_{L^2(\Omega)}^2 = \lim_{i \to \infty} (g_j, z_{h_i}) = 0,$$

which contradicts our assumption $g_j \neq 0$.

Finally, for any $h < h_0 = \min\{\hat{h}, \tilde{h}\}$ the assumption *(A)* holds. If not, then there exists a vector $(a_i)_{j=i}^m \subset \mathbb{R}^m$ such that

$$\sum_{i=1}^m (g_i, z_h) a_i = 0 \quad \forall z_h \in Y_h.$$

For any j we choose $e_k \in Y_h$ such that $(g_j, e_k) \neq 0$. Hence, $\mathrm{supp}(e_k) \cap \omega_j \neq \emptyset$, and $\mathrm{supp}(e_k) \cap \omega_i = \emptyset$ holds for every $i \neq j$. Then,

$$0 = \sum_{i=1}^m (g_i, e_k) a_i = (g_j, e_k) a_j,$$

which implies that $a_j = 0$. Since j was arbitrary in $\{1, \ldots, m\}$ we get a contradiction. □

THEOREM 4.5. *Let us assume that (A) holds. Then problem (P_σ) has at least one solution. Moreover, if $\tilde{u}$ is a solution of (P_σ), then $\bar{u}_\tau = (\Lambda_\tau \tilde{u}_j)_{j=1}^m$ is also a solution of (P_σ). In addition, if f is affine with respect to y, then $\bar{u}_\tau$ is the unique solution belonging to U_τ^m.*

Proof. To establish the existence of a solution $\tilde{u}$ we follow the lines of the proof of Theorem 3.1. The only concern is the boundedness of the sequence $\{a_k\}_{k=1}^\infty$ in $\mathbb{R}^m$. For this purpose we express the difference $z_{\sigma,k} = y_{\sigma,k} - \hat{y}_{\sigma,k}$, where $y_{\sigma,k}$ and $\hat{y}_{\sigma,k}$ are the solutions to (4.8) corresponding to u_k and $\hat{u}_k$, respectively. Thus, $z_{\sigma,k}$ is solution of the following system

$$\left\{ \begin{aligned} &\Big(\frac{z_{i,h;k} - z_{i-1,h;k}}{\tau_i}, z_h\Big) + a(z_{i,h;k}, z_h) + \frac{1}{\tau_i} \int_{I_i} (\partial_y f(\cdot, t, \xi_{i,h;k}) z_{i,h;k}, z_h)\, dt \\ &\qquad\qquad = \sum_{j=1}^m (g_j, z_h) a_{k,j} \quad \forall z_h \in Y_h,\ 1 \le i \le N_\tau \\ &z_{0,h;k} = 0, \end{aligned} \right. \tag{4.10}$$

where $\xi_{i,h;k} = \hat{y}_{i,h;k} + \theta_{i,h;k}(x,t) z_{i,h;k}$ with $0 \le \theta_{i,h;k}(x,t) \le 1$.

As in the proof of Theorem 3.1, we have that $\{y_{\sigma,k}\}_{k=1}^\infty$ and $\{\hat{y}_{\sigma,k}\}_{k=1}^\infty$ are bounded in $L^2(Q)$. Since $\mathcal{Y}_\sigma \subset L^\infty(Q)$ and since it is finite dimensional, we deduce that $\{\hat{y}_{\sigma,k}\}_{k=1}^{\infty}$ and $\{y_{\sigma,k}\}_{k=1}^{\infty}$ are also bounded in $L^\infty(Q)$. Therefore, the sequences $\{\xi_{i,h;k}\}_{k=1}^{\infty}$ are bounded in $L^\infty(\Omega \times I_i)$ as well. Again we argue by contradiction and we assume that $\rho_k = \max\{|a_{k,j}| : 1 \leq j \leq m\} \to \infty$ as $k \to \infty$. Then, we define $\zeta_{\sigma,k} = \frac{1}{\rho_k} z_{\sigma,k}$ and $\hat{a}_{k,j} = \frac{a_{k,j}}{\rho_k}$. By taking a subsequence, we have that $\zeta_{\sigma,k} \to 0$ in $L^\infty(Q)$ and $\hat{a}_{k,j} \to \hat{a}_j$, $1 \leq j \leq m$ for some $\{\hat{a}_j\}_{j=1}^m \subset \mathbb{R}$. We observe that by definition of ρ_k the vector $\hat{a} \neq 0$. Dividing (4.10) by ρ_k we obtain the mentioned subsequence

$$\begin{cases} \Big(\dfrac{\zeta_{i,h;k} - \zeta_{i-1,h;k}}{\tau_i}, z_h\Big) + a(\zeta_{i,h;k}, z_h) + \dfrac{1}{\tau_i}\displaystyle\int_{I_i} (\partial_y f(\cdot, t, \xi_{i,h;k})\zeta_{i,h;k}, z_h)\, dt \\ \qquad\qquad\qquad = \displaystyle\sum_{j=1}^m (g_j, z_h)\hat{a}_{k,j} \quad \forall z_h \in Y_h,\ 1 \leq i \leq N_\tau \\ z_{0,h;k} = 0. \end{cases}$$

Passing to the limit in this system as $k \to \infty$ we get that

$$\sum_{j=1}^m (g_j, z_h)\hat{a}_j = 0 \quad \forall z_h \in Y_h.$$

Hence, assumption *(A)* implies that $\hat{a} = 0$, which is the desired contradiction. Consequently, the sequence $\{a_k\}_{k=1}^{\infty}$ is bounded, and then the existence of a solution $\tilde{u}$ follows by standard arguments.

The fact that $\bar{u}_\tau = (\Lambda_\tau \tilde{u}_j)_{j=1}^m$ is also a solution of (P_σ) is an immediate consequence of Remark 4.3 and inequality (4.4). Finally, we prove the uniqueness of a solution in U_τ^m if f is affine with respect to the state. First we observe that both terms in the cost functional are convex in this case. Moreover, the first term is strictly convex on U_τ^m provided that the affine mapping $u_\tau \to y_\sigma$ is injective. To this end we assume that for some $u_\tau = (u_j)_{j=1}^m \in U_\tau^m$, with $u_j = \sum_{k=1}^{N_\tau} u_{j,k}\chi_k$, the associated discrete state y_σ is identically zero. Then from (4.8) we have that

$$\sum_{j=1}^m (g_j, z_h)u_{j,k} = 0 \ \ \forall z_h \in Y_h,\ \forall\, 1 \leq k \leq N_\tau.$$

Again by assumption *(A)* we infer that $u_j = 0$ for every $1 \leq j \leq m$, hence $u_\tau = 0$. ☐

Remark 4.6. *In the case that $\beta_j > 0$ for all $1 \leq j \leq m$, condition (A) is not needed to establish the existence of a solution of* (P_σ)*. However, it is still necessary for the uniqueness in the case that f is affine with respect to y.*

The rest of this section is devoted to the formulation of the first order optimality conditions for the problem (P_σ). Arguing in a similar way as for the continuous problem (P), we separate the smooth and the convex parts of J_σ

$$J_\sigma(u) = F_\sigma(u) + G(u), \quad \text{with } F_\sigma(u) = \frac{1}{2}\|y_\sigma - y_d\|^2_{L^2(Q_h)} + \sum_{j=1}^m \frac{\beta_j}{2}\Big(\int_0^T u_j(t)\, dt\Big)^2,$$

where y_σ is related to u by the equation (4.8). The derivative of F_σ is expressed by

$$F'_\sigma(u)v = \sum_{j=1}^m \int_0^T \Big(\int_{\omega_j} \varphi_\sigma(x,t)g_j(x)\, dx + \beta_j \int_0^T u_j(s)\, ds\Big) v_j(t)\, dt, \tag{4.11}$$

where $\varphi_\sigma \in \mathcal{Y}_\sigma$ is the adjoint state associated to u, i.e.

$$\left\{\begin{array}{l} \Big(\dfrac{\varphi_{k,h} - \varphi_{k+1,h}}{\tau_k}, z_h\Big) + a(\varphi_{k,h}, z_h) + \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (\partial_y f(\cdot, t, y_{k,h})\varphi_{k,h}, z_h)\, dt \\ \qquad\qquad\qquad = \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (y_{k,h} - y_d, z_h)\, dt, \quad \forall z_h \in Y_h,\ k = N_\tau, \dots 1 \\ \varphi_{N_\tau+1,h} = 0. \end{array}\right. \tag{4.12}$$

Using this expression for F'_σ and arguing exactly as in the proof of Theorem 3.3 we obtain the first order optimality conditions for a local solution $\bar u_\tau \in BV(0,T)^m$ of (P_σ). For this purpose we introduce the functions

$$\bar\Phi_{\sigma,j}(t) = \int_0^t \int_{\omega_j} \bar\varphi_\sigma(x,s) g_j(x)\, dx\, ds + \beta_j t \int_0^T u_j(s)\, ds, \quad 1 \le j \le m, \tag{4.13}$$

where $\bar\varphi_\sigma \in \mathcal{Y}_\sigma$ is the adjoint state associated to $\bar u_\tau$.

THEOREM 4.7. *If $\bar u_\tau$ is a local solution of (P_σ), then $\bar\Phi_{\sigma,j} \in C^1[0,T] \cap C_0(0,T)$ for $1 \le j \le m$, $\frac{1}{\alpha_j}\bar\Phi_{\sigma,j} \in \partial g(\bar u'_{\tau,j})$, and hence they satisfy*

$$\|\bar\Phi_{\sigma,j}\|_{C_0(0,T)} \left\{\begin{array}{ll} = \alpha_j & \text{if } \bar u_{\tau,j} \ne 0, \\ \le \alpha_j & \text{if } \bar u_{\tau,j} = 0, \end{array}\right. \tag{4.14}$$

$$\int_0^T \bar\Phi_{\sigma,j}\, d\bar u'_{\tau,j} = \|\bar\Phi_{\sigma,j}\|_{C_0(0,T)} \|\bar u'_{\tau,j}\|_{\mathcal{M}(0,T)}. \tag{4.15}$$

In the case where $\bar u_\tau$ is a local solution of (P_σ) belonging to U_τ^m (see Theorem 4.5), we have the following sparsity result analogous to Corollary 3.4.

COROLLARY 4.8. *Let $\bar u_\tau = (\bar u_{\tau,j})_{j=1}^m \in U_\tau^m$ be a local solution of (P_σ). Then, for each $j \in \{1, \dots, m\}$ such that $\bar u_{\tau,j}$ is not a constant function on $[0,T]$, we have*

$$\left\{\begin{array}{l} \bar u'^{+}_{\tau,j} = \displaystyle\sum_{k \in \mathcal{J}_\sigma^+} (\bar u_{j,k+1} - \bar u_{j,k})\delta_{t_k} \text{ with } \mathcal{J}_\sigma^+ = \{k \in \{1, \dots, N_\tau - 1\} : \bar\Phi_{\sigma,j}(t_k) = +\alpha_j\}, \\ \bar u'^{-}_{\tau,j} = \displaystyle\sum_{k \in \mathcal{J}_\sigma^-} (\bar u_{j,k+1} - \bar u_{j,k})\delta_{t_k} \text{ with } \mathcal{J}_\sigma^- = \{k \in \{1, \dots, N_\tau - 1\} : \bar\Phi_{\sigma,j}(t_k) = -\alpha_j\}, \end{array}\right.$$

where $\bar u'_{\tau,j} = \bar u'^{+}_{\tau,j} - \bar u'^{-}_{\tau,j}$ is the Jordan decomposition of the measure $\bar u'_{\tau,j}$.

Proof. The proof of this result is a consequence of the representation formula for $\bar u'_\tau$ given in (4.2). In addition, we use $\frac{1}{\alpha_j}\bar\Phi_{\sigma,j} \in \partial g(\bar u'_{\tau,j})$ along with Proposition 2.4, and the fact that $\bar u'_{\tau,j} \ne 0$ by assumption. Finally, we take into account that $\Phi_{\sigma,j}$ is piecewise linear and continuous, and $\Phi_{\sigma,j}(0) = \Phi_{\sigma,j}(T) = 0$. Consequently its maximal and minimal values are attained at the interior grid points $\{t_k\}_{k=1}^{N_\tau - 1}$. ☐

5. Convergence Analysis. The goal of this section is to prove the convergence of solutions of (P_σ) to solutions of (P) as $\sigma \to 0$. Additionally we give some error estimates for the difference between the optimal discrete and continuous states.

THEOREM 5.1. *Let us assume that either f is affine with respect to y or $\beta_j > 0$ for every $1 \le j \le m$, and let $\{\bar u_\tau\}_\tau \subset BV(0,T)^m$ be a family of global solutions of problems (P_σ), $\sigma = (h, \tau)$. Then this family is bounded in $BV(0,T)^m$. In addition, if f is affine or $n < 3$, then any weak* limit $\bar u$ of a subsequence when $\sigma \to 0$ is a global solution of (P). For such a subsequence we have*

$$\|\bar u'_\tau\|_{\mathcal{M}(0,T)^m} \to \|\bar u'\|_{\mathcal{M}(0,T)^m} \text{ and } \|\bar u - \bar u_\tau\|_{L^p(0,T)^m} \to 0\ \forall p \in [1, +\infty) \tag{5.1}$$

$$\|\bar y - \bar y_\sigma\|_{L^2(Q)} \to 0 \text{ and } J_\sigma(\bar u_\tau) \to J(\bar u) \tag{5.2}$$

where $\bar{y}$ and $\bar{y}_\sigma$ are the continuous and discrete states associated to $\bar{u}$ and $\bar{u}_\tau$, respectively.

For the proof we will use the following lemma.

LEMMA 5.2. *Let $d_\sigma \in L^2(Q)$ and take $y_\sigma \in \mathcal{Y}_\sigma$ solution of*

$$\begin{cases} \Big(\dfrac{y_{k,h} - y_{k-1,h}}{\tau_k}, z_h\Big) + a(y_{k,h}, z_h) + \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (f(\cdot, t, y_{k,h}), z_h)\, dt \\ \qquad\qquad = \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (d_\sigma(t), z_h)\, dt, \quad \forall z_h \in Y_h,\ 1 \le k \le N_\tau \\ y_{0,h} = y_{0h}. \end{cases} \tag{5.3}$$

Then, there exists a constant $C_\Omega > 0$ dependent only on Ω such that

$$\|y_\sigma\|_{L^\infty(0,T;L^2(\Omega))} + \|\nabla_x y_\sigma\|_{L^2(Q)} \le C_\Omega(\|d_\sigma\|_{L^2(Q)} + \|f(\cdot,\cdot,0)\|_{L^2(Q)} + \|y_{0h}\|_{L^2(\Omega)}). \tag{5.4}$$

Proof. The proof is standard, except for the nonlinear term. Choosing $z_h = y_{k,h}$ in (5.3), we obtain

$$(y_{k,h} - y_{k-1,h}, y_{k,h}) + \tau_k a(y_{k,h}, y_{k,h}) + \int_{I_k} (f(\cdot, t, y_{k,h}) - f(\cdot, t, 0), y_{k,h})\, dt$$
$$= \int_{I_k} (d_\sigma(t) - f(\cdot, t, 0), y_{k,h})\, dt.$$

Now, using the monotonicity of f with respect to y we deduce

$$(y_{k,h} - y_{k-1,h}, y_{k,h}) + \tau_k a(y_{k,h}, y_{k,h}) \le \int_{I_k} (d_\sigma(t) - f(\cdot, t, 0), y_{k,h})\, dt.$$

The rest of the proof can be completed as in the standard linear case. □

Proof of Theorem 5.1. Let us set

$$a_\tau = \frac{1}{\tau}\int_0^T \bar{u}_\tau\, dt \quad\text{and}\quad \hat{u}_\tau = \bar{u}_\tau - a_\tau.$$

Let $\hat{y}_\tau$ be the discrete state associated with $\hat{u}_\tau$. The proof is divided into three steps.

Step 1. $\{\bar{y}_\sigma\}_\sigma$ and $\{\bar{u}_\tau\}_\tau$ are bounded in $L^2(Q)$ and $BV(0,T)^m$.

From the global optimality of $\bar{u}_\tau$ we have that $J_\sigma(\bar{u}_\tau) \le J_\sigma(0)$ for every σ. From Lemma 5.2, we obtain that the discrete states y_σ associated to 0 are uniformly bounded in $L^2(Q)$. Hence, $\{J_\sigma(0)\}_\sigma$ is bounded and consequently $\{\bar{y}_\sigma\}_\sigma$ and $\{\bar{u}'_\tau\}_\tau$ are bounded in $L^2(Q)$ and $\mathcal{M}(0,T)^m$, respectively. According to (2.1), it is enough to prove the boundedness of $\{a_\tau\}_\tau$ in $\mathbb{R}^m$ to conclude the boundedness of $\{\bar{u}_\tau\}_\tau$ in $BV(0,T)^m$. This is obvious if $\beta_j > 0$ for $1 \le j \le m$. Otherwise, by assumption we have that $f = c_0 y + d_0$ with $c_0 \ge 0$, $c_0 \in L^\infty(Q)$, and $d_0 \in L^{\hat{p}}(Q)$.

Let us put $z_\sigma = \bar{y}_\sigma - \hat{y}_\sigma$. Using again (2.1) we get that $\{\hat{u}_\tau\}_\tau$ is bounded $BV(0,T)^m \subset L^2(Q)^m$. Then, Lemma 5.2 implies the boundedness of $\{\hat{y}_\sigma\}_\sigma$ in $L^2(Q)$. Thus, we also have the boundedness of $\{z_\sigma\}_\sigma$ in $L^2(Q)$. Subtracting the discrete equations satisfied by $\bar{y}_\sigma$ and $\hat{y}_\sigma$ we get

$$\begin{cases} \Big(\dfrac{z_{k,h} - z_{k-1,h}}{\tau_k}, z_h\Big) + a(z_{k,h}, z_h) + \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (c_0(\cdot, t) z_{k,h}, z_h)\, dt \\ \qquad\qquad = \displaystyle\sum_{j=1}^m (g_j, z_h) a_{\tau,j} \quad \forall z_h \in Y_h,\ 1 \le k \le N_\tau \\ z_{0,h} = 0, \end{cases} \tag{5.5}$$

where $a_\tau = (a_{\tau,j})_{j=1}^m$. We argue by contradiction and we assume that

$$\rho_\tau = \max_{1\le j\le m} |a_{\tau,j}| \to +\infty \quad \text{as} \quad k \to \infty.$$

Then, introducing $\zeta_\sigma = \frac{1}{\rho_\tau} z_\sigma$ and $\bar a_{\tau,j} = \frac{a_{\tau,j}}{\rho_\tau}$, we deduce from (5.5)

$$\begin{cases} \Big(\dfrac{\zeta_{k,h} - \zeta_{k-1,h}}{\tau_k}, z_h\Big) + a(\zeta_{k,h}, z_h) + \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (c_0(\cdot,t)\zeta_{k,h}, z_h)\,dt \\ \qquad\qquad = \displaystyle\sum_{j=1}^m (g_j, z_h)\bar a_{\tau,j} \quad \forall z_h \in Y_h,\ 1 \le k \le N_\tau \\ \zeta_{0,h} = 0, \end{cases} \tag{5.6}$$

By taking a subsequence, that we denote in the same way, we can assume that $\bar a_{\tau,j} \to \bar a_j$ as $\tau \to 0$ for every $1 \le j \le m$, and $\bar a = (\bar a_j)_{j=1}^m \neq 0$. Let us denote by $\bar\zeta_\sigma$ the solution of (5.6) with $\bar a_\tau$ replaced by $\bar a$. From Lemma 5.2 we deduce that $\|\zeta_\sigma - \bar\zeta_\sigma\|_{L^2(Q)} \to 0$ as $\sigma \to 0$. Let $\bar\zeta \in H^{2,1}(Q)$ be the solution to

$$\begin{cases} \dfrac{\partial\bar\zeta}{\partial t}(x,t) - \Delta\bar\zeta(x,t) + c_0\bar\zeta &= \displaystyle\sum_{j=1}^m \bar a_j g_j & \text{in } Q = \Omega\times(0,T), \\ \bar\zeta(x,t) &= 0 & \text{on } \Sigma = \Gamma\times(0,T), \\ \bar\zeta(x,0) &= 0 & \text{in } \Omega. \end{cases} \tag{5.7}$$

From Proposition 4.2, we deduce that $\|\bar\zeta - \bar\zeta_\sigma\|_{L^2(Q)} \to 0$ as $\sigma \to 0$. By the boundedness of $\{z_\sigma\}_\sigma$ in $L^2(Q)$ and the definition of ζ_σ we deduce that $\zeta_\sigma \to 0$ in $L^2(Q)$. Hence, $\bar\zeta_\sigma = \zeta_\sigma + (\bar\zeta_\sigma - \zeta_\sigma) \to 0$ in $L^2(Q)$ as well. This implies that $\bar\zeta = 0$ and consequently $\sum_{j=1}^m \bar a_j g_j = 0$. From our assumptions on $\{g_j\}_{j=1}^m$, this implies that $\bar a = 0$, which gives the desired contradiction. Therefore, $\{\bar u_\tau\}_\tau$ is bounded in $BV(0,T)^m$.

Let us take a subsequence of $\{\bar u_\tau\}_\tau$ denoted in the same way, such that $\bar u_\tau \stackrel{*}{\rightharpoonup} \bar u$ as $\sigma \to 0$.

Step 2. $\bar u$ is a global solution of (P)*, and* (5.1)-(5.2) *hold.*

The compactness of the embedding $BV(0,T) \subset L^p(0,T)$ for every $p \in [1,+\infty)$ implies the strong convergence $\bar u_\tau \to \bar u$ in $L^p(0,T)^m$. Let us denote by $\bar y$ and $\bar u$ and $\hat y_\sigma$ the continuous and discrete states corresponding to $\bar u$. From Proposition 4.2 we get that $\hat y_\sigma \to \bar y$ in $L^2(Q)$ as $\sigma \to 0$. Subtracting the equations satisfied by $\bar y_\sigma$ and $\hat y_\sigma$ we get for $\zeta_\sigma = \bar y_\sigma - \hat y_\sigma$

$$\begin{cases} \Big(\dfrac{\zeta_{k,h} - \zeta_{k-1,h}}{\tau_k}, z_h\Big) + a(\zeta_{k,h}, z_h) + \dfrac{1}{\tau_k}\displaystyle\int_{I_k} (\partial_y f(\cdot,t,\xi_{k,h})\zeta_{k,h}, z_h)\,dt \\ \qquad\qquad = \displaystyle\sum_{j=1}^m (g_j, z_h)\frac{1}{\tau_k}\int_{I_k} (\bar u_{\tau,j} - \bar u_j)\,dt \quad \forall z_h \in Y_h,\ 1 \le k \le N_\tau \\ \zeta_{0,h} = 0, \end{cases} \tag{5.8}$$

where $\xi_{k,h}(x,t) = \hat y_{k,h} + \theta_{k,h}(x,t)\zeta_{k,h}$ with $0 \le \theta_{k,h}(x,t) \le 1$. In the case of an affine function f, we simply have $\partial_y f(x,t,\xi_{k,h}) = c_0(x,t)$. Arguing as in Lemma 5.2 and using that $\partial_y f \ge 0$ we get

$$\|\zeta_\sigma\|_{L^2(Q)} \le C_\Omega \max_{1\le j\le m} \|g_j\|_{L^\infty(Q)} \|\bar u - \bar u_\tau\|_{L^2(Q)^m} \to 0 \text{ as } \sigma \to 0.$$

Hence, $\bar{y}_\sigma = \hat{y}_\sigma + \zeta_\sigma \to \bar{y}$ in $L^2(Q)$. Now, the following relations hold

$$\begin{aligned}
J(\bar{u}) &\le F(\bar{u}) + \liminf_{\sigma\to 0} G(\bar{u}_\tau) \le F(\bar{u}) + \limsup_{\sigma\to 0} G(\bar{u}_\tau) \\
&= \lim_{\sigma\to 0} F_\sigma(\bar{u}_\tau) + \limsup_{\sigma\to 0} G(\bar{u}_\tau) = \limsup_{\sigma\to 0} J_\sigma(\bar{u}_\tau) \\
&\le \limsup_{\sigma\to 0} J_\sigma(\bar{u}) = J(\bar{u}) = F(\bar{u}) + G(\bar{u}).
\end{aligned}$$

As a consequence we have $G(\bar{u}) = \lim_{\tau\to 0} G(\bar{u}_\tau)$. Finally, taking into account that $\|\bar{u}'_j\|_{\mathcal{M}(0,T)} \le \liminf_{\tau\to 0} \|\bar{u}'_{\tau,j}\|_{\mathcal{M}(0,T)}$ for $1 \le j \le m$, we deduce $\|\bar{u}'_{\tau,j}\|_{\mathcal{M}(0,T)} \to \|\bar{u}'_j\|_{\mathcal{M}(0,T)}$ for $1 \le j \le m$. This completes the proof. □

The next theorem addresses the approximation of local solutions of (P) by local minima of (P_σ). It is in some sense a converse of previous theorem.

THEOREM 5.3. *Assume that either f is affine or $n < 3$ and let $\bar{u}$ be a strict $L^p(0,T)^m$-local minimum of* (P) *with $p \in [1,+\infty)$. Then there exist an $L^p(0,T)^m$-ball $B_\rho(\bar{u})$ such that J_σ has a global minimum $\bar{u}_\tau$ in $\bar{B}_\rho(\bar{u}) \cap BV(0,T)^m$ for every σ. The family $\{\bar{u}_\tau\}_\tau$ converges to $\bar{u}$ in the sense of* (5.1)-(5.2). *Consequently, there exists σ_0 such that $\bar{u}_\tau$ is a local solution of* (P_σ) *for every $|\sigma| \le |\sigma_0|$.*

Proof. Since $\bar{u}$ is a strict $L^p(0,T)^m$-local minimum of (P), there exists $\rho > 0$ such that

$$J(\bar{u}) < J(u) \quad \forall u \in \bar{B}_\rho(\bar{u}) \setminus \{\bar{u}\} \tag{5.9}$$

We consider the problems

$$(\mathrm{P}_{\sigma,\rho}) \ \min\{J_\sigma(u) : u \in BV(0,T)^m \cap \bar{B}_\rho(\bar{u})\}.$$

The existence of at least one solution $\bar{u}_\tau$ for ($\mathrm{P}_{\sigma,\rho}$), $\sigma = (h,\tau)$, is obvious. Now, we can argue as in the proof of the previous theorem to deduce that $\{\bar{u}_\tau\}_\tau$ has converging subsequences and any of these limits is a solution of the problem

$$(\mathrm{P}_\rho) \ \min\{J(u) : u \in BV(0,T)^m \cap \bar{B}_\rho(\bar{u})\}.$$

Since $\bar{u}$ is the unique solution of (P_ρ), it follows that the whole family $\{\bar{u}_\tau\}_\tau$ converges to $\bar{u}$ in the sense of (5.1) and (5.2). Due to the convergence $\|\bar{u} - \bar{u}_\tau\|_{L^p(0,T)^m} \to 0$, we deduce the existence of σ_0 such that $\bar{u}_\tau \in B_\rho(\bar{u})$ for every $|\sigma| \le |\sigma_0|$, and hence $\bar{u}_\tau$ is a local minimum of (P_σ) in the ball $B_\rho(\bar{u})$. □

The rest of the section is devoted to the analysis of the rate of convergence for the states $\|\bar{y} - \bar{y}_\sigma\|_{L^2(Q)}$. Let $\bar{u}$ be a local solution of (P) such that the sufficient second order conditions (SSOC) (3.15) holds. Theorem 3.8 implies that $\bar{u}$ is a strict strong local solution, and hence it is a strict $L^p(0,T)^m$-local solution as well. Let $\rho > 0$ such that $\bar{u}$ is a global minimum of J in $\bar{B}_\rho(\bar{u}) \cap BV(0,T)^m$. Let $\{\bar{u}_\tau\}_\tau$ be a family of global minima of J_σ on $\bar{B}_\rho(\bar{u}) \cap BV(0,T)^m$ converging to $\bar{u}$ in $L^p(0,T)^m$, for $p > 1$. Then we have the following rate of convergence of the associated states.

THEOREM 5.4. *Let us assume that $\bar{u}$ satisfies the (SSOC) and that either f is affine or $n < 3$ holds. Then, under the above notations, there exists $C > 0$ independent of σ such that*

$$\|\bar{y} - \bar{y}_\sigma\|_{L^2(Q)} \le C(\sqrt{\tau} + h). \tag{5.10}$$

Proof. Since $\bar u_\tau \to \bar u$ in $L^p(0,T)^m$ with $p > 1$, we have that $\|y_{\bar u_\tau} - \bar y\|_{L^\infty(Q)} \to 0$ as $\sigma \to 0$, where $y_{\bar u_\tau}$ is the continuous state corresponding to $\bar u_\tau$. Let $\epsilon > 0$ be as introduced in Corollary 3.4. Then there exists σ_ε such that $\|y_{\bar u_\tau} - \bar y\|_{L^\infty(Q)} < \varepsilon$ for every $|\sigma| \le |\sigma_\varepsilon|$. Now, utilizing (3.17) we have

$$\begin{aligned}&\frac{\delta}{2}\|y_{\bar u_\tau} - \bar y\|^2_{L^2(Q)} \le J(\bar u_\tau) - J(\bar u)\\ &= [J(\bar u_\tau) - \hat J_\sigma(\bar u_\tau)] + [\hat J_\sigma(\bar u_\tau) - \hat J_\sigma(\bar u)] + [\hat J_\sigma(\bar u) - J(\bar u)],\end{aligned} \tag{5.11}$$

where

$$\hat J_\sigma(u) = \frac12\|y_\sigma(u) - y_d\|^2_{L^2(Q)} + \sum_{j=1}^M \frac{\beta_j}{2}\Big(\int_0^T u_j(t)\,dt\Big)^2 + G(u).$$

Let us estimate these terms. For the first term we use Proposition 4.2 as follows

$$\begin{aligned}&J(\bar u_\tau) - \hat J_\sigma(\bar u_\tau) = \frac12\|y_{\bar u_\tau} - y_d\|^2_{L^2(Q)} - \frac12\|\bar y_\sigma - y_d\|^2_{L^2(Q_h)}\\ &\le C_1\|y_{\bar u_\tau} - \bar y_\sigma\|_{L^2(Q)} \le C_2(\tau + h^2).\end{aligned}$$

The third term is estimated in the same way, and for the second it is enough to observe

$$\hat J_\sigma(\bar u_\tau) - \hat J_\sigma(\bar u) = J_\sigma(\bar u_\tau) - J_\sigma(\bar u) \le 0,$$

the last inequality being consequence of the fact that J_σ achieves the minimum value in the ball $B_\rho(\bar u) \cap BV(0,T)^m$ at $\bar u_\tau$. All together this leads to

$$\|y_{\bar u_\tau} - \bar y\|_{L^2(Q)} \le C_3(\sqrt{\tau} + h).$$

Finally

$$\|\bar y - \bar y_\sigma\|_{L^2(Q)} \le \|\bar y - y_{\bar u_\tau}\|_{L^2(Q)} + \|y_{\bar u_\tau} - \bar y_\sigma\|_{L^2(Q)} \le C_3(\sqrt{\tau} + h) + C_4(\tau + h^2),$$

where we have used again Proposition 4.2. ☐

Remark 5.5. *Under the assumptions of the above theorem, and supposing that* $y_d \in L^2(0,T;L^4(\Omega))$, *and using Proposition 4.2, we can argue as in [4, Theorem 5.1] to deduce that* $|J(\bar u) - J_\sigma(\bar u_\tau)| \le C(\sqrt{\tau} + h)$.

6. Numerical Solution. In this section we show how (P_σ) can be solved numerically. We consider the case of a linear state equation with zero state at the initial time, i.e., $f \equiv 0$ and $y_0 \equiv 0$ in (1.1).

6.1. A fully discrete formulation. Defining $y_{d,\sigma}$ as the $L^2(Q_h)$ projection of y_d onto $\mathcal{Y}_\sigma$ we have

$$\|y_\sigma - y_d\|^2_{L^2(Q_h)} = \|y_\sigma - y_{d,\sigma}\|^2_{L^2(Q_h)} + \|y_{d,\sigma} - y_d\|^2_{L^2(Q_h)},$$

hence (P_σ) is equivalent to

$$\min_{u\in BV(0,T)^m} \frac12\|y_\sigma - y_{d,\sigma}\|^2_{L^2(Q_h)} + \sum_{j=1}^m\left(\alpha_j\|u_j'\|_{\mathcal{M}(0,T)} + \frac{\beta_j}{2}\Big(\int_0^T u_j(t)\,dt\Big)^2\right).$$

Therefore, Theorem 4.5 guarantees that we can find a solution for (P_σ) by solving

$$(\mathrm{Q}_\sigma) \qquad \min_{u_\tau \in U_\tau^m} \frac{1}{2}\|y_\sigma - y_{d,\sigma}\|^2_{L^2(Q_h)} + \sum_{j=1}^m \left(\alpha_j \|u'_{\tau,j}\|_{\mathcal{M}(0,T)} + \frac{\beta_j}{2}\Big(\int_0^T u_{\tau,j}(t)\,dt\Big)^2 \right).$$

We now reformulate (Q_σ) as an optimization problem in $\mathbb{R}^{mN_\tau}$.

By definition, $u_\tau \in U_\tau$ has the representation $u_\tau = \sum_{k=1}^{N_\tau} u_k \chi_k$ with coefficient vector $\hat{u}_\tau = (u_k)_{k=1}^{N_\tau} \in \mathbb{R}^{N_\tau}$. Analogously, $u_\tau = (u_{\tau,1}, \ldots, u_{\tau,m}) \in U_\tau^m$ can be represented by its coefficient vector $\hat{u}_\tau = (u_{11}, \ldots, u_{1N_\tau}, u_{21}, \ldots, u_{mN_\tau})^T \in \mathbb{R}^{mN_\tau}$. Subsequently, let us denote $N_\rho = mN_\tau$. In view of (4.7) the coefficient vector of $y_\sigma \in \mathcal{Y}_\sigma$ is $\hat{y}_\sigma = (y_{11}, \ldots, y_{1N_h}, y_{21}, \ldots, y_{N_\tau N_h})^T \in \mathbb{R}^{N_\sigma}$, with $N_\sigma = N_\tau N_h$. Similarly, $y_{d,\sigma}$ has coefficient vector $\hat{y}_{d,\sigma}$. Denoting by $M_h \in \mathbb{R}^{N_h \times N_h}$ the mass matrix $M_h = ((e_i, e_j))_{i,j=1}^{N_h}$, by $A_h \in \mathbb{R}^{N_h \times N_h}$ the stiffness matrix $A_h = (a(e_i, e_j))_{i,j=1}^{N_h}$, and for $1 \le j \le m$ by $G_j \in \mathbb{R}^{N_h}$ the vector $G_j = ((g_j, e_k))_{k=1}^{N_h}$, the discrete state equation (4.8) can be expressed as $L_\sigma \hat{y}_\sigma = C\hat{u}_\tau$, where

$$L_\sigma = \begin{pmatrix} \tau_1^{-1} M_h + A_h & 0 & 0 \\ -\tau_2^{-1} M_h & \tau_2^{-1} M_h + A_h & 0 \\ 0 & \ddots & \ddots \end{pmatrix} \in \mathbb{R}^{N_\sigma \times N_\sigma},$$

$$C = \begin{pmatrix} C_1 & \ldots & C_m \end{pmatrix} \in \mathbb{R}^{N_\sigma \times N_\rho}$$

with

$$C_j = \begin{pmatrix} G_j & & \\ & \ddots & \\ & & G_j \end{pmatrix} \in \mathbb{R}^{N_\sigma \times N_\tau} \qquad \text{for } 1 \le j \le m.$$

Invoking (4.2) and introducing

$$M_\sigma = \begin{pmatrix} \tau_1 M_h & & \\ & \ddots & \\ & & \tau_{N_\tau} M_h \end{pmatrix} \in \mathbb{R}^{N_\sigma \times N_\sigma}$$

we obtain that (Q_σ) is equivalent to

$$\min_{\hat{u}_\tau \in \mathbb{R}^{N_\rho}} \Bigg[\frac{1}{2}\left(L_\sigma^{-1} C\hat{u}_\tau - \hat{y}_{d,\sigma}\right)^T M_\sigma \left(L_\sigma^{-1} C\hat{u}_\tau - \hat{y}_{d,\sigma}\right) + \sum_{j=1}^m \left(\alpha_j \sum_{k=2}^{N_\tau} |u_{jk} - u_{j(k-1)}| + \frac{\beta_j}{2}\Big(\sum_{k=1}^{N_\tau} \tau_k u_{jk}\Big)^2 \right) \Bigg].$$

Defining the vector $\hat{d}_\tau = (d_{11}, d_{12}, \ldots, d_{1N_\tau}, d_{21}, \ldots, d_{mN_\tau})^T \in \mathbb{R}^{N_\rho}$ by $d_{j1} = u_{j1}$ and $d_{jk} = u_{jk} - u_{j(k-1)}$ for $1 \le j \le m$ and $2 \le k \le N_\tau$, this problem becomes

$$\min_{\hat{d}_\tau \in \mathbb{R}^{N_\rho}} \Bigg[\frac{1}{2}\big(L_\sigma^{-1} CT\hat{d}_\tau - \hat{y}_{d,\sigma}\big)^T M_\sigma \big(L_\sigma^{-1} CT\hat{d}_\tau - \hat{y}_{d,\sigma}\big) + \sum_{j=1}^m \left(\alpha_j \sum_{k=2}^{N_\tau} |d_{jk}| + \frac{\beta_j}{2}\big(v_\tau^T T_\tau \hat{d}_{\tau,j}\big) \cdot \big(v_\tau^T T_\tau \hat{d}_{\tau,j}\big) \right) \Bigg],$$

where

$$T = \begin{pmatrix} T_\tau & & \\ & \ddots & \\ & & T_\tau \end{pmatrix} \in \mathbb{R}^{N_\rho \times N_\rho} \qquad \text{with} \qquad T_\tau = \begin{pmatrix} 1 & & \\ \vdots & \ddots & \\ 1 & \dots & 1 \end{pmatrix} \in \mathbb{R}^{N_\tau \times N_\tau},$$

$v_\tau = (\tau_1, \dots, \tau_{N_\tau})^T \in \mathbb{R}^{N_\tau}$, and $\hat{d}_{\tau,j} = (d_{j1}, \dots, d_{jN_\tau})^T \in \mathbb{R}^{N_\tau}$. Employing the further definitions $S = L_\sigma^{-1} CT$,

$$Q = \hat{Q}^T \hat{Q} \qquad \text{with} \qquad \hat{Q} = \begin{pmatrix} \sqrt{\beta_1} v_\tau^T T_\tau & & \\ & \ddots & \\ & & \sqrt{\beta_m} v_\tau^T T_\tau \end{pmatrix} \in \mathbb{R}^{m \times N_\rho},$$

$$\psi_{jk} : \mathbb{R}^{N_\rho} \to \mathbb{R}, \qquad \psi_{jk}(\hat{d}_\tau) = |d_{jk}|$$

for $1 \le j \le m$, $2 \le k \le N_\tau$, and

$$\Psi : \mathbb{R}^{N_\rho} \to \mathbb{R}, \qquad \Psi(\hat{d}_\tau) = \sum_{j=1}^{m} \alpha_j \sum_{k=2}^{N_\tau} \psi_{jk}(\hat{d}_\tau),$$

we arrive at the fully discrete optimization problem

$$(\mathrm{Q}_\rho) \qquad \min_{\hat{d}_\tau \in \mathbb{R}^{N_\rho}} J_\rho(\hat{d}_\tau) = \frac{1}{2}\big(S\hat{d}_\tau - \hat{y}_{d,\sigma}\big)^T M_\sigma \big(S\hat{d}_\tau - \hat{y}_{d,\sigma}\big) + \frac{1}{2}\hat{d}_\tau^T Q \hat{d}_\tau + \Psi(\hat{d}_\tau).$$

6.2. Discrete Optimality Conditions and Regularization. Since J_ρ is convex, $\hat{d}_\tau^* \in \mathbb{R}^{N_\rho}$ is optimal for (Q_ρ) if and only if $0 \in \partial J_\rho(\hat{d}_\tau^*)$. Since both the differentiable and the non-differentiable part of J_ρ are continuous, we obtain from the sum rule that $0 \in \partial J_\rho(\hat{d}_\tau^*)$ is equivalent to

$$0 \in S^T M_\sigma (S\hat{d}_\tau^* - \hat{y}_{d,\sigma}) + Q\hat{d}_\tau^* + \partial\Psi(\hat{d}_\tau^*),$$

where we have used that M_σ and Q are symmetric. Thus, $\hat{d}_\tau^*$ is optimal for (Q_ρ) if and only if there exists $\hat{\lambda}_\tau^* \in \mathbb{R}^{N_\rho}$ such that

$$S^T M_\sigma (S\hat{d}_\tau^* - \hat{y}_{d,\sigma}) + Q\hat{d}_\tau^* - \hat{\lambda}_\tau^* = 0 \qquad \text{and} \qquad -\hat{\lambda}_\tau^* \in \partial\Psi(\hat{d}_\tau^*). \tag{6.1}$$

The sum rule and the chain rule, cf. [10, Chapter I, Proposition 5.7], yield that $\partial\Psi(\hat{d}_\tau^*) \subset \mathbb{R}^{N_\rho}$ is given by

$$\partial\Psi(\hat{d}_\tau^*) = \{0\} \times \alpha_1 \partial\psi(d_{12}^*) \times \ldots \times \alpha_1 \partial\psi(d_{1N_\tau}^*) \times \{0\} \times \alpha_2 \partial\psi(d_{22}^*) \times \ldots \times \alpha_m \partial\psi(d_{mN_\tau}^*),$$

where $\psi : \mathbb{R} \to \mathbb{R}$ denotes $\psi(x) = |x|$. We recognize in $S^T M_\sigma (S\hat{d}_\tau^* - \hat{y}_{d,\sigma}) + Q\hat{d}_\tau^*$ the discrete version of $(\bar{\Phi}_j)_{j=1}^m$, cf. (3.4), which indicates that first-discretize-then-optimize and first-optimize-then-discretize coincide. To enable the use of semismooth Newton methods we proceed in two steps. The first step is to apply a regularization to (Q_ρ). More precisely, instead of (Q_ρ) we consider for $\gamma > 0$ the problem

$$(\mathrm{Q}_{\rho,\gamma}) \qquad \min_{\hat{d}_\tau \in \mathbb{R}^{N_\rho}} J_{\rho,\gamma}(\hat{d}_\tau) = \frac{1}{2}\big(S\hat{d}_\tau - \hat{y}_{d,\sigma}\big)^T M_\sigma \big(S\hat{d}_\tau - \hat{y}_{d,\sigma}\big) + \frac{1}{2}\hat{d}_\tau^T Q \hat{d}_\tau + \Psi_\gamma(\hat{d}_\tau),$$

where Ψ_γ is defined by

$$\Psi_\gamma : \mathbb{R}^{N_\rho} \to \mathbb{R}, \qquad \Psi_\gamma(\hat{d}_\tau) = \sum_{j=1}^{m} \alpha_j \sum_{k=2}^{N_\tau} \psi_{jk}^{\gamma}(\hat{d}_\tau)$$

with

$$\psi_{jk}^{\gamma} : \mathbb{R}^{N_\rho} \to \mathbb{R}, \qquad \psi_{jk}^{\gamma}(\hat{d}_\tau) = |d_{jk}| + \frac{\gamma}{2\tau_k}|d_{jk}|^2$$

for $1 \leq j \leq m$ and $2 \leq k \leq N_\tau$. Note that $(\mathrm{Q}_{\rho,\gamma})$ can be interpreted as the discrete counterpart of

$$\min_{u\in H^1(0,T)^m} \frac{1}{2}\|y_u - y_d\|_{L^2(Q)}^2 + \sum_{j=1}^{m}\Big(\alpha_j\big(\|u_j'\|_{L^1(0,T)} + \frac{\gamma}{2}\|u_j'\|_{L^2(0,T)}^2\big) + \frac{\beta_j}{2}\big(\int_0^T u_j(t)\,dt\big)^2\Big).$$

Since there holds $\|u_j'\|_{L^1(0,T)} = \|u_j'\|_{\mathcal{M}(0,T)}$ for this problem due to $u_j' \in L^1(0,T)$, this problem can be regarded as a regularized version of (P).

Arguing as above we obtain that $(\mathrm{Q}_{\rho,\gamma})$ has the optimality conditions (6.1), but with $\partial\Psi$ replaced by $\partial\Psi_\gamma$. In addition, $\partial\Psi_\gamma$ has the same structure as $\partial\Psi$, but with $\partial\psi$ in the component jk replaced by $\partial\psi_\gamma^k$, where $\psi_\gamma^k : \mathbb{R} \to \mathbb{R}$ denotes $\psi_\gamma^k(x) = |x| + \frac{\gamma}{2\tau_k}|x|^2$. In the second step, we rewrite $-\hat{\lambda}_\tau^* \in \partial\Psi_\gamma(\hat{d}_\tau^*)$ componentwise as $-\lambda_{jk}^*/\alpha_j \in \partial\psi_\gamma^k(d_{jk}^*)$ for $1 \leq j \leq m$ and $2 \leq k \leq N_\tau$, and replace each of these conditions equivalently by $d_{jk}^* \in \partial(\psi_\gamma^k)^*(-\lambda_{jk}^*/\alpha_j)$, where $(\psi_\gamma^k)^*$ denotes the conjugate function of ψ_γ^k, given by $(\psi_\gamma^k)^*(y) = \sup_{x\in\mathbb{R}}(yx - \psi_\gamma^k(x))$; cf. [10, Chapter I, Corollary 5.2]. Note that for $k = 1$ we keep the conditions $\lambda_{jk}^* = 0$. A straightforward computation reveals that $(\psi_\gamma^k)^*$ is the continuously differentiable function

$$(\psi_\gamma^k)^*(y) = \frac{\tau_k}{2\gamma} \cdot \begin{cases} (y+1)^2 & \text{if } y \leq -1, \\ 0 & \text{if } -1 < y < 1, \\ (y-1)^2 & \text{if } y \geq 1. \end{cases}$$

Therefore, the optimality conditions of $(\mathrm{Q}_{\rho,\gamma})$ can be recast as

$$S^T M_\sigma(S\hat{d}_\tau^* - \hat{y}_{d,\sigma}) + Q\hat{d}_\tau^* - \hat{\lambda}_\tau^* = 0, \qquad -\frac{\lambda_{j1}^*}{\alpha_j} = 0 \qquad \text{and} \qquad d_{jk}^* = \big((\psi_\gamma^k)^*\big)'\left(-\frac{\lambda_{jk}^*}{\alpha_j}\right)$$

for $1 \leq j \leq m$ and $2 \leq k \leq N_\tau$. This reads $F_\gamma(\hat{d}_\tau^*, \hat{\lambda}_\tau^*) = 0$ if we choose

$$F_\gamma : \mathbb{R}^{N_\rho} \times \mathbb{R}^{N_\rho} \to \mathbb{R}^{N_\rho} \times \mathbb{R}^{N_\rho}, \qquad F_\gamma(\hat{d}_\tau, \hat{\lambda}_\tau) = \begin{pmatrix} S^T M_\sigma(S\hat{d}_\tau - \hat{y}_{d,\sigma}) + Q\hat{d}_\tau + \hat{\lambda}_\tau^\alpha \\ F_{\gamma,1}(\hat{d}_\tau, \hat{\lambda}_\tau) \\ \vdots \\ F_{\gamma,m}(\hat{d}_\tau, \hat{\lambda}_\tau) \end{pmatrix},$$

where we have employed the definition $(\hat{\lambda}_\tau^\alpha)_{jk} = \alpha_j \lambda_{jk}$ for $1 \leq j \leq m$ and $1 \leq k \leq N_\tau$, as well as for $1 \leq j \leq m$ the mappings $F_{\gamma,j} : \mathbb{R}^{N_\rho} \times \mathbb{R}^{N_\rho} \to \mathbb{R}^{N_\tau}$ given by

$$F_{\gamma,j}(\hat{d}_\tau, \hat{\lambda}_\tau) = \gamma \begin{pmatrix} 0 \\ \frac{d_{j2}}{\tau_2} \\ \vdots \\ \frac{d_{jN_\tau}}{\tau_{N_\tau}} \end{pmatrix} - \begin{pmatrix} \lambda_{j1} \\ (\lambda_{j2}+1)^- + (\lambda_{j2}-1)^+ \\ \vdots \\ (\lambda_{jN_\tau}+1)^- + (\lambda_{jN_\tau}-1)^+ \end{pmatrix}.$$

Since F_γ is semismooth, we can apply a semismooth Newton method to solve $F_\gamma = 0$. For later reference we note that the Newton step $\hat{s}_\tau = (\hat{s}_d, \hat{s}_\lambda)$ at $(\hat{d}_\tau, \hat{\lambda}_\tau)$ solves

$$F'_\gamma(\hat{d}_\tau, \hat{\lambda}_\tau)\hat{s}_\tau = -F_\gamma(\hat{d}_\tau, \hat{\lambda}_\tau) \quad \text{with} \quad F'_\gamma(\hat{d}_\tau, \hat{\lambda}_\tau) = \begin{pmatrix} S^T M_\sigma S + Q & \operatorname{diag}(\hat{\alpha}) \\ \gamma \operatorname{diag}(w) & -\operatorname{diag}(e(\hat{\lambda}_\tau)) \end{pmatrix}. \tag{6.2}$$

Here, we have used $\hat{\alpha}, w, e(\hat{\lambda}_\tau) \in \mathbb{R}^{N_\rho}$, defined componentwise by $(\hat{\alpha})_{jk} = \alpha_j$,

$$(w)_{jk} = \begin{cases} 0 & \text{if } k = 1, \\ \frac{1}{\tau_k} & \text{if } k \neq 1, \end{cases} \qquad \text{and} \qquad (e(\hat{\lambda}_\tau))_{jk} = \begin{cases} 1 & \text{if } k = 1, \\ 0 & \text{if } k \neq 1 \text{ and } \lambda_{jk} \in (-1,1), \\ 1 & \text{if } k \neq 1 \text{ and } \lambda_{jk} \notin (-1,1) \end{cases}$$

for $1 \leq j \leq m$ and $1 \leq k \leq N_\tau$.

6.3. Path-Following Algorithm. Since we have approximated (Q_ρ) by $(\mathrm{Q}_{\rho,\gamma})$, we consider a path-following algorithm that drives γ to zero.

Algorithm BV: Path-following method to solve (Q_ρ)

Input: $\hat{d}^0_\tau, \hat{\lambda}^0_\tau \in \mathbb{R}^{N_\rho}$, $\gamma_0 > 0$, $\mathrm{TOL}_\gamma > 0$, $\mathrm{TOL_F} > 0$, $\nu \in (0,1)$.
Set $k = 0$.
while $\gamma_k > \mathrm{TOL}_\gamma$ **do**
  Set $i = 0$ and $(\tilde{d}^0_\tau, \tilde{\lambda}^0_\tau) = (\hat{d}^k_\tau, \hat{\lambda}^k_\tau)$.
  while $\|F_{\gamma_k}(\tilde{d}^i_\tau, \tilde{\lambda}^i_\tau)\|_{L^2(0,T)^{2m}} > \mathrm{TOL_F}$ **do**
    Compute the Newton step $(\tilde{s}^i_d, \tilde{s}^i_\lambda)$ at $(\tilde{d}^i_\tau, \tilde{\lambda}^i_\tau)$ according to (6.2).
    Define $(\tilde{d}^{i+1}_\tau, \tilde{\lambda}^{i+1}_\tau) = (\tilde{d}^i_\tau + \tilde{s}^i_d, \tilde{\lambda}^i_\tau + \tilde{s}^i_\lambda)$; set $i = i+1$.
  end
  Define $(\hat{d}^{k+1}_\tau, \hat{\lambda}^{k+1}_\tau) = (\tilde{d}^i_\tau, \tilde{\lambda}^i_\tau)$ and $\gamma_{k+1} = \nu\gamma_k$; set $k = k+1$.
end
Output: $(\hat{d}^k_\tau, \hat{\lambda}^k_\tau)$.

In Algorithm BV we have used the definition $\|\hat{v}\|_{L^2(0,T)^{2m}} = \sum_{j=1}^{2m}(\sum_{k=1}^{N_\tau}\tau_k v_{jk}^2)^{\frac{1}{2}}$ for $\hat{v} = (v_{11}, \ldots, v_{1N_\tau}, v_{21}, \ldots, v_{(2m)N_\tau}) \in \mathbb{R}^{N_\rho} \times \mathbb{R}^{N_\rho}$.

Several variants of this algorithm are conceivable. For instance, a damping strategy could be included, $\mathrm{TOL_F}$ could depend on γ_k, and ν could vary with k.

Regarding the convergence behavior of Algorithm BV we point out that the semismooth Newton method for F_γ converges locally at a q-superlinear rate to the unique solution of $(\mathrm{Q}_{\rho,\gamma})$. To prove this it suffices to establish that $(\hat{d}_\tau, \hat{\lambda}_\tau) \mapsto \|F'_\gamma(\hat{d}_\tau, \hat{\lambda}_\tau)^{-1}\|$ is bounded, cf. [23, Proposition 2.12]. Using (6.2) it can be shown that F'_γ is invertible and that $\{F'_\gamma(\hat{d}_\tau, \hat{\lambda}_\tau) : (\hat{d}_\tau, \hat{\lambda}_\tau) \in \mathbb{R}^{N_\rho} \times \mathbb{R}^{N_\rho}\} \subset \mathbb{R}^{2N_\rho \times 2N_\rho}$ contains only a finite number of elements. This implies, in particular, the asserted boundedness.

7. Numerical Examples. We illustrate our findings by three examples. Our goal is to exemplify the structure of optimal controls for (P). Throughout, we treat the case where $f \equiv 0$, $\beta_j = 0$ for all j, and $y_0 \equiv 0$. In particular, (P) is convex and Theorem 3.1 yields the existence of a unique and global optimal solution.

In all examples we consider controls defined on $(0,T) = (0,2)$ and employ uniformly spaced temporal and spatial grids. We found $\gamma_0 = 1$, $\mathrm{TOL_F} = 10^{-12}m$, $\mathrm{TOL}_\gamma = 10^{-12}$, and $\nu = 0.1$ to be reliable choices in Algorithm BV. We use $\hat{d}^0_\tau = 0$ and take $\hat{\lambda}_\tau^0$ such that $(\hat{d}_\tau^0, \hat{\lambda}_\tau^0)$ satisfies the condition $S^T M_\sigma(S\hat{d}_\tau - \hat{y}_{d,\sigma}) + \hat{\lambda}_\tau^\alpha = 0$ in the optimality system $F_\gamma = 0$. When γ_k reaches TOL_γ, the inner while loop in Algorithm BV is executed until $|\|F_{\gamma_k}(\tilde{d}_\tau^i, \tilde{\lambda}_\tau^i)\|_{L^2(0,T)^{2m}} - \|F_{\gamma_k}(\tilde{d}_\tau^{i-1}, \tilde{\lambda}_\tau^{i-1})\|_{L^2(0,T)^{2m}}| \le \mathrm{TOL_F}$ and $\|F_{\gamma_k}(\tilde{d}_\tau^i, \tilde{\lambda}_\tau^i)\|_{L^2(0,T)^{2m}} \le \mathrm{TOL_F}$ are satisfied for three consecutive i. We use GMRES to solve the nonsymmetric linear system (6.2) to an accuracy of 10^{-12}. Each iteration of GMRES requires to solve two PDEs with differential operator L_σ. These PDE solves are performed to an accuracy of 10^{-12} using preconditioned GMRES.

7.1. Example 1: One control and one spatial dimension. We start with an example in which $m = 1$, $\Omega = (-1, 1)$, and $\omega = (0, 1)$. The remaining specifications are made such that an exact analytic solution $\bar{u}$ of (P) is known. The optimal control $\bar{u}$ exhibits $l \in \mathbb{N}$ jumps and it is constant apart from these jumps. Consider

$$\min_{u \in BV(0,T)} \frac{1}{2}\|y_u - y_d\|_{L^2(Q)}^2 + \bar{\alpha}\|u'\|_{\mathcal{M}(0,T)},$$

where y_u is the solution to the parabolic state equation

$$\begin{cases} \dfrac{\partial y}{\partial t}(x,t) - \Delta y(x,t) &= ug \quad \text{in } Q = (-1,1) \times (0,2), \\ y(-1,t) = y(1,t) &= 0 \quad \text{on } (0,2), \\ y(x,0) &= 0 \quad \text{in } (-1,1). \end{cases} \tag{7.1}$$

We take $g \equiv 1$ in ω and $g \equiv 0$ elsewhere, i.e., $g = \chi_\omega$. Let $\kappa > 0$, $l \in \mathbb{N}$, and $c_k \ge 0$ for $1 \le k \le l$. Define

$$\bar{\alpha} = \frac{4\kappa}{l\pi^2}, \qquad \bar{\varphi}(x,t) = \kappa \sin(l\pi t)\cos\Big(\frac{\pi}{2}x\Big), \qquad \bar{u} = \begin{cases} 0 & \text{if } t < \frac{1}{l}, \\ c_1 & \text{if } \frac{1}{l} < t < \frac{3}{l}, \\ \vdots & \qquad \vdots \\ \sum_{k=1}^l c_k & \text{if } \frac{2l-1}{l} < t < 2. \end{cases}$$

In particular, this implies $\bar{u}' = \sum_{k=1}^l c_k \delta_{\frac{2k-1}{l}}$ and $\|\bar{u}'\|_{\mathcal{M}(0,T)} = \sum_{k=1}^l c_k$. Denoting by L the differential operator $\frac{\partial}{\partial t} - \Delta$ we set $y_d = \bar{y} - L^*\bar{\varphi}$, where $\bar{y} = y_{\bar{u}}$. To conclude that $\bar{u}$ is the optimal solution of the above optimization problem, we check if $\bar{u}$ satisfies the necessary optimality conditions of Theorem 3.3. Since we are dealing with a convex problem, this is already sufficient for global optimality. Alternatively, the optimality of $\bar{u}$ can be established using the conditions from Theorem 3.8, in particular the condition (SSOC). Considering the first order conditions from Theorem 3.3 we first note that the adjoint equation $L^*\varphi_{\bar{u}} = y_{\bar{u}} - y_d$ together with boundary conditions is satisfied by construction. Second, we confirm that

$$\bar{\Phi}(t) = \int_0^t \int_\omega \bar{\varphi}(x,s)\,dx\,ds = \frac{2\kappa}{l\pi^2}\big(1 - \cos(l\pi t)\big) = \frac{\bar{\alpha}}{2}\big(1 - \cos(l\pi t)\big),$$

which demonstrates $\bar{\Phi} \in C^1[0,2] \cap C_0(0,2)$ and $\bar{\Phi}(t) \in [0, \bar{\alpha}]$ for all $t \in [0,2]$, with $\bar{\Phi}(t) = \bar{\alpha}$ exactly for $t = \frac{2k-1}{l}$ with $1 \le k \le l$. Hence, we have $\|\bar{\Phi}\|_{C_0(0,T)} = \bar{\alpha}$ and

$$\int_0^T \bar{\Phi}\,d\bar{u}' = \sum_{k=1}^l c_k \bar{\Phi}\Big(\frac{2k-1}{l}\Big) = \bar{\alpha}\sum_{k=1}^l c_k = \|\bar{\Phi}\|_{C_0(0,T)}\|\bar{u}'\|_{\mathcal{M}(0,T)},$$

which establishes (3.5) and (3.6). Thus, $\bar{u}$ is optimal. In view of Corollary 3.4 we note

$$\begin{cases} \operatorname{supp}(\bar{u}'^{+}) = \operatorname{supp}(\bar{u}') \subset \{t \in [0, T] : \bar{\Phi}(t) = \bar{\alpha}\}, \\ \operatorname{supp}(\bar{u}'^{-}) = \{t \in [0, T] : \bar{\Phi}(t) = -\bar{\alpha}\} = \emptyset, \end{cases}$$

where the inclusion is an equality if and only if all c_k are positive. Since we have

$$J(\bar{u}) = \frac{1}{2}\|y_{\bar{u}} - y_d\|^2_{L^2(Q)} + \bar{\alpha}\|\bar{u}'\|_{\mathcal{M}(0,T)} = \frac{1}{2}\|L^*\bar{\varphi}\|^2_{L^2(Q)} + \frac{4\kappa}{l\pi^2}\sum_{k=1}^{l} c_k$$

and we easily compute $\|L^*\bar{\varphi}\|^2_{L^2(Q)} = \kappa^2\pi^2(l^2 + \frac{\pi^2}{16})$, the optimal value is given by

$$J(\bar{u}) = \frac{\kappa^2\pi^2}{2}\left(l^2 + \frac{\pi^2}{16}\right) + \frac{4\kappa}{l\pi^2}\sum_{k=1}^{l} c_k.$$

For the numerical experiments we take $\bar{y}$ as solution of the state equation computed on a grid with $N_t = 10240$ and $N_h = 511$ nodes. This is possible since we know $\bar{u}$. We also use this $\bar{y}$ to obtain $y_{d,\sigma}$. We choose $l = 5$, $\kappa = 0.01$, $c_1 = c_3 = c_5 = 2$, and $c_2 = c_4 = 1$, which yields $\bar{\alpha} = 1/(125\pi^2) \approx 8.1 \cdot 10^{-4}$ and $J(\bar{u}) \approx 1.9 \cdot 10^{-2}$. Furthermore, it implies that $\bar{u}$ exhibits five jumps, which occur exactly at those t where $\bar{\Phi}(t) = \bar{\alpha}$. Unless indicated otherwise we employ $N_t = 1000$ and $N_h = 159$, which corresponds to $\tau = 0.002$ and $h = 0.0125$. Application of Algorithm BV yields $\bar{y}_\sigma$, $\bar{u}_\tau$, and the optimal dual variable $\bar{\lambda}_\tau$, which can be interpreted as discretization of $\bar{\lambda} = \frac{1}{\bar{\alpha}}\bar{\Phi} = \frac{1}{2}(1 - \cos(5\pi t))$. These quantities—more precisely, linear interpolations of them—are depicted together with $y_{d,\sigma}$ in Figure 7.1. We observe that $\bar{u}_\tau$ and $\bar{\lambda}_\tau$ resemble closely their continuous counterparts $\bar{u}$ and $\bar{\lambda}$. In particular, $\bar{u}_\tau$ clearly displays the five distinct jumps of $\bar{u}$.

To assess the discretization error we apply Algorithm BV on different grids, where each grid satisfies $N_\tau = 10((N_h + 1)/16)^2$. We use $N_h + 1 = 80, 96, 112, \ldots, 240$. The resulting errors $\|\bar{y} - \bar{y}_\sigma\|_{L^2(Q)}$ and $|J(\bar{u}) - J_\sigma(\bar{u}_\tau)|$ are plotted in Figure 7.2. For the error of the state we observe approximately linear convergence, which agrees with Theorem 5.4. The optimal objective value appears to converge superlinearly. This is better than we would expect from Remark 5.5, but matches previous observations and results for optimal control problems involving measures, cf. [3, 4, 14, 17].

Next we investigate the influence of α on solutions of (P). For this purpose we continue to work with $l = 5$, $\kappa = 0.01$, $c_1 = c_3 = c_5 = 2$, and $c_2 = c_4 = 1$. In particular, we keep the corresponding y_d. However, instead of $\bar{\alpha} = 1/(125\pi^2)$ we use

$$\alpha_\theta = \theta\bar{\alpha} \quad \text{with} \quad \theta \in [10^{-3}, 10^2]$$

in the objective. We stress that for $\theta \neq 1$ we do not know exact solutions of (P). Employing $L^*\bar{\varphi} = \kappa(\frac{\pi^2}{4}\sin(l\pi t)\cos(\frac{\pi}{2}x) - l\pi\cos(l\pi t)\cos(\frac{\pi}{2}x))$ it follows from the definition that y_d does not satisfy the initial condition $y(x, 0) \equiv 0$ of the state equation. This implies $\bar{y} \neq y_d$ regardless of the value of θ. Figure 7.3, Figure 7.4 and Figure 7.5 show $\bar{y}_\sigma = \bar{y}^\theta_\sigma$, $\bar{u}_\tau = \bar{u}^\theta_\tau$ and $\bar{\lambda}_\tau = \bar{\lambda}^\theta_\tau$ for different values of θ. We observe that $\bar{u}^\theta_\tau$ is constant for $\theta = 100$. Although not depicted, this is also true for every $\theta > 100$ that we tested. Hence, in accordance with Remark 3.5 the optimal control is constant for sufficiently large values of α. As θ decreases, the number of jumps of $\bar{u}^\theta_\tau$ increases. For $\theta < 1$ jumps with negative height occur. Approximately around $\theta = 0.1$ the measures

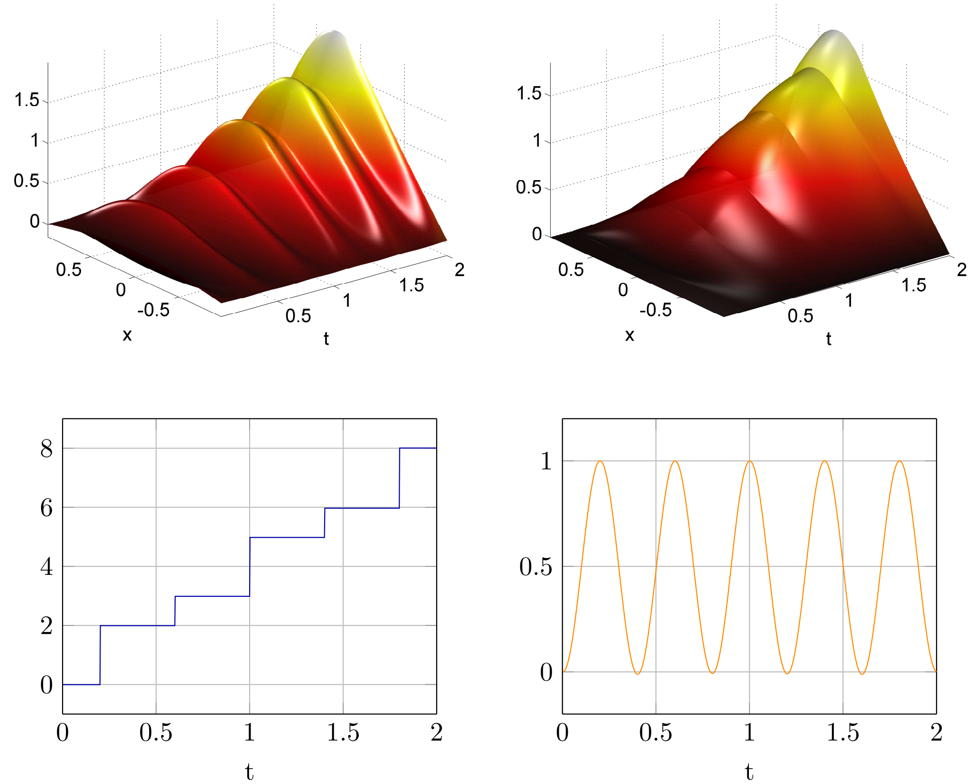

FIG. 7.1. *Example 1:* $y_{d,\sigma}$ *and* $\bar{y}_\sigma$ *(top left and right),* $\bar{u}_\tau$ *and* $\bar{\lambda}_\tau$ *(bottom left and right).*

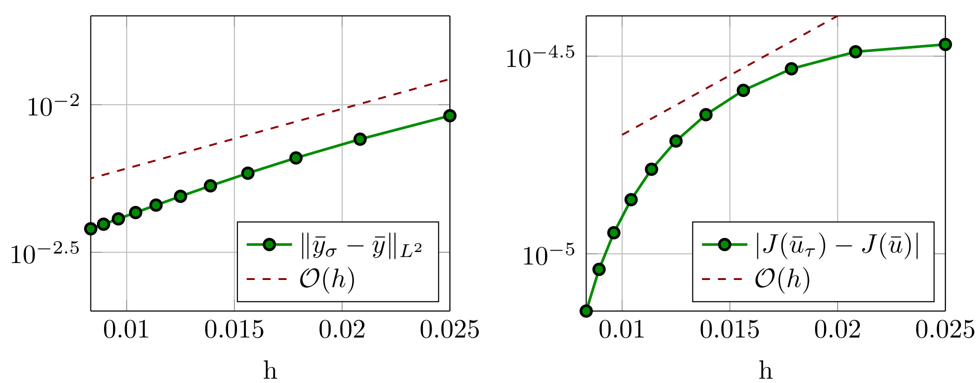

FIG. 7.2. *Example 1: Discretization errors for optimal state and optimal objective value.*

of $\operatorname{supp}((\bar{u}_\tau^\theta)')$ and $\{t \in (0,T) : \bar{\lambda}_\tau^\theta(t) = \pm 1\}$ become positive. As θ decreases further, these measures increase further.

To draw a comparison between (P) and the classical L^2-regularized tracking problem, we now replace $\alpha_\theta \|u'\|_{\mathcal{M}(0,T)}$ in the objective by $\frac{\alpha_\theta}{2}\|u\|^2_{L^2(0,T)}$. Recalling that

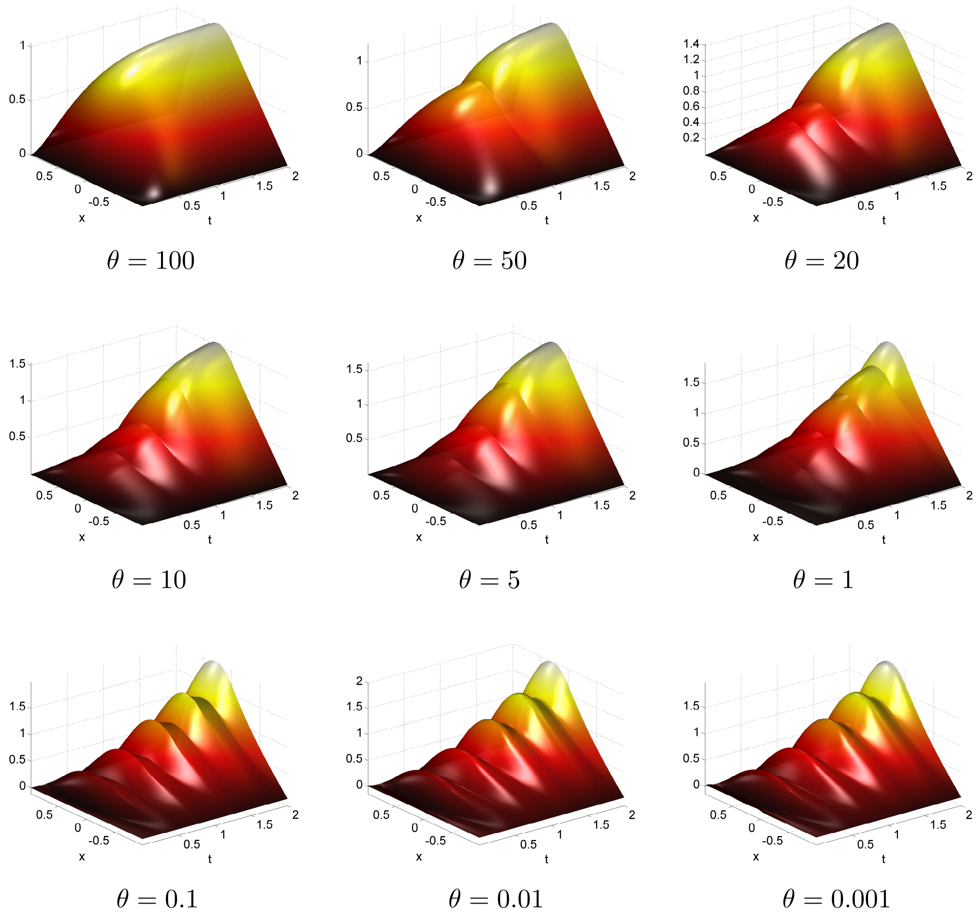

FIG. 7.3. *Example 1:* $\bar{y}_\sigma^\theta$ *for different values of* θ.

we have $m = 1$ in this example, the discretization of $\frac{\alpha_\theta}{2}\|u\|^2_{L^2(0,T)}$ is given by

$$\frac{1}{2}\hat{d}_\tau^T \tilde{Q}^T \tilde{Q} \hat{d}_\tau, \qquad \text{where} \qquad \tilde{Q} = \sqrt{\alpha_\theta}\,\mathrm{diag}(\sqrt{\tau_1}, \ldots, \sqrt{\tau_{N_\tau}})T_\tau \in \mathbb{R}^{N_\tau \times N_\tau}.$$

Figure 7.6 depicts the optimal controls $\bar{u}^\theta_{\tau,L^2}$ that we obtain for $\alpha_\theta = \theta\bar{\alpha}$ and various values of θ. Figure 7.7 shows the corresponding tracking errors $\frac{1}{2}\|\bar{y}^\theta_{\sigma,L^2} - y_{d,\sigma}\|^2_{L^2(0,T)}$ as well as the tracking errors for (P). It also displays the norms of the controls as they appear in the objective. The missing data point for the norm of the BV-control at $\theta = 100$ results from the fact that the corresponding control is constant, hence its BV-seminorm equals zero. We observe that the tracking errors for both control problems have a similar order of magnitude. From a practical point of view, however, the controls of (P) have a simpler structure. We note, in particular, that for $\theta \approx 5$ the tracking errors are approximately equal for the L^2 and BV-seminorm cases. The BV-control, however, is cheaper and also reproduces 4 jumps, whereas the L^2-control has a complicated structure.

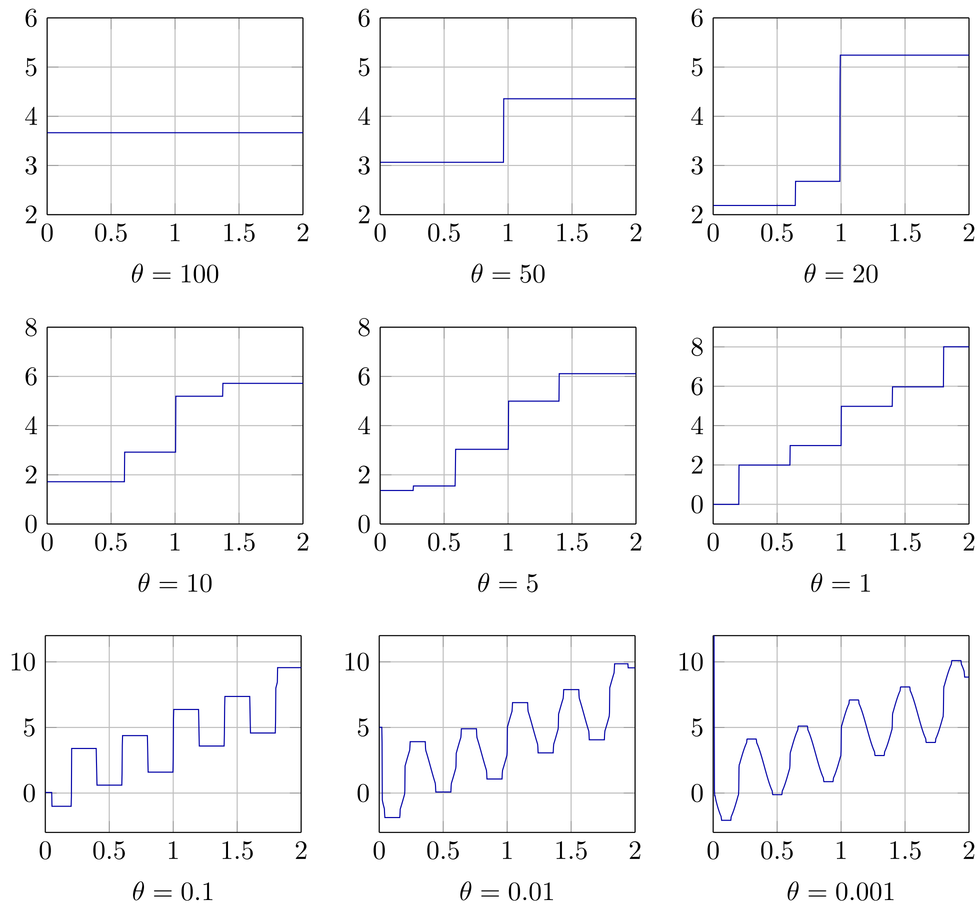

FIG. 7.4. *Example 1: $\bar{u}_\tau^\theta$ for different values of θ.*

7.2. Example 2: Three controls and one spatial dimension. The second example generalizes the first one. Here, we have $m \in \mathbb{N}$ controls, $\Omega = (-1, 1)$, and $\omega_j = (a_j, b_j)$ for $1 \le j \le m$ with $-1 \le a_1 < b_1 \le a_2 < b_2 \le \ldots \le a_m < b_m \le 1$. The following construction ensures that for every j the optimal control $\bar{u}_j$ has exactly $0 \le l_j \le m$ jumps and is constant apart from these jumps. We consider

$$\min_{u \in BV(0,T)^m} \frac{1}{2}\|y_u - y_d\|^2_{L^2(Q)} + \sum_{j=1}^{m} \alpha_j \|u_j'\|_{\mathcal{M}(0,T)},$$

where y_u denotes the solution to (7.1), but with ug replaced by $\sum_{j=1}^m u_j g_j$. We take $g_j = \chi_{\omega_j}$ for all j. Let $\kappa > 0$ and $c_{jk} \ge 0$ for $1 \le j, k \le m$. Define

$$\alpha_j = \frac{4\kappa}{m\pi^2}\left(\sin\Big(\frac{\pi}{2}b_j\Big) - \sin\Big(\frac{\pi}{2}a_j\Big)\right) \qquad \text{and} \qquad \bar{\varphi}(x,t) = \kappa \sin(m\pi t)\cos\Big(\frac{\pi}{2}x\Big),$$

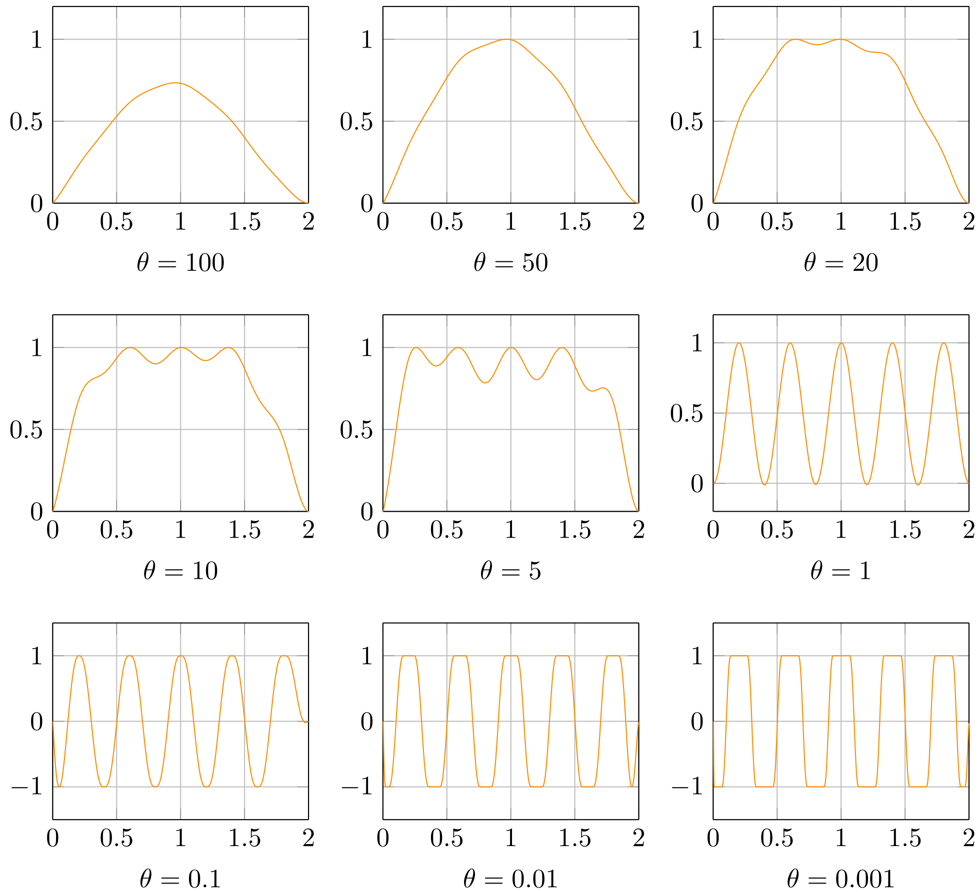

FIG. 7.5. *Example 1:* $\bar\lambda_\tau^\theta$ *for different values of* θ.

as well as for $1 \le j \le m$

$$\bar u_j = \begin{cases} 0 & \text{if } t < \frac{1}{m}, \\ c_{j1} & \text{if } \frac{1}{m} < t < \frac{3}{m}, \\ \vdots & \qquad \vdots \\ \sum_{k=1}^m c_{jk} & \text{if } \frac{2m-1}{m} < t < 2, \end{cases} \qquad \text{and} \qquad y_d = \bar y - L^*\bar\varphi,$$

where $L = \frac{\partial}{\partial t} - \Delta$ and $\bar y = y_{\bar u}$. Observing $\bar\Phi_j(t) = \frac{\alpha_j}{2}(1 - \cos(m\pi t))$ for all j we readily confirm the optimality of $\bar u = (\bar u_j)_{j=1}^m$ in a similar manner as in the first example.

The numerical results that follow are obtained by choosing $m = 3$, $\omega_1 = (-1, -\frac{1}{2})$, $\omega_2 = (-\frac{1}{4}, \frac{1}{4})$, $\omega_3 = (\frac{1}{2}, 1)$, $\kappa = 10^{-2}$, $c_{11} = 5$, $c_{22} = 3$, $c_{33} = 1$, and all other c_{jk} equal to zero. This implies that $\bar u_1$, $\bar u_2$ and $\bar u_3$ each have exactly one jump. These choices are specifically made to study the numerical behavior in situations where the inclusion $\mathrm{supp}(\bar{\mathrm{u}}_{\mathrm{j}}'^{+}) \subset \{\mathrm{t} \in [0, \mathrm{T}] : \bar\Phi_{\mathrm{j}}(\mathrm{t}) = \alpha_{\mathrm{j}}\}$ is strict. Similar to Example 1, we take $\bar y$ as solution of the state equation computed on a grid with $N_t = 12288$ and $N_h = 511$ nodes. In addition, we use $\bar y$ to obtain $y_{d,\sigma}$. We apply Algorithm BV with $N_t = 6144$ and $N_h = 255$, which corresponds to $\tau \approx 0.000326$ and $h = 0.0078125$.

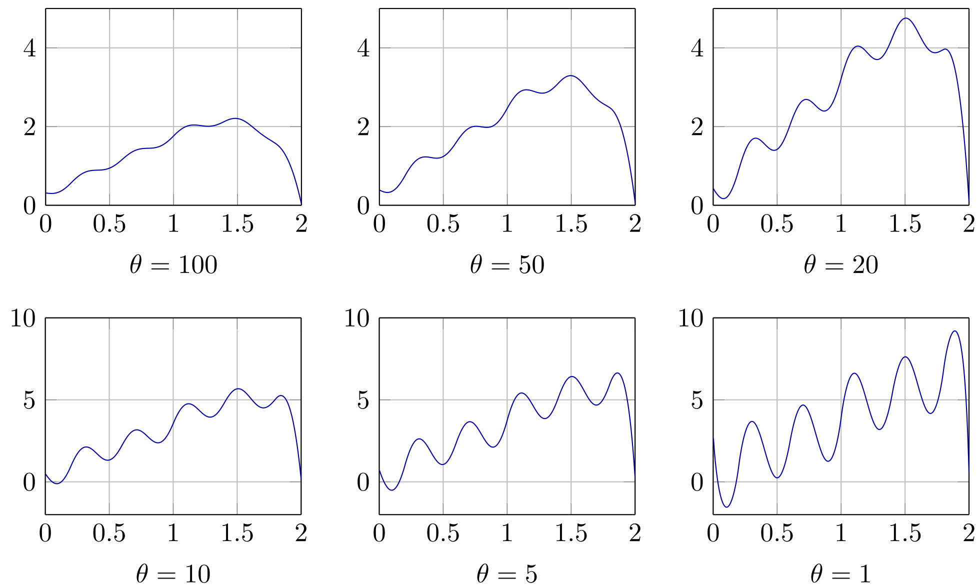

FIG. 7.6. *Example 1:* $\bar{u}^{\theta}_{\tau,L^2}$ *for different values of* θ.

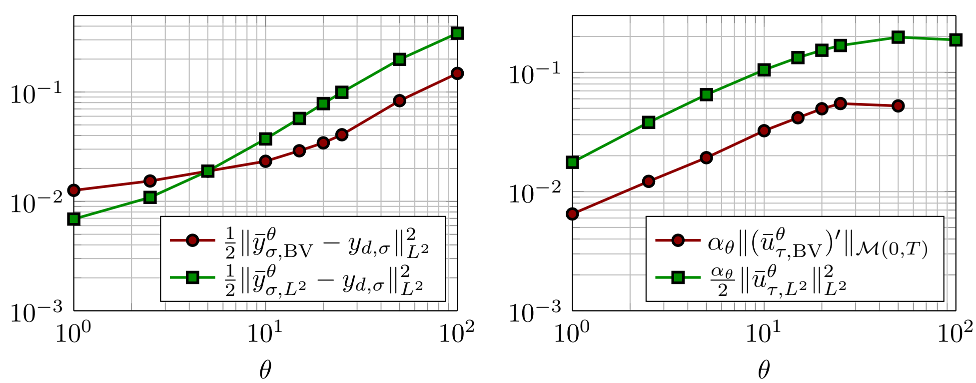

FIG. 7.7. *Example 1: Tracking errors and control norms for* (P) *and* L^2*-regularization.*

Figure 7.8 displays $y_{d,\sigma}$, $\bar{y}_\sigma$, $(\bar{u}_{\tau,j})_j$, and $(\bar{\lambda}_{\tau,j})_j$. The dual variables $(\bar{\lambda}_{\tau,j})_j$ resemble their continuous counterparts $(\bar{\lambda}_j)_j = \frac{1}{\alpha_j}\bar{\Phi}_j = \frac{1}{2}(1-\cos(3\pi t))$. In particular, each of them has three isolated maximums with value approximately 1. We recall from Corollary 4.8 that the approximated optimal controls $(\bar{u}_{\tau,j})_j$ are allowed to jump in at most these three points. We observe that $\bar{u}_{\tau,1}$ and $\bar{u}_{\tau,2}$ exhibit exactly one jump. This corresponds to the behavior of the analytical controls $\bar{u}_1$ and $\bar{u}_2$. The approximation $\bar{u}_{\tau,3}$ displays a minor jump at $t = 1$ which is not present in $\bar{u}_3$. At $t \approx 5/3$ it has a jump of height 1, just as $\bar{u}_3$. Summarizing we conclude from this example and other experiments that the case of strict inclusion $\mathrm{supp}(\bar{u}_j'^{+}) \subset \{t \in [0,T] : \bar{\Phi}_j(t) = \alpha_j\}$ is numerically challenging.

7.3. Example 3: One control and two spatial dimensions. In the third example we choose $m = 1$, $\Omega = (-1,1)^2$ and $\omega = (0,1)^2$. We consider the same

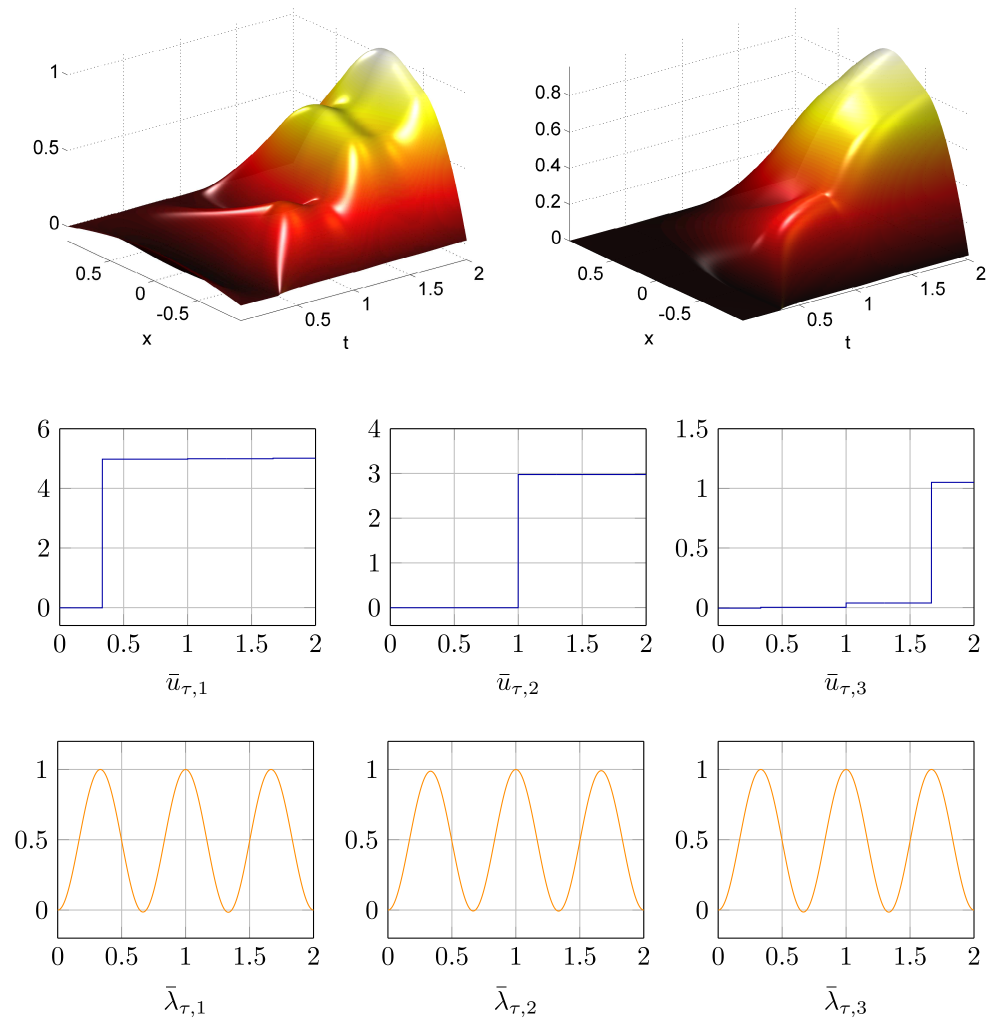

FIG. 7.8. *Example 2:* $y_{d,\sigma}$ *and* $\bar{y}_\sigma$ *(top left and right),* $(\bar{u}_{\tau,j})_j$*, and* $(\bar{\lambda}_{\tau,j})_j$.

objective function and state equation as in the first example, except that Ω and ω are different. We take $g = \chi_\omega$, $y_d(x_1, x_2, t) = (x_1 - 1.2)(x_1 + 1)(x_2 + 1)(x_2 - 0.9)te^{-t}$, and $\bar{\alpha} = 10^{-3}$. The choice of y_d yields $\bar{y} \neq y_d$ since y_d does not satisfy the boundary conditions of the state equation. We apply Algorithm BV with $N_t = 256$ and $N_h = 63^2$, which corresponds to $\tau = 0.0078125$ and $h = \sqrt{2}/32 \approx 0.0442$. Figure 7.9 shows $y_{d,\sigma}$ and $\bar{y}_\sigma$ at different points in time. Moreover, it depicts $\bar{u}_\tau = \bar{u}_{\tau,\mathrm{BV}}$ and $\bar{\lambda}_\tau$, as well as the optimal control $\bar{u}_{\tau,L^2}$ obtained through classical L^2-regularization (analogously as for Example 1). Apparently, in this example $\mathrm{supp}(\bar{u}')$ and $\{t \in [0,T] : \bar{\Phi}(t) = \bar{\alpha}\}$ do not consist of a finite number of points, but have positive measure.

While the tracking errors associated to the controls in Figure 7.9 are comparable, $\frac{1}{2}\|\bar{y}_{\sigma,\mathrm{BV}} - y_{d,\sigma}\|^2_{L^2(Q_h)} \approx 0.04026$ and $\frac{1}{2}\|\bar{y}_{\sigma,L^2} - y_{d,\sigma}\|^2_{L^2(Q_h)} \approx 0.04012$, the structure of the BV-control is simpler than that of the L^2-control. For the control terms in the objectives we have $\bar{\alpha}\|(\bar{u}_{\tau,\mathrm{BV}})'\|_{\mathcal{M}(0,T)} \approx 5 \cdot 10^{-4}$ and $\frac{\bar{\alpha}}{2}\|\bar{u}_{\tau,L^2}\|^2_{L^2(0,T)} \approx 8 \cdot 10^{-3}$.

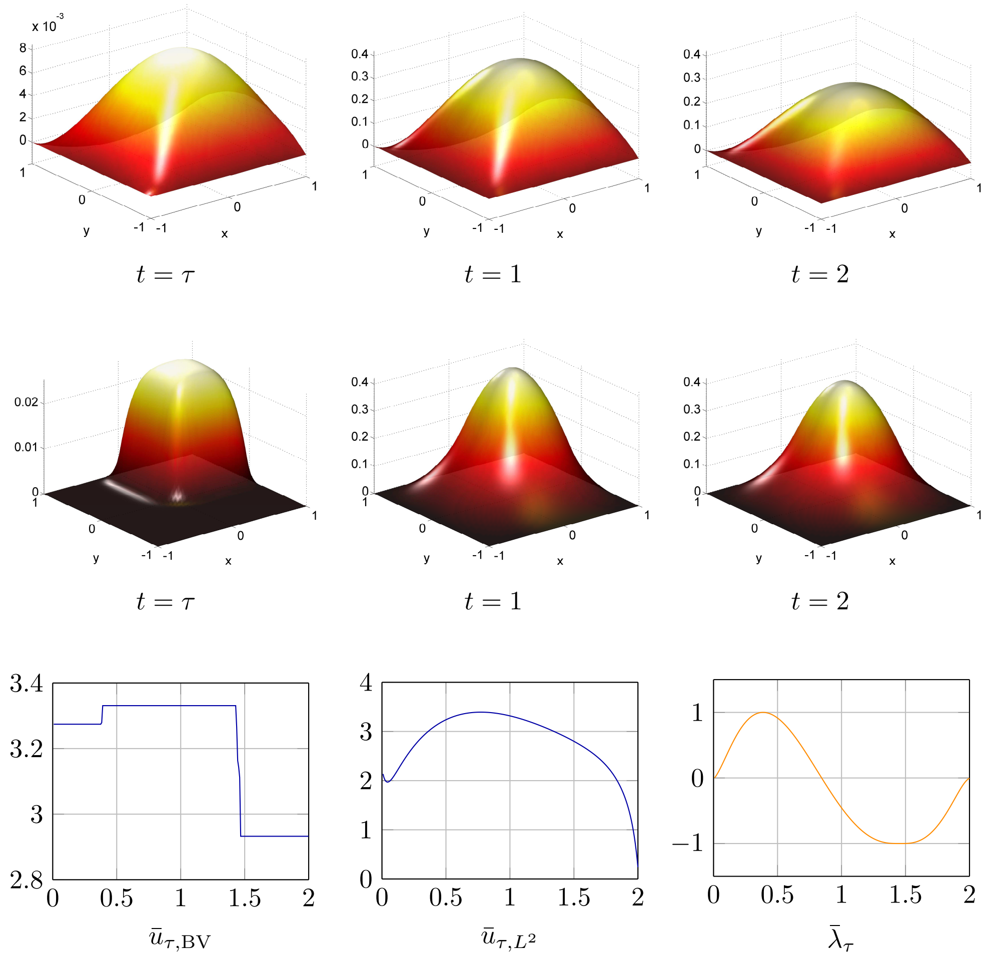

FIG. 7.9. *Example 3:* $y_{d,\sigma}$ *(top) and* $\bar{y}_\sigma$ *(middle) for different* t*,* $\bar{u}_{\tau,\mathrm{BV}}$*,* $\bar{u}_{\tau,L^2}$*, and* $\bar{\lambda}_\tau$*.*

REFERENCES

[1] L. Ambrosio, N. Fusco, and D. Pallara, *Functions of bounded variation and free discontinuity problems*, Oxford University Press, New York, 2000.

[2] E. Casas, *Pontryagin's principle for state-constrained boundary control problems of semilinear parabolic equations*, SIAM J. Control Optim., 35 (1997), pp. 1297–1327.

[3] E. Casas, C. Clason, and K. Kunisch, *Approximation of elliptic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 50 (2012), pp. 1735–1752.

[4] ———, *Parabolic control problems in measure spaces with sparse solutions*, SIAM J. Control Optim., 51 (2013), pp. 28–63.

[5] E. Casas, P.I. Kogut, and G. Leugering, *Approximation of optimal control problems in the coefficient for the p-Laplace equation. I. Convergence result*, SIAM J. Control Optim., *to appear*.

[6] E. Casas and K. Kunisch, *Optimal control of semilinear elliptic equations in measure spaces*, SIAM J. Control Optim., 52 (2014), pp. 339–364.

[7] E. Casas and F. Tröltzsch, *Second order analysis for optimal control problems: Improving results expected from abstract theory*, SIAM J. Optim., 22 (2012), pp. 261–279.

[8] ———, *Second order optimality conditions for weak and strong local solutions of parabolic optimal control problems*, Vietnam J. Math., (2015).

[9] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM: Control, Optimisation and Calculus of Variations, 17 (2011), pp. 243–266.
[10] I. Ekeland and R. Temam, *Convex analysis and variational problems. Unabridged, corrected republication of the 1976 English original*, SIAM, Philadelphia, 1999.
[11] E. Giusti, *Minimal surfaces and functions of bounded variation*, Birkhäuser, Boston, 1984.
[12] D. Hafemayr and B. Vexler, *Optimale Steuerung von Differentialgleichungen mit BV-Funktionen*, TU Munich, 2016.
[13] R. Herzog, G. Stadler, and G. Wachsmuth, *Directional sparsity in optimal control of partial differential equations*, SIAM J. Control Optim., 50 (2012), pp. 943–963.
[14] K. Kunisch, K. Pieper, and B. Vexler, *Measure valued directional sparsity for parabolic optimal control problems*, SIAM J. Control Optim., 52 (2014), pp. 3078–3108.
[15] D. Meidner and B. Vexler, *A priori error estimates for the space-time finite element discretization of parabolic optimal control problems. Part I: Problems without control constraints*, SIAM J. Control Optim., 47 (2008), pp. 1150–1177.
[16] I. Neitzel and B. Vexler, *A priori error estimates for space-time finite element discretization of semilinear parabolic optimal control problems*, Numer. Math., 120 (2012), pp. 345–386.
[17] K. Pieper and B. Vexler, *A priori error analysis for discretization of sparse elliptic optimal control problems in measure space*, SIAM J. Control Optim., 51 (2013), pp. 2788–2808.
[18] P.A. Raviart and J.M. Thomas, *Introduction à l'analyse numérique des equations aux dérivées partielles*, Masson, Paris, 1983.
[19] W. Rudin, *Real and complex analysis*, McGraw-Hill, London, 1970.
[20] R.E. Showalter, *Monotone operators in Banach space and nonlinear partial differential equations*, vol. 49 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, 1997.
[21] V. Thomée, *Galerkin finite element methods for parabolic problems*, Spinger-Verlag, Berlin, second ed., 2006.
[22] F. Tröltzsch, *Optimal control of partial differential equations: Theory, methods and applications*, vol. 112 of Graduate Studies in Mathematics, American Mathematical Society, Philadelphia, 2010.
[23] M. Ulbrich, *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces*, MOS-SIAM Series on Optimization, SIAM, Philadelphia, 2011.