Pré-Publicações do Departamento de Matemática
Universidade de Coimbra
Preprint Number 06–42

ELLIPTIC OPTIMAL CONTROL PROBLEMS WITH L^1-CONTROL COST AND APPLICATIONS FOR THE PLACEMENT OF CONTROL DEVICES

GEORG STADLER

ABSTRACT: Elliptic optimal control problems with L^1-control cost are analyzed. Due to the nonsmooth objective functional the optimal controls are identically zero on large parts of the control domain. For applications, in which one cannot put control devices (or actuators) all over the control domain, this provides information about where it is most efficient to put them. We analyze structural properties of L^1-control cost solutions. For solving the non-differentiable optimal control problem we propose a semismooth Newton method that can be stated and analyzed in function space and converges locally with a superlinear rate. Numerical tests on model problems show the usefulness of the approach for the location of control devices and the efficiency of our algorithm.

KEYWORDS: optimal control, nonsmooth regularization, optimal actuator location, placement of control devices, semismooth Newton, primal-dual method.

1. Introduction

In this paper, we analyze elliptic optimal control problems with L^1-control cost and argue their use for the placement of actuators (i.e. control devices). Due to the non-differentiability of the objective functional for L^1-control cost (in the sequel also called L^1-regularization), the structure of optimal controls differs significantly from what one obtains for the usual smooth regularization. If one cannot or does not want to distribute control devices all over the control domain, but wants to place available devices in an optimal way, the L^1-solution gives information about the optimal location of control devices. As model problems, we consider the following constrained elliptic optimal control problems with L^1 control cost.

$$\left\{\begin{array}{l} \text{minimize } J(y,u) := \frac{1}{2}\|y-y_d\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L^2}^2 + \beta\|u\|_{L^1} \\ \qquad\qquad \text{over } (y,u) \in H_0^1(\Omega)\times L^2(\Omega) \\ \text{subject to } Ay = u + f \in \Omega, \\ \qquad\qquad a \le u \le b \text{ almost everywhere in } \Omega, \end{array}\right. \tag{$\mathcal{P}$}$$

Received September 11, 2006.

where $\Omega \subset \mathbb{R}^n$ is a bounded domain with sufficiently smooth boundary $\Gamma = \partial\Omega$, $y_d, f \in L^2(\Omega)$, $a, b \in L^2(\Omega)$ with $a < 0 < b$ almost everywhere and $\alpha, \beta > 0$. Moreover, $A : H_0^1(\Omega) \mapsto H^{-1}(\Omega)$ is a second-order linear elliptic differential operator, and $\|\cdot\|_{L^2}$ and $\|\cdot\|_{L^1}$ denote the $L^2(\Omega)$ and $L^1(\Omega)$-norm, respectively. In the sequel, y is called state and u the control variable, and y_d is referred to as desired stated. Note that the novelty in the above problem is the introduction of the L^1-regularization term $\beta\|u\|_{L^1}$.

Nonsmooth regularization for PDE-constrained optimization has mainly been used for inverse problems, see e.g. [2,5,21,22,26]. In particular, the use of the L^1-norm of the gradient as regularization has led to better results for the recovery of data from noisy measurements than smooth regularization. As mentioned above, our main motivation for the use of nonsmooth regularization for optimal control problems is a different one, namely its ability to provide information about the optimal location of control devices and actuators. Although intuition and experience might help in this design issue, this approach fails when prior experience is lacking or the physical system modelled by the PDE is too complex. Provided only a finite number of control locations is possible, one might use a discrete method for the location problem, but clearly the number of possible configurations grows combinatorially as the number of devices or the number of possible locations increase. To overcome these problems, we propose the use of a L^1-norm control cost. As will be shown in this paper, this results in optimal controls that are identically zero in regions where they are not able to decrease the cost functional significantly (this significance is controlled by the size of $\beta > 0$); we may think of these sets as sets where no control devices need to be put. By these means, using the nonsmooth L^1-regularization term (even if in combination with the squared L^2-norm such as in $(\mathcal{P})$), one can treat the somewhat discrete-type question of *where* to place control devices and actuators.

An application problem that has partly motivated this research is the optimal placement of actuators on piezoelectric plates [8, 11]. Here, engineers want to know where to put electrodes in order to achieve a certain displacement of the plate. For linear material laws, this problem fits into the framework of our model problem $(\mathcal{P})$. Obviously, there are many other applications in which similar problems arise.

Additional motivation for considering $(\mathcal{P})$ is due to the fact that in certain applications the L^1-norm has a more interesting physical interpretation than the squared L^2-norm. For instance, the total fuel consumption of vehicles corresponds to a L^1-norm term, see [27]. We remark that the L^1-term $\|u\|_{L^1}$ is nothing else than the $L^1(\Omega)$-norm of u, while the L^2-term $\|u\|_{L^2}^2$ (which is the *squared* L^2-norm) is not a norm.

As mentioned above, the use of nonsmooth functionals in PDE-constrained optimization not standard and has mainly been used in the context of edge-preserving image processing (see [22, 26]) and other inverse problems where nonsmooth data have to be recovered (see e.g. [2,5]). An interesting comparison of the properties of various nonsmooth regularization terms in finite dimensions is given [9,20]. One of the few contributions using L^1-regularization in optimal control is [27]. Here, a free-flying robot whose dynamical behavior is governed by a system of nonlinear ordinary differential equations is navigated to a given final state. The optimal control is characterized as minimizer of an L^1-functional, which corresponds to the total fuel consumption. Finally, we mention the paper [14] that deals with elliptic optimal control problems with supremum-norm functional.

Clearly, the usage of a nonsmooth cost functional introduces severe difficulties into the problem, both theoretically as well as for a numerical algorithm. As mentioned above, a solution of $(\mathcal{P})$ with $\beta > 0$ obeys properties significantly different from the classical elliptic optimal control model problem

$$\begin{cases} \text{minimize } J_2(y,u) := \frac{1}{2}\|y-y_d\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L_2}^2 \\ \qquad\qquad\text{over } (y,u) \in H_0^1(\Omega)\times L^2(\Omega) \\ \text{subject to } Ay = u + f \in \Omega, \\ \qquad\qquad a \le u \le b \text{ almost everywhere in } \Omega, \end{cases} \tag{$\mathcal{P}_2$}$$

with $\alpha > 0$. One aim of this paper is to compare the structure of solutions of $(\mathcal{P})$ to those of $(\mathcal{P}_2)$ and to explore their different properties. Moreover, we propose and analyze an algorithm for the efficient solution of $(\mathcal{P})$.

Clearly, setting $\alpha := 0$ in $(\mathcal{P})$ results in the problem

$$\begin{cases} \text{minimize } J_1(y,u) := \frac{1}{2}\|y-y_d\|_{L^2}^2 + \beta\|u\|_{L^1} \\ \qquad\qquad\text{over } (y,u) \in W_0^{1,1}(\Omega)\times L^1(\Omega) \\ \text{subject to } Ay = u + f \in \Omega, \\ \qquad\qquad a \le u \le b \text{ almost everywhere in } \Omega. \end{cases} \tag{$\mathcal{P}_1$}$$

Now, the optimal control has to be searched for in the larger space $L^1(\Omega)$. The smoothing property of the elliptic operator A guarantees that the state y corresponding to $u \in L^1(\Omega)$ is an element in $L^2(\Omega)$, provided $n \le 4$. However, for the above problem to have a solution, the inequality constraints on the control are essential. In absence of (one of) the box constraints on u, $(\mathcal{P}_1)$ may or may not have a solution. This is due to the fact that $L^1(\Omega)$ is not a reflexive function space.

As a remedy for the difficulties that arise for $(\mathcal{P}_1)$, in the sequel we focus on $(\mathcal{P})$ with small $\alpha > 0$. Note that whenever $\beta > 0$, the cost functional in $(\mathcal{P})$ obeys a kink at points where $u = 0$ independently from $\alpha \geq 0$. In particular, $\alpha > 0$ does not regularize the non-differentiability of the functional $J(\cdot\,,\cdot)$. However, α influences the regularity of the solution and also plays an important role for the numerical method we propose for the solution of $(\mathcal{P})$. This algorithm is based on the combination of semismooth Newton methods, a condensation of Lagrange multipliers and certain complementarity functions. Moreover, it is related to the primal-dual active set method [3, 15]. Its fast local convergence can be proved in function space, which allows certain statements about the algorithm's dependence (or independence) of the fineness of the discretization [17]. We remark that the analysis of algorithms for PDE-constrained optimization in function space has recently gained a considerably amount of attention; we refer for instance to [15, 16, 23–25, 28, 29].

This paper is organized as follows. In the next section, we derive necessary optimality conditions for $(\mathcal{P})$ using Lagrange multipliers. In Section 3, we study structural properties of solutions of $(\mathcal{P})$. The algorithm we propose for solving elliptic optimal control problems with L^1-control cost is presented and analyzed in Section 4. Finally, in the concluding section, we report on numerical tests, where we discuss structural properties of the solutions as well as the performance of our algorithms.

2. First-order optimality system

In this section, we derive first-order necessary optimality conditions for $(\mathcal{P})$. For that purpose, we replace $(\mathcal{P})$ by a reduced problem formulation. This reduction to a problem that involves the control variable u only is possible due to the existence of the inverse $A^{-1} : H^{-1}(\Omega) \to H_0^1(\Omega)$ of the differential operator A. The reduced problem is

$$\left\{\begin{aligned} &\text{minimize } \hat{J}(u) := \tfrac{1}{2}\|A^{-1}u + A^{-1}f - y_d\|_{L^2}^2 + \tfrac{\alpha}{2}\|u\|_{L^2}^2 + \beta\|u\|_{L^1} \\ &\qquad\qquad\text{over } u \in U_{ad} := \{u \in L^2(\Omega) : a \leq u \leq b \text{ a.e. in } \Omega\}. \end{aligned}\right. \tag{$\hat{\mathcal{P}}$}$$

This is a convex optimization problem posed in the Hilbert space $L^2(\Omega)$. Its unique solvability follows from standard arguments [13, 24], and its solution

$\bar{u} \in U_{ad}$ is characterized (see e.g. [7, 10, 18]) by the variational inequality

$$\big(A^{-\star}(A^{-1}\bar{u} + A^{-1}f - y_d) + \alpha\bar{u}, u - \bar{u}\big) + \varphi(u) - \varphi(\bar{u}) \geq 0 \text{ for all } u \in U_{ad}, \quad (1)$$

where $A^{-\star}$ denotes the inverse of the transposed operator, i.e. $A^{-\star} = (A^\star)^{-1}$; moreover, $\varphi(v) := \beta \int |v(x)|\, dx = \beta\|v\|_{L^1}$. In the sequel, we denote by $\partial\varphi(\bar{u})$ the subdifferential of φ at $\bar{u}$, i.e. $\partial\varphi(\bar{u}) = \{u \in L^2(\Omega) : \varphi(v) - \varphi(\bar{u}) \geq (u, v - \bar{u}) \text{ for all } v \in L^2(\Omega)\}$. It follows from results in convex analysis that, for $\bar{\lambda} \in \partial\varphi(\bar{u})$, the equation (1) implies

$$\big(A^{-\star}(A^{-1}\bar{u} + A^{-1}f - y_d) + \alpha\bar{u} + \bar{\lambda}, u - \bar{u}\big) \geq 0 \text{ for all } u \in U_{ad}. \quad (2)$$

The differential inclusion $\bar{\lambda} \in \partial\varphi(\bar{u})$ yields, in particular, that

$$\bar{\lambda} \in \Lambda_{ad} := \{\lambda \in L^2(\Omega) : |\lambda| \leq \beta \text{ a.e. in } \Omega\}. \quad (3)$$

A pointwise (almost everywhere) discussion of the variational inequality (2) such as in [24, p. 57] allows to show that there exist nonnegative functions $\bar{\lambda}_a, \bar{\lambda}_b \in L^2(\Omega)$ that act as Lagrange multipliers for the inequality constraints in U_{ad}. Moreover, evaluating the differential inclusion $\bar{\lambda} \in \partial\varphi(\bar{u})$ relates $\bar{\lambda}$ to the sign of $\bar{u}$ (see also [10, 18, 23]). This leads to the optimality system for the reduced problem $(\hat{\mathcal{P}})$ summarized in the next theorem.

Theorem 1. *The optimal solution $\bar{u}$ of $(\hat{\mathcal{P}})$ is characterized by the existence of $(\bar{\lambda}, \bar{\lambda}_a, \bar{\lambda}_b) \in \Lambda_{ad} \times L^2(\Omega) \times L^2(\Omega)$ such that*

$$A^{-\star}(A^{-1}\bar{u} + A^{-1}f - y_d) + \alpha\bar{u} + \bar{\lambda} + \bar{\lambda}_b - \bar{\lambda}_a = 0, \quad (4a)$$

$$\bar{\lambda}_b \geq 0, b - \bar{u} \geq 0, \bar{\lambda}_b(b - \bar{u}) = 0, \quad (4b)$$

$$\bar{\lambda}_a \geq 0, \bar{u} - a \geq 0, \bar{\lambda}_a(\bar{u} - a) = 0, \quad (4c)$$

$$\bar{\lambda} = \beta \text{ a.e. on } \{x \in \Omega : \bar{u} > 0\}, \quad (4d)$$

$$|\bar{\lambda}| \leq \beta \text{ a.e. on } \{x \in \Omega : \bar{u} = 0\}, \quad (4e)$$

$$\bar{\lambda} = -\beta \text{ a.e. on } \{x \in \Omega : \bar{u} < 0\}. \quad (4f)$$

Above, (4b) - (4c) are the complementarity conditions for the inequality constraints in U_{ad}. Moreover, $\bar{\lambda} \in \Lambda_{ad}$ together with (4d) - (4f) is an equivalent expression for $\bar{\lambda} \in \partial\varphi(\bar{u})$.

Next, we derive an optimality system for $(\mathcal{P})$ using (4), i.e. the optimality conditions for $(\hat{\mathcal{P}})$. We introduce the adjoint variable $\bar{p}$ by

$$\bar{p} := -A^{-\star}(A^{-1}\bar{u} + A^{-1}f - y_d). \quad (5)$$

Then, (4a) becomes

$$-\bar{p} + \alpha\bar{u} + \bar{\lambda} + \bar{\lambda}_b - \bar{\lambda}_a = 0. \tag{6}$$

Applying the operator $A^\star$ to equation (5) and using the state variable $\bar{y} := A^{-1}(\bar{u} + f)$, we obtain the adjoint equation

$$A^\star \bar{p} = y_d - \bar{y}. \tag{7}$$

Next, we study the complementarity conditions (4b)–(4f). Surprisingly, it will turn out that we can write these conditions in a very compact form, namely as one (non-differentiable) operator equation. To do so, first we condense the Lagrange multipliers $\bar{\lambda}, \bar{\lambda}_a$ and $\bar{\lambda}_b$ into one multiplier

$$\bar{\mu} := \bar{\lambda} - \bar{\lambda}_a + \bar{\lambda}_b. \tag{8}$$

Now, we utilize a complementarity functions to reformulate the system (4b)–(4f) together with the condition $\bar{\lambda} \in \Lambda_{ad}$. Namely, we use

$$\begin{aligned} C(\bar{u}, \bar{\mu}) := \bar{u} - \max(0, \bar{u} + c(\bar{\mu} - \beta)) - \min(0, \bar{u} + c(\bar{\mu} + \beta)) \\ + \max(0, (\bar{u} - b) + c(\bar{\mu} - \beta)) + \min(0, (\bar{u} - a) + c(\bar{\mu} + \beta)) = 0 \end{aligned} \tag{9}$$

for some $c > 0$. Above, the min- and max-functions are to be understood pointwise. In the next lemma we clarify the relationship between (9) and (4b)–(4f).

Lemma 2. *For $(\bar{u}, \bar{\lambda}, \bar{\lambda}_a, \bar{\lambda}_b) \in (L^2(\Omega))^4$, the following two statements are equivalent:*

1. *The quadruple $(\bar{u}, \bar{\lambda}, \bar{\lambda}_a, \bar{\lambda}_b)$ satisfies the conditions* (4b)–(4f)*, and $\bar{\lambda} \in \Lambda_{ad}$.*
2. *There exists a function $\bar{\mu} \in L^2(\Omega)$ such that $(\bar{\mu}, \bar{u})$ satisfies* (9) *and the functions $\bar{\lambda}, \bar{\lambda}_a$ and $\bar{\lambda}_b$ can be derived as follows:*

$$\begin{cases} \bar{\lambda} & = \min(\beta, \max(-\beta, \bar{\mu})), \\ \bar{\lambda}_a & = -\min(0, \bar{\mu} + \beta), \\ \bar{\lambda}_b & = \max(0, \bar{\mu} - \beta). \end{cases} \tag{10}$$

Proof: First, we prove that $1 \Rightarrow 2$. We set $\bar{\mu} := \bar{\lambda} + \bar{\lambda}_b - \bar{\lambda}_a$. It requires a straightforward calculation to deduce from (4b)-(4f) and the assumption $a < 0 < b$ that (10) holds. It remains to show that $(\bar{u}, \bar{\mu})$ satisfies $C(\bar{u}, \bar{\mu}) = 0$. To prove that, we separately discuss subsets of Ω where $\mu(x) > \beta$, $\mu(x) = \beta$,

$|\mu(x)| < \beta$, $\mu(x) = -\beta$ and $\mu(x) < -\beta$. The argumentation below is to be understood in a pointwise almost everywhere sense.

- $\bar{\mu} > \beta$: From the construction of $\bar{\mu}$ we obtain, using (4b)-(4f) that $\bar{\mu} > \beta$ is only possible if $\bar{\lambda} = \beta$, $\bar{\lambda}_a = 0$ and $\bar{\lambda}_b > 0$. Thus, from (4b) we get $\bar{u} = b$. Therefore,
$$C(\bar{u}, \bar{\mu}) = \bar{u} - (\bar{u} + c(\bar{\mu} - \beta)) + ((\bar{u} - b) + c(\bar{\mu} - \beta)) = 0.$$
- $\bar{\mu} = \beta$: Again, from (4b)-(4f) it follows that $\bar{\lambda} = \beta$ and $\bar{\lambda}_a = \bar{\lambda}_b = 0$. The conditions (4b) and (4d) imply that $0 \le \bar{u} \le b$ and thus
$$C(\bar{u}, \bar{\mu}) = \bar{u} - (\bar{u} + c(\bar{\mu} - \beta)) = 0.$$
- $|\bar{\mu}| < \beta$: In this case we deduce from (4b)-(4f) that $\bar{\lambda} = \bar{\mu}$ and $\bar{\lambda}_a = \bar{\lambda}_b = 0$. From (4e) we obtain $\bar{u} = 0$ and $C(\bar{u}, \bar{\mu}) = 0$ is trivially satisfied.
- Since the verification of the condition $C(\bar{u}, \bar{\mu}) = 0$ for the two remaining sets where $\mu = -\beta$ or $\mu < -\beta$ is very similarly to the cases $\mu = \beta$ and $\mu > \beta$, it is skipped here.

This ends the first part of the proof.

Now, we turn to the implication $2 \Rightarrow 1$. We suppose given $(\bar{u}, \bar{\mu}) \in (L^2(\Omega))^2$ that satisfy $C(\bar{u}, \bar{\mu}) = 0$ and derive $\bar{\lambda}$, $\bar{\lambda}_a$ and $\bar{\lambda}_b$ from (10). By definition, it follows that $\bar{\lambda} \in \Lambda_{ad}$ and that $\bar{\mu} = \bar{\lambda} - \bar{\lambda}_a + \bar{\lambda}_b$ holds. To prove the conditions (4b)-(4f), we again distinguish different cases:

- $\bar{u} + c(\bar{\mu} - \beta) > b$: In this case, only the two max-terms in (9) contribute to the sum. We obtain
$$0 = C(\bar{u}, \bar{\mu}) = \bar{u} - b$$
and thus $\bar{u} = b$. Now, from $\bar{u} + c(\bar{\mu} - \beta) > b$ we obtain $\bar{\mu} > \beta$, which implies that $\bar{\lambda} = \beta$, $\bar{\lambda}_b > 0$ and $\bar{\lambda}_a = 0$ and the conditions (4b)-(4f) are satisfied.
- $0 < \bar{u} + c(\bar{\mu} - \beta) \le b$: Here, only the first max-term in (9) is different from zero. Hence,
$$0 = C(\bar{u}, \bar{\mu}) = \bar{\mu} - \beta.$$
This implies $\bar{\mu} = \beta$ and $0 < \bar{u} \le b$. Clearly, $\bar{\lambda} = \beta$ and $\bar{\lambda}_a = \bar{\lambda}_b = 0$, and again (4b)-(4f) hold.

- $|\bar{u}+c\bar{\mu}| \le c\beta$: In this case, all the max and min-terms in (9) are equal to zero, which implies $\bar{u}=0$. This shows that $|\bar{\mu}| \le \beta$, and thus $\bar{\lambda}=\bar{\mu}$ and $\bar{\lambda}_a=\bar{\lambda}_b=0$ which proves the conditions (4b)-(4f).
- $a \le \bar{u}+c(\bar{\mu}+\beta) < 0$: This case is very similar to the case $0 < \bar{u}+c(\bar{\mu}-\beta) \le b$ discussed above.
- $\bar{u}+c(\bar{\mu}+\beta) < a$: Analogous to the case $\bar{u}+c(\bar{\mu}-\beta) > b$.

Since for every point of Ω exactly one of the above five conditions holds, this finishes the proof of the implication $2 \Rightarrow 1$ and ends the proof. ■

In the next theorem we summarize the first-order optimality conditions for $(\mathcal{P})$.

Theorem 3. *The solution $(\bar{y},\bar{u}) \in H_0^1(\Omega) \times L^2(\Omega)$ of $(\mathcal{P})$ is characterized by the existence of $(\bar{p},\bar{\mu}) \in H_0^1(\Omega) \times L^2(\Omega)$ such that*

$$A\bar{y} - \bar{u} - f = 0, \tag{11a}$$

$$A^\star p + \bar{y} - y_d = 0, \tag{11b}$$

$$-\bar{p} + \alpha\bar{u} + \bar{\mu} = 0, \tag{11c}$$

$$\begin{aligned}\bar{u} &- \max(0, \bar{u}+c(\bar{\mu}-\beta)) - \min(0, \bar{u}+c(\bar{\mu}+\beta)) \\ &+ \max(0, (\bar{u}-b)+c(\bar{\mu}-\beta)) + \min(0, (\bar{u}-a)+c(\bar{\mu}+\beta)) = 0,\end{aligned} \tag{11d}$$

with $c > 0$.

Note that, from (11) one obtains an optimality system for $(\mathcal{P}_2)$ simply by setting $\beta = 0$ in (11): While the equations (11a)–(11c) remain unchanged, the condition (11d) becomes

$$\bar{\mu} + \max(0, \bar{\mu}+c^{-1}(\bar{u}-b)) + \min(0, \bar{\mu}+c^{-1}(\bar{u}-a)) = 0. \tag{12}$$

This formulation has been used for the construction of an algorithm for bilaterally control constraint optimal control problems of the form $(\mathcal{P}_2)$, see [19]. The next section is concerned with structural properties of solutions of $(\mathcal{P})$ in comparison with those of $(\mathcal{P}_2)$.

3. Properties of solutions of $(\mathcal{P})$

For simplicity of the presentation, in this section we dismiss the control constraints in $(\mathcal{P})$ and $(\mathcal{P}_2)$, i.e. we choose $a := -\infty$ and $b := \infty$ and thus $U_{ad} = L^2(\Omega)$. We think of $\alpha > 0$ being fixed and study the dependence of the optimal control on β. To emphasize this dependence, in the rest of this section we denote the solution of $(\mathcal{P})$ by $(\bar{y}_\beta, \bar{u}_\beta)$. The first lemma states that, if β is sufficiently large, the optimal control is $\bar{u}_\beta \equiv 0$.

Lemma 4. *If $\beta \geq \beta_0 := \|A^{-\star}(y_d - A^{-1}f)\|_{L^\infty}$, the unique solution of $(\mathcal{P})$ is $(\bar{y}_\beta, \bar{u}_\beta) = (A^{-1}f, 0)$.*

Proof: For the proof we use the reduced form $(\hat{\mathcal{P}})$ of $(\mathcal{P})$. For arbitrary $u \in L^2(\Omega)$ we consider

$$\begin{aligned}
\hat{J}(u) - \hat{J}(0) &= \frac{1}{2}\|A^{-1}u\|_{L^2}^2 - (y_d - A^{-1}f, A^{-1}u)_{L^2} + \beta\|u\|_{L^1} + \frac{\alpha}{2}\|u\|_{L^2}^2 \\
&\geq \frac{1}{2}\|A^{-1}u\|_{L^2}^2 - \|u\|_{L^1}\|A^{-\star}(y_d - A^{-1}f)\|_{L^\infty} + \beta\|u\|_{L^1} + \frac{\alpha}{2}\|u\|_{L^2}^2 \\
&= \frac{1}{2}\|A^{-1}u\|_{L^2}^2 + (\beta - \beta_0)\,\|u\|_{L^1} + \frac{\alpha}{2}\|u\|_{L^2}^2.
\end{aligned}$$

Clearly, the latter expression is nonnegative if $\beta \geq \beta_0$. Thus, for $\beta \geq \beta_0$, $\hat{J}(u) - \hat{J}(0) \geq 0$ for all $u \in U_{ad}$, which proves that the optimal control is $\bar{u}_\beta \equiv 0$. Using (11a) the corresponding state is obtained as $A^{-1}f$. ■

An analogous result with respect to the parameter α in $(\mathcal{P}_2)$ does not hold, i.e. in general optimal controls for $(\mathcal{P}_2)$ will only approach zero as α tends to infinity. Lemma (4) is also a consequence of the fact that the L^1-term in the objective functional can be seen as exact penalization (see e.g. [4]) for the constraint $u = 0$.

To gain more insight in the structure of solutions of $(\mathcal{P})$ and in the role of the cost weight parameters α and β, we next discuss the behavior of $\bar{u}_\beta$ as β changes (while $\alpha > 0$ is kept fixed), i.e. we investigate the mapping

$$\Phi : [0, \infty) \to L^2(\Omega), \;\; \Phi(\beta) := \bar{u}_\beta.$$

For that purpose, we derive the sensitivity of the optimal control $\bar{u}_\beta$ with respect to β. We will show that the function Φ is directionally differentiable, and that its derivative is discontinuous at boundaries of regions where $\bar{u}_\beta = 0$. Moreover, we discuss the influence of α on the discontinuity. We start our study with continuity properties of Φ.

Lemma 5. *The mapping Φ is Lipschitz continuous.*

Proof: Let $\beta, \beta' \geq 0$ and denote the solution variables corresponding to β and β' by $(\bar{y}_\beta, \bar{u}_\beta, \bar{p}_\beta, \bar{\mu}_\beta)$ and $(\bar{y}_{\beta'}, \bar{u}_{\beta'}, \bar{p}_{\beta'}, \bar{\mu}_{\beta'})$, respectively. From (11a)–(11c) we obtain

$$A^{-\star}A^{-1}u + \alpha u - A^{-\star}y_d + A^{-\star}A^{-1}f + \mu = 0 \tag{13}$$

for both $(u,\mu) = (\bar u_\beta, \bar\mu_\beta)$ and $(u,\mu) = (\bar u_{\beta'}, \bar\mu_{\beta'})$. Deriving the difference between these two equations leads to

$$A^{-\star}A^{-1}(\bar u_\beta - \bar u_{\beta'}) + \alpha(\bar u_\beta - \bar u_{\beta'}) + \bar\mu_\beta - \bar\mu_{\beta'} = 0,$$

and taking the inner product with $\bar u_\beta - \bar u_{\beta'}$ results in

$$\|A^{-1}(\bar u_\beta - \bar u_{\beta'})\|_{L^2}^2 + \alpha\|(\bar u_\beta - \bar u_{\beta'})\|_{L^2}^2 = (\bar\mu_\beta - \bar\mu_{\beta'}, \bar u_{\beta'} - \bar u_\beta)_{L^2}. \tag{14}$$

We now estimate the right hand side of (14) pointwise (almost everywhere). From the complementarity conditions, we deduce that the following cases can occur:

- $\bar\mu_\beta = \beta, u_\beta \geq 0, \bar\mu_{\beta'} = \beta', u_{\beta'} \geq 0$: Here, we obtain
$$(\bar\mu_\beta - \bar\mu_{\beta'})(\bar u_{\beta'} - \bar u_\beta) = (\beta - \beta')(\bar u_{\beta'} - \bar u_\beta) \leq |\beta - \beta'||\bar u_{\beta'} - \bar u_\beta|.$$
- $\bar\mu_\beta = \beta, u_\beta \geq 0, |\bar\mu_{\beta'}| < \beta', u_{\beta'} = 0$: In this case, we find $\bar u_{\beta'} - \bar u_\beta = -\bar u_\beta \leq 0$ and thus
$$(\bar\mu_\beta - \bar\mu_{\beta'})(\bar u_{\beta'} - \bar u_\beta) \leq (\beta - \beta')(\bar u_{\beta'} - \bar u_\beta) \leq |\beta - \beta'||\bar u_{\beta'} - \bar u_\beta|.$$
- $\bar\mu_\beta = \beta, u_\beta \geq 0, \bar\mu_{\beta'} = -\beta', u_{\beta'} \leq 0$: From the sign structure of the variables one obtains the estimate
$$(\bar\mu_\beta - \bar\mu_{\beta'})(\bar u_{\beta'} - \bar u_\beta) \leq 0.$$
- $|\bar\mu_\beta| < \beta, u_\beta = 0, |\bar\mu_{\beta'}| < \beta', u_{\beta'} = 0$: In this case, trivially
$$(\bar\mu_\beta - \bar\mu_{\beta'})(\bar u_{\beta'} - \bar u_\beta) = 0.$$
- There are five more cases that may occur. Since they are very similar to those discussed above, their discussion is skipped here and we only remark that in all remaining cases the pointwise estimate
$$(\bar\mu_\beta - \bar\mu_{\beta'})(\bar u_{\beta'} - \bar u_\beta) \leq |\beta - \beta'||\bar u_{\beta'} - \bar u_\beta|$$

holds as well. Summarizing, we obtain

$$(\bar\mu_\beta - \bar\mu_{\beta'})(\bar u_{\beta'} - \bar u_\beta) \leq |\beta - \beta'||\bar u_{\beta'} - \bar u_\beta| \text{ almost everywhere in } \Omega.$$

Integrating result in

$$(\bar\mu_\beta - \bar\mu_{\beta'}, \bar u_{\beta'} - \bar u_\beta)_{L^2} \leq |\Omega|^{1/2}|\beta - \beta'|\|\bar u_{\beta'} - \bar u_\beta\|_{L^2}, \tag{15}$$

and combing (14) with (15) leads to

$$\|\bar u_{\beta'} - \bar u_\beta\|_{L^2} \leq \frac{1}{\alpha}|\Omega|^{1/2}|\beta - \beta'|,$$

which proves Lipschitz continuity of Φ and thus ends the proof. ∎

Clearly, from the above lemma we get $L^2(\Omega)$-boundedness of the sequence

$$\frac{1}{\beta-\beta'}(\bar u_\beta - \bar u_{\beta'}) \ \text{ as } \beta' \to \beta,\ \beta' > \beta. \tag{16}$$

Let $\dot u_\beta$ denote a weak accumulation point of this sequence, i.e. a weak limit of a subsequence. On the same subsequence $(\bar y_\beta - \bar y_{\beta'})/(\beta-\beta') = (A^{-1}\bar u_\beta - A^{-1}\bar u_{\beta'})/(\beta-\beta') \to \dot y_\beta$ weakly in $H_0^1(\Omega)$ and strongly in $L^2(\Omega)$. Moreover, $(\bar p_\beta - \bar p_{\beta'})/(\beta-\beta') = (-A^{-\star}A^{-1}\bar u_\beta - -A^{-\star}A^{-1}\bar u_{\beta'})/(\beta-\beta') \to \dot p_\beta$ strongly in $L^2(\Omega)$. We will characterize $\dot u_\beta$ as right directional derivative of Φ and give its explicit form. Similar arguments can be used to derive the left directional derivative; for related results we also refer to [16].

For the following discussion we consider the optimality system (11) taking into account that $U_{ad} = L^2(\Omega)$ and choosing $c := \alpha^{-1}$. This results in

$$A\bar y_\beta - \bar u_\beta - f = 0, \tag{17a}$$
$$A^\star \bar p_\beta + \bar y_\beta - y_d = 0, \tag{17b}$$
$$-\bar\mu_\beta + \bar p_\beta - \alpha \bar u_\beta = 0, \tag{17c}$$
$$\bar u_\beta - \max(0, \bar u_\beta + \alpha^{-1}(\bar\mu_\beta - \beta)) - \min(0, \bar u_\beta + \alpha^{-1}(\bar\mu_\beta + \beta)) = 0. \tag{17d}$$

Using (17c), we can replace $\bar\mu_\beta$ in (17d) and, due to the choice of $c = \alpha^{-1}$ obtain

$$\bar u_\beta - \max(0, \alpha^{-1}(\bar p_\beta - \beta)) - \min(0, \alpha^{-1}(\bar p_\beta + \beta)) = 0. \tag{18}$$

Besides the fact that we have eliminated the variable $\bar\mu_\beta$ from the optimality system (17), the formulation (18) also has the advantage that the expressions in the pointwise max- and min-operators enjoy additional regularity. The choice $c = \alpha^{-1}$ will be essential for the reasoning in the next section, where we present an algorithm for the solution of $(\mathcal{P})$. Next, we introduce the functions

$$g^-(\beta) := \frac{1}{\alpha}(\bar p_\beta + \beta) = \frac{1}{\alpha}(A^{-1}y_d - A^{-\star}A^{-1}(\bar u_\beta + f) + \beta), \tag{19a}$$
$$g^+(\beta) := \frac{1}{\alpha}(\bar p_\beta - \beta) = \frac{1}{\alpha}(A^{-1}y_d - A^{-\star}A^{-1}(\bar u_\beta + f) - \beta). \tag{19b}$$

On the same subsequence as above we have

$$\frac{1}{\beta-\beta'}(g^-(\beta) - g^-(\beta')) \longrightarrow -\frac{1}{\alpha}A^{-\star}A^{-1}\dot u_\beta + \frac{1}{\alpha} =: \dot g^-(\beta),$$

where, due to the smoothing properties of the operator A^{-1} and its transposed $A^{-\star}$ the latter convergence is strong in $L^2(\Omega)$. Analogously, one obtains

$$\frac{1}{\beta-\beta'}(g^+(\beta)-g^+(\beta')) \longrightarrow -\frac{1}{\alpha}A^{-\star}A^{-1}\dot{u}_\beta - \frac{1}{\alpha} =: \dot{g}^+(\beta)$$

strongly in $L^2(\Omega)$. We now introduce the sets

$$\begin{aligned}\mathcal{S}_\beta^+ &= \{x\in\Omega : g^+(\beta)>0 \text{ or } (g^+(\beta)=0 \wedge \dot{g}^+(\beta)\geq 0)\},\\ \mathcal{S}_\beta^- &= \{x\in\Omega : g^-(\beta)<0 \text{ or } (g^-(\beta)=0 \wedge \dot{g}^-(\beta)\leq 0)\}\end{aligned}$$

and denote by $\chi_{\mathcal{S}}$ the characteristic function for a set $\mathcal{S}\subset\Omega$. We are now prepared to prove the following theorem.

Theorem 6. *Let $\dot{u}_\beta$ be a weak accumulation point of the sequence* (16). *Then* $(A^{-1}\dot{u}_\beta, \dot{u}_\beta)$ *solves the auxiliary optimal control problem*

$$\left\{\begin{array}{l} \textit{minimize } J_{aux}(\dot{y}_\beta,\dot{u}_\beta) := \frac{1}{2}\|\dot{y}_\beta\|_{L^2}^2 + \frac{\alpha}{2}\|\dot{u}_\beta\|_{L^2}^2 + (\chi_{\mathcal{S}_\beta^+}-\chi_{\mathcal{S}_\beta^-}, \dot{u}_\beta)\\ \qquad\qquad \textit{over } (\dot{y}_\beta,\dot{u}_\beta)\in H_0^1(\Omega)\times L^2(\Omega)\\ \textit{subject to } A\dot{y}_\beta = \dot{u}_\beta \in \Omega,\ \ \dot{u}_\beta = 0 \textit{ in } \Omega\setminus(\mathcal{S}_\beta^-\cup\mathcal{S}_\beta^+).\end{array}\right. \tag{21}$$

Proof: From (13) for $(\bar{u}_\beta,\bar{\mu}_\beta)$ and $(\bar{u}_{\beta'},\bar{\mu}_{\beta'})$, we obtain

$$A^{-\star}A^{-1}\dot{u}_\beta + \alpha\dot{u}_\beta + \dot{\mu}_\beta = 0, \tag{22}$$

where $\dot{\mu}_\beta$ is the weak limit of $\frac{1}{\beta-\beta'}(\bar{\mu}_\beta-\bar{\mu}_{\beta'})$ on the subsequence chosen above. Using (19b) and separately arguing for sets with $g^+(\beta)>0$, $g^+(\beta)=0$ and $g^+(\beta)<0$, one obtains that

$$\frac{1}{\beta-\beta'}\left(\max(0,g^+(\beta))-\max(0,g^+(\beta'))\right) \to \frac{1}{\alpha}\left(\dot{p}_\beta-1\right)\chi_{\mathcal{S}_\beta^+}$$

strongly in $L^2(\Omega)$. Similarly, with (19a)

$$\frac{1}{\beta-\beta'}\left(\min(0,g^-(\beta))-\max(0,g^-(\beta'))\right) \to \frac{1}{\alpha}\left(\dot{p}_\beta+1\right)\chi_{\mathcal{S}_\beta^-}.$$

Hence, from (18) we obtain

$$\dot{u}_\beta - \frac{1}{\alpha}\left(\dot{p}_\beta-1\right)\chi_{\mathcal{S}_\beta^+} - \frac{1}{\alpha}\left(\dot{p}_\beta+1\right)\chi_{\mathcal{S}_\beta^-} = 0, \tag{23}$$

which yields that $\dot{u}_\beta = 0$ on $\Omega\setminus(\mathcal{S}_\beta^-\cup\mathcal{S}_\beta^+)$. Moreover, from (23)

$$-\dot{p}_\beta + \alpha\dot{u}_\beta + \chi_{\mathcal{S}_\beta^+} - \chi_{\mathcal{S}_\beta^-} = 0 \ \text{ on } \ \mathcal{S}_\beta^-\cup\mathcal{S}_\beta^+. \tag{24}$$

Comparing (24) with (22) shows that $\dot\mu_\beta = 1$ on $\mathcal{S}_\beta^+$ and $\dot\mu_\beta = -1$ on $\mathcal{S}_\beta^-$. Now, it is easy to see that $(\dot y_\beta, \dot u_\beta)$ solves (21) and that $\dot\mu_\beta - (\chi_{\mathcal{S}_\beta^+} - \chi_{\mathcal{S}_\beta^-})$ acts as Lagrange multiplier for the equality constraint $\dot u_\beta = 0$ on $\Omega \setminus (\mathcal{S}_\beta^- \cup \mathcal{S}_\beta^+)$. ■

Let us conclude this section with a discussion of structural properties of $\dot u_\beta$. We consider the boundary of $\mathcal{S}_\beta^+$. On the one hand, $\dot u_\beta = 0$ on $\Omega \setminus (\mathcal{S}_\beta^- \cup \mathcal{S}_\beta^+)$. On the other hand, from (24) we obtain that

$$-\dot p_\beta + \alpha \dot u_\beta - 1 = 0 \;\text{ in }\; \mathcal{S}_\beta^+. \tag{25}$$

In our numerical tests we observe that usually $\dot p_\beta = A^{-\star}A^{-1}\dot u_\beta$ is small compared to $\alpha \dot u_\beta$, and thus (25) shows that $\dot u_\beta$ is of the order of magnitude of $1/\alpha$ in $\mathcal{S}_\beta^+$. Hence, $\dot u_\beta$ obeys a jump of approximate magnitude α^{-1} along the boundary of $\mathcal{S}_\beta^+$. This jump in the (directional) derivative $\dot u_\beta$ partly explains the following behavior we observe in our test problems as we solve them for decreasing β: Starting with large β, the optimal control equals zero until β drops below a certain value, when the optimal control becomes different from zero. This behavior can partly be explained with the L^1-term in the cost functional.

Moreover, for $n \le 3$, parts of Ω where $\bar u$ has a different sign are always separated by regions with positive measure, in which $\bar u_\beta$ is identically zero. This follows easily from standard regularity results for elliptic equations [12], which show that $\bar u_\beta \in H^2(\Omega)$: The Sobolev embedding theorem [1, p. 97] yields that $H^2(\Omega)$ embeds into the space of continuous functions, which implies that $\bar p_\beta$ is continuous. Now, the assertion follows from (18).

4. Primal-dual active set method

Here, we present a numerical technique to find the solution of $(\mathcal{P})$ or, equivalently, of the reformulated first-order optimality conditions (11). Obviously, an algorithm based on (11) has to cope with the pointwise min- and max-terms in (11d). One possibility to deal with these non-differentiabilities is utilizing smooth approximations of the max- and min-operators. This leads to a smoothed L^1-norm term in $(\mathcal{P})$ and thus has the disadvantage that the typical properties of solutions of $(\mathcal{P})$ (e.g. , the splitting into sets with $u = 0$ and $u \neq 0$) are lost. Hence, we prefer solving (11) directly instead of dealing with a smoothed version. Since we focus on fast second-order methods, we require an appropriate linearization of the nonlinear and nonsmooth system (11). We use a recent generalized differentiability concept, which, for the convenience of the reader is briefly summarized in the next section. We point out that this concept of semismoothness and generalized Newton methods holds in a function space setting. Such an infinite-dimensional analysis has several advantages over purely finite-dimensional approaches: The regularity (or, more important, the non-regularity) of variables often explains the behavior of algorithms dealing with the discretized problem; Moreover, the well-posedness of a method in infinite dimensions is the basis for the investigation of mesh-independence properties.

4.1. Semismoothness in function space. Since we want to apply the differentiability concept introduced in [6, 15, 25] to (11), we briefly recall its notion and some results here.

Definition 7. *Let X, Y be Banach spaces, $D \subset X$ be open and $\mathcal{F} : D \longrightarrow Y$ be a nonlinear mapping. Then, we call the mapping $\mathcal{F}$ generalized differentiable on the open subset $U \subset D$ if there exists a mapping $\mathcal{G} : U \longrightarrow \mathcal{L}(X, Y)$ such that*

$$\lim_{h\to 0} \frac{1}{\|h\|_X} \|\mathcal{F}(x+h) - \mathcal{F}(x) - \mathcal{G}(x+h)h\|_Y = 0 \tag{26}$$

for every $x \in U$.

The above introduced mapping $\mathcal{G}$, which need not be unique, is referred to as *generalized derivative*. Note that in (26) $\mathcal{G}$ is evaluated at the point $x + h$ rather than at x and thus might change as $h \to 0$. Focusing on (11d), we are interested in the differentiability properties of the pointwise max- and min-operators. To state the next result, which is due to [15], we restrict ourselves to the pointwise max-operator. The corresponding result for the min-operator follows easily from $\min(0, \cdot) = -\max(0, -\cdot)$.

Lemma 8. *The pointwise* max*-operator $\mathcal{F}_{max} : L^r(\Omega) \to L^s(\Omega)$ defined by $\mathcal{F}_{max}(v) = \max(0, v)$ for $v \in L^r(\Omega)$ is generalized differentiable for $1 \le s < r \le \infty$. The mapping*

$$\mathcal{G}_{max}(y)(x) = \begin{cases} 1 & \text{if } y(x) \ge 0, \\ 0 & \text{if } y(x) < 0, \end{cases} \tag{27}$$

is a generalized derivative of $\mathcal{F}_{max}$ at y.

Above, $\mathcal{G}_{max}$ is chosen as an element of the subgradient of convex analysis (see [7]). However, generalized derivatives of $\mathcal{F}_{max}$ need not be elements of the subgradient, see [15]. A generalized derivative of the pointwise min-operator $\mathcal{F}_{min} : L^r(\Omega) \to L^s(\Omega)$ is

$$\mathcal{G}_{min}(y)(x) = \begin{cases} 1 & \text{if } y(x) \le 0, \\ 0 & \text{if } y(x) > 0. \end{cases}$$

We point out that the norm gap (i.e. , $r < s$) is essential for generalized differentiability of $\mathcal{F}_{min}$ and $\mathcal{F}_{max}$. Assume now we intend to find a root $\bar{x}$ of

$$\mathcal{F}(x) = 0 \tag{28}$$

employing a Newton iteration. That is, given an iterate x^k, the next iterate x^{k+1} is computed from

$$\mathcal{F}(x^k) + \mathcal{G}(x^k)(x^{k+1} - x^k) = 0.$$

Then, following [6, 15, 25] the following local convergence result holds:

Theorem 9. *Suppose that $\bar{x} \in D$ is a solution of* (28) *and that $\mathcal{F}$ is semismooth in an open neighborhood U of $\bar{x}$ with generalized derivative $\mathcal{G}$. If $\mathcal{G}(x)^{-1}$ exists for all $x \in U$ and $\{\|\mathcal{G}(x)^{-1}\| : x \in U\}$ is bounded, the Newton iteration*

$$x^0 \in U \text{ given, } x^{k+1} = x^k - \mathcal{G}(x^k)^{-1}\mathcal{F}(x^k)$$

is well-defined and, provided x^0 is sufficiently close to $\bar{x}$, converges at superlinear rate.

Now, we apply the above calculus to derive and analyze a solution algorithm for $(\mathcal{P})$.

4.2. Application to $(\mathcal{P})$**.** ¿From (11c), we infer that

$$\bar{\mu} = \bar{p} - \alpha\bar{u}.$$

Inserting this identity in (11d) results in

$$\begin{aligned} &\bar{u} - \max(0, \bar{u} + c(\bar{p} - \alpha\bar{u} - \beta)) - \min(0, \bar{u} + c(\bar{p} - \alpha\bar{u} + \beta)) \\ &\quad + \max(0, (\bar{u} - b) + c(\bar{p} - \alpha\bar{u} - \beta)) + \min(0, (\bar{u} - a) + c(\bar{p} - \alpha\bar{u} + \beta)) = 0. \end{aligned} \tag{29}$$

With the choice $c := \alpha^{-1}$, (29) becomes

$$\begin{aligned} &\bar{u} - \alpha^{-1}\max(0, \bar{p} - \beta) - \alpha^{-1}\min(0, \bar{p} + \beta) \\ &\quad + \alpha^{-1}\max(0, \bar{p} - \beta - \alpha b) + \alpha^{-1}\min(0, \bar{p} + \beta - \alpha a) = 0. \end{aligned} \tag{30}$$

Note that, now the unknown that appears inside the pointwise max- and min-operators is $\bar{p}$, which, compared to $\bar{u} \in L^2(\Omega)$ obeys additional regularity. To make this more precise, we introduce the operator $\mathcal{S} := -A^{-\star}A^{-1} : L^2(\Omega) \to H_0^1(\Omega)$ and denote by $h := -A^{-\star}(A^{-1}f - y_d) \in H_0^1(\Omega)$. Then, (5) can be written as

$$\bar{p} = \mathcal{S}\bar{u} + h.$$

Let us consider the mapping

$$\mathcal{T} : L^2(\Omega) \to L^s(\Omega) \text{ with } s \in \begin{cases} (2, \infty] & \text{for } n = 1, \\ (2, \infty) & \text{for } n = 2, \\ (2, \frac{2n}{(n-2)}] & \text{for } n \geq 3, \end{cases} \tag{31}$$

defined by $\mathcal{T}u = p = \mathcal{S}u + h$. Strictly speaking, $\mathcal{T}u = \mathcal{I}(\mathcal{S}u + h)$ with $\mathcal{I}$ denoting the Sobolev embedding (see e.g. [1]) of $H_0^1(\Omega)$ into $L^s(\Omega)$ with s as defined in (31). From the above considerations, it follows that $\mathcal{T}$ is well-defined and continuous. Since $\mathcal{T}$ is affine, it is also Fréchet differentiable from $L^2(\Omega)$ to $L^s(\Omega)$. Replacing $\bar{p}$ in (30) by $\mathcal{T}\bar{u}$ motivates to define $\mathcal{F} : L^2(\Omega) \to L^2(\Omega)$ by

$$\begin{aligned} \mathcal{F}(u) := u &- \alpha^{-1}\max(0, \mathcal{T}u - \beta) - \alpha^{-1}\min(0, \mathcal{T}u + \beta) \\ &+ \alpha^{-1}\max(0, \mathcal{T}u - \beta - \alpha b) + \alpha^{-1}\min(0, \mathcal{T}u + \beta - \alpha a). \end{aligned} \tag{32}$$

This allows to express the optimality system (11) in the compact form

$$\mathcal{F}(u) = 0. \tag{33}$$

After the preparations above, we are now able to argue generalized differentiability of the function $\mathcal{F}$ and derive a generalized Newton iteration for the solution of (33) and thus for $(\mathcal{P})$.

Theorem 10. *The function $\mathcal{F}$ as defined in* (32) *is generalized differentiable in the sense of Definition 7. A generalized derivative is given by*

$$\mathcal{G}(u)(v) = v - \alpha^{-1}\chi_{\mathcal{A}}(\mathcal{T}v) - \alpha^{-1}\chi_{\mathcal{B}}(\mathcal{T}v) + \alpha^{-1}\chi_{\mathcal{C}}(\mathcal{T}v) + \alpha^{-1}\chi_{\mathcal{D}}(\mathcal{T}v), \tag{34}$$

where

$$\begin{aligned} \mathcal{A} &= \{x \in \Omega : \mathcal{T}u - \beta \geq 0 \text{ a.e. in } \Omega\}, \\ \mathcal{B} &= \{x \in \Omega : \mathcal{T}u + \beta \leq 0 \text{ a.e. in } \Omega\}, \\ \mathcal{C} &= \{x \in \Omega : \mathcal{T}u - \beta - \alpha b \geq 0 \text{ a.e. in } \Omega\}, \\ \mathcal{D} &= \{x \in \Omega : \mathcal{T}u + \beta - \alpha a \leq 0 \text{ a.e. in } \Omega\}. \end{aligned}$$

Proof: After the above discussion, the prove is an application of the general theory from Section 4.1 to (32). For showing that the conditions required for the application of Theorem 9 are satisfied, we restrict ourselves to the first max-term in (32). Analogous reasoning yields generalized differentiability of the remaining terms in (32) and thus of $\mathcal{F}$. From the smoothing property of the affine operator $\mathcal{T}$ we obtain for each $u \in L^2(\Omega)$ that $\mathcal{T}u \in L^s(\Omega)$ with some $s > 2$. Thus, from Lemma 8 it follows that

$$\mathcal{F}_1 : u \to \max(0, \mathcal{T}u - \beta)$$

is semismooth in the sense of Definition 7 if considered as mapping from $L^2(\Omega)$ into $L^2(\Omega)$; Moreover, for its generalized derivative we obtain

$$\mathcal{G}_1(u)(v) = \chi_{\mathcal{A}}(\mathcal{T}v),$$

where $\chi_{\mathcal{A}}$ denotes the characteristic function for the set $\mathcal{A} = \{x \in \Omega : \mathcal{T}u - \beta \geq 0 \text{ a.e. in } \Omega\}$, compare with (27). A similar argumentation for the remaining max- and min-operators in (32) shows that the whole function $\mathcal{F}$ is generalized differentiable and that a generalized derivative is given by (34), which ends the proof. ■

We can now state our algorithm for the solution of $(\mathcal{P})$.

Algorithm 1 (semismooth Newton).

(1) Initialize $u^0 \in L^2(\Omega)$ and set $k := 0$.

(2) Unless some stopping criterion is satisfied, compute the generalized derivative $\mathcal{G}(u^k)$ as given in (34) and derive δu^k from

$$\mathcal{G}(u^k)\delta u^k = -\mathcal{F}(u^k). \tag{35}$$

(3) Update $u^{k+1} := u^k + \delta u^k$, set $k := k + 1$ and return to Step 1.

Now, Theorem 9 applies and yields the following convergence result for Algorithm 1:

Theorem 11. *Let the initialization u^0 be sufficiently close to the solution $\bar{u}$ of $(\mathcal{P})$. Then the iterates u^k of Algorithm 1 converge superlinearly to $\bar{u}$ in $L^2(\Omega)$. Moreover, the corresponding states y^k converge superlinearly to $\bar{y}$ in $H_0^1(\Omega)$.*

Proof: To apply Theorem 9, it remains to show that the generalized derivative (34) is invertible and that the norms of the inverse linear mappings are bounded. Both things follow with some calculations (that we skip here) from the fact that the system solved in each step constitutes the optimality system for a uniquely solvable auxiliary problem. ∎

We conclude this section with the statement of a different, more explicit form for the Newton step (35) in Algorithm 1. This alternative formulation also justifies the name "primal-dual active set strategy" for Algorithm 1.

Utilizing (34), the explicit statement of the Newton step (35) is

$$\begin{aligned}\delta u^k - \alpha^{-1}\chi_{\mathcal{B}_-^k}(\mathcal{T}\delta u^k) - \alpha^{-1}\chi_{\mathcal{B}_+^k}(\mathcal{T}\delta u^k) + \alpha^{-1}\chi_{\mathcal{A}_a^k}(\mathcal{T}\delta u^k) + \alpha^{-1}\chi_{\mathcal{A}_b^k}(\mathcal{T}\delta u^k)&\\ = -u^k + \alpha^{-1}\chi_{\mathcal{B}_-^k}(\mathcal{T}u^k + \beta) + \alpha^{-1}\chi_{\mathcal{B}_+^k}(\mathcal{T}u^k - \beta)&\\ - \alpha^{-1}\chi_{\mathcal{A}_a^k}(\mathcal{T}u^k + \beta - \alpha b) + \alpha^{-1}\chi_{\mathcal{A}_b^k}(\mathcal{T}u^k - \beta - \alpha a),&\end{aligned} \tag{36}$$

where

$$\mathcal{B}_-^k = \{x \in \Omega : \mathcal{T}u^k + \beta \leq 0 \text{ a.e. in } \Omega\}, \tag{37a}$$

$$\mathcal{B}_+^k = \{x \in \Omega : \mathcal{T}u^k - \beta \geq 0 \text{ a.e. in } \Omega\}, \tag{37b}$$

$$\mathcal{A}_a^k = \{x \in \Omega : \mathcal{T}u^k + \beta - \alpha a \leq 0 \text{ a.e. in } \Omega\}, \tag{37c}$$

$$\mathcal{A}_b^k = \{x \in \Omega : \mathcal{T}u^k - \beta - \alpha b \geq 0 \text{ a.e. in } \Omega\}. \tag{37d}$$

Note that $\mathcal{A}_a^k \subset \mathcal{B}_-^k$ and $\mathcal{A}_b^k \subset \mathcal{B}_+^k$. To obtain a disjoint splitting of Ω, we introduce the sets

$$\mathcal{I}_-^k := \mathcal{B}_-^k \setminus \mathcal{A}_a^k, \quad \mathcal{I}_+^k := \mathcal{B}_+^k \setminus \mathcal{A}_b^k, \quad \mathcal{A}_o^k := \Omega \setminus (\mathcal{B}_-^k \cup \mathcal{B}_+^k). \tag{38}$$

These definitions result in the following disjoint splitting of Ω:

$$\Omega = \mathcal{A}_a^k \,\dot\cup\, \mathcal{I}_-^k \,\dot\cup\, \mathcal{A}_o^k \,\dot\cup\, \mathcal{I}_+^k \,\dot\cup\, \mathcal{A}_b^k. \tag{39}$$

Observe that only the iterates for the control u^k appear explicity in Algorithm 1. The corresponding iterates for the state, the adjoint state and the multiplier are, in terms of u^k, given by

$$p^k = \mathcal{T}u^k, \quad y^k = A^{-1}(u^k + f) \text{ and } \mu^k = \mathcal{T}u^k - \alpha u^k, \tag{40}$$

compare with the definition of $\mathcal{T}$ on page 16 and with (11a)–(11c).

Concerning the sets defined in (37), we remark that for $c = \alpha^{-1}$ we obtain

$$\alpha^{-1}\mathcal{T}u^k = \alpha^{-1}p^k = u^k + c(p^k - \alpha u^k) = u^k + c\mu^k. \tag{41}$$

Therefore, replacing $\mathcal{T}u^k$ by $u^k + c\mu^k$ in (37), the sets $\mathcal{B}_-^k, \mathcal{B}_+^k, \mathcal{A}_a^k$ and $\mathcal{A}_b^k$ can equivalently be determined using both the primal variable u^k and the dual variable (i.e. the Lagrange multiplier) μ^k. This relates (37) to the NCP function (11d) and also explains the notion *primal-dual* in the restatement of Algorithm 1 below. Utilizing the above preparations, we can dicuss the iteration rule (36) separately on the subsets of the splitting (39). To start with, on $\mathcal{A}_a^k$ we obtain

$$\delta u^k - \alpha^{-1}\mathcal{T}\delta u^k + \alpha^{-1}\mathcal{T}\delta u^k = -u^k + \alpha^{-1}(\mathcal{T}u^k + \beta) - \alpha^{-1}(\mathcal{T}u^k + \beta - \alpha a),$$

where we used that $\mathcal{A}_a^k \subset \mathcal{B}_-^k$. Since $\delta u^k = u^{k+1} - u^k$, this results in the setting

$$u^{k+1} = a \quad \text{on} \quad \mathcal{A}_a^k.$$

Next we turn to $\mathcal{I}_-^k$. From (36) we obtain

$$\delta u^k - \alpha^{-1}\mathcal{T}\delta u^k = -u^k + \alpha^{-1}(\mathcal{T}u^k + \beta)$$

and, with $\delta u^k = u^{k+1} - u^k$, that $u^{k+1} - \alpha^{-1}\mathcal{T}u^{k+1} = \alpha^{-1}\beta$. Multiplying with α and using (40) yields

$$\mu^{k+1} = -\beta \ \text{ on } \ \mathcal{I}_-^k.$$

Continuing the evaluation of (36) separately on the remaining sets of the splitting (39) and using (40) yields

$$\begin{aligned} u^{k+1} &= 0 \ \text{ on } \ \mathcal{A}_o^k, \\ \mu^{k+1} &= \beta \ \text{ on } \ \mathcal{I}_+^k, \\ u^{k+1} &= b \ \text{ on } \ \mathcal{A}_b^k. \end{aligned}$$

Thus, we can restate Algorithm 1 as primal-dual active set method.

Algorithm 2 (primal-dual active set method).

(1) Initialize $u^0 \in L^2(\Omega)$, compute y^0, p^0 and μ^0 as in (40) and set $k := 0$.
(2) Unless some stopping criterion is satisfied, derive the sets $\mathcal{A}_a^k, \mathcal{I}_-^k, \mathcal{A}_o^k, \mathcal{I}_+^k$ and $\mathcal{A}_b^k$ following (37), (38) and (41), i.e.

$$\begin{aligned} \mathcal{A}_a^k &= \{x \in \Omega : u^k + \alpha^{-1}(\mu^k + \beta) \le a \text{ a.e. in } \Omega\}, \\ \mathcal{I}_-^k &= \{x \in \Omega : a < u^k + \alpha^{-1}(\mu^k + \beta) \le 0 \text{ a.e. in } \Omega\}, \\ \mathcal{A}_o^k &= \{x \in \Omega : |u^k + \alpha^{-1}\mu^k| < \beta \text{ a.e. in } \Omega\}, \\ \mathcal{I}_+^k &= \{x \in \Omega : 0 \le u^k + \alpha^{-1}(\mu^k - \beta) < b \text{ a.e. in } \Omega\}, \\ \mathcal{A}_b^k &= \{x \in \Omega : u^k + \alpha^{-1}(\mu^k - \beta) \ge b \text{ a.e. in } \Omega\}. \end{aligned}$$

(3) Solve for $(u^{k+1}, y^{k+1}, p^{k+1}, \mu^{k+1})$:

$$\begin{aligned}
&Ay^{k+1} = u^{k+1} + f,\\
&A^\star p^{k+1} = y_d - y^{k+1},\\
&-p^{k+1} + \alpha u^{k+1} + \mu^{k+1} = 0,\\
&u^{k+1} = a \quad \text{on} \quad \mathcal{A}_a^k,\\
&\mu^{k+1} = -\beta \ \text{ on } \ \mathcal{I}_-^k,\\
&u^{k+1} = 0 \ \text{ on } \ \mathcal{A}_o^k,\\
&\mu^{k+1} = \beta \ \text{ on } \ \mathcal{I}_+^k,\\
&u^{k+1} = b \ \text{ on } \ \mathcal{A}_b^k.
\end{aligned}$$

(4) Set $k := k + 1$ and return to Step 2.

5. Numerical Examples

We end this paper with a numerical study. Our aim is twofold: Firstly, we examine the influence of the L^1-norm on the structure of solutions of $(\mathcal{P})$ and numerically verify our theoretical findings in Section 3. Secondly, we study the performance of our algorithm for the solution of $(\mathcal{P})$.

As initialization u^0 for Algorithm 1 (or equivalently Algorithm 2) we choose the solution of (11) with $\bar{\mu} = 0$, that is, the solution of $(\mathcal{P})$ with $\beta = 0$ and $U_{ad} = L^2(\Omega)$. We terminate the algorithm if the sets $\mathcal{A}_a^k$,$\mathcal{I}_-^k$,$\mathcal{A}_o^k$,$\mathcal{I}_+^k$,$\mathcal{A}_b^k$ coincide for two consecutive iterations or as soon as the discrete analogue of $\|u^{k+1} - u^k\|_{L^2}$ drops below the tolerance $\varepsilon = 10^{-10}$. If the linear systems in Algorithm 2 are solved exactly, the first stopping criterion yields the exact solution of $(\mathcal{P})$. To be more precise, after discretization this stopping rule results in the exact solution of the discrete analogue of $(\mathcal{P})$.

Subsequently, we focus on the following test problems. We use $\Omega = [0, 1]^2$ and, unless otherwise specified, $A = -\Delta$.

Example 1. The data for this example are as follows: $a \equiv -30$, $b \equiv 30$, $y_d = \sin(2\pi x)\sin(2\pi y)\exp(2x)/6$, $f = 0$ and $\alpha = 0.0001$. We study the influence of the parameter β onto the solution. For $\beta = 0.001$, the optimal control $\bar{u}$ with corresponding condensed multiplier $\bar{\mu}$, the optimal state and the splitting into the active/inactive sets are shown in Figure 1.

Example 2. This second example is constructed in order to obtain sets $\mathcal{A}_a$, $\mathcal{I}_-$, $\mathcal{A}_o$, $\mathcal{I}_+$, $\mathcal{A}_b$ that have a more complex structure at the solution. Moreover,

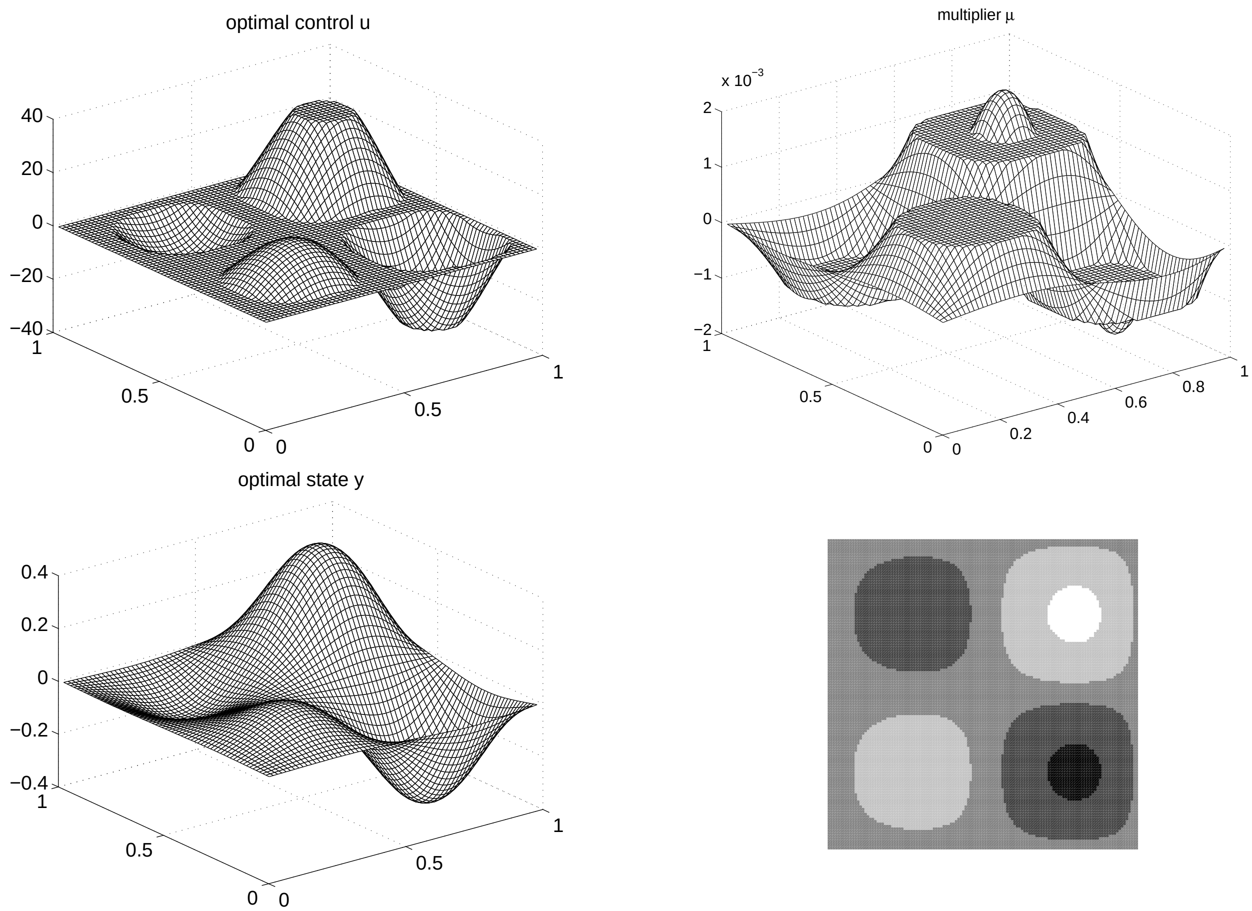

FIGURE 1. Example 1: Optimal control $\bar{u}$ (upper left), corresponding multiplier $\bar{\mu}$ (upper right), optimal state $\bar{y}$ (lower left), and visualization of the splitting (lower right) according to (39) into $\mathcal{A}_a$ (in black), $\mathcal{I}_-$ (in dark grey), $\mathcal{A}_o$ (in middle grey), $\mathcal{I}_+$ (in light grey) and $\mathcal{A}_b$ (in white).

the upper control bound b is zero on a part of Ω with positive measure. The exact data are $a \equiv -10$,

$$b = \begin{cases} 0 & \text{for } (x, y) \in [0, 1/4] \times [0, 1], \\ -5 + 20x & \text{for } (x, y) \in [1/4, 1] \times [0, 1], \end{cases}$$

$y_d = \sin(4\pi x)\cos(8\pi y)\exp(2x)$, $f = 10\cos(8\pi x)\cos(8\pi y)$, $\alpha = 0.0002$ and $\beta = 0.002$. The solution $(\bar{y}, \bar{u})$, as well as the corresponding Lagrange multiplier are shown in (2). Moreover we visualize the splitting (39) at the solution.

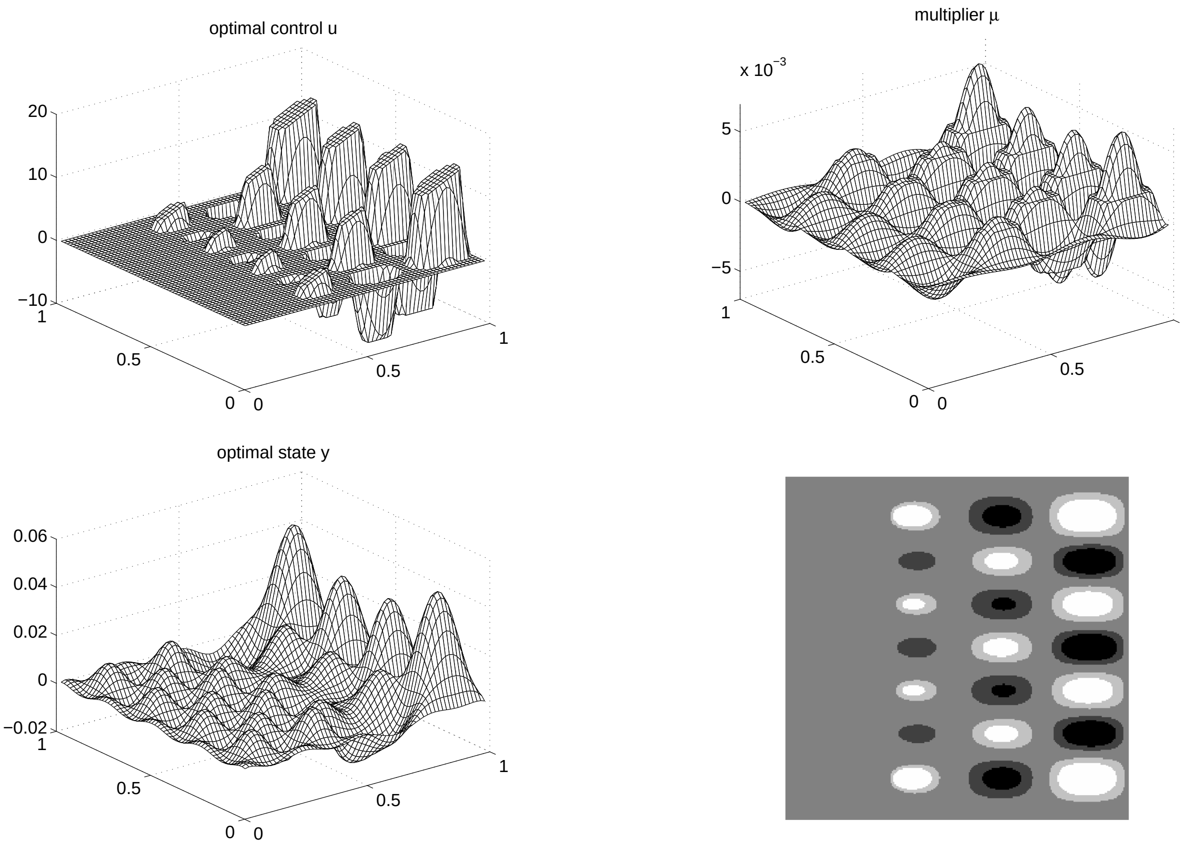

FIGURE 2. Example 2: Optimal control $\bar{u}$ (upper left), corresponding multiplier $\bar{\mu}$ (upper right), optimal state $\bar{y}$ (lower left), and visualization of the splitting (lower right) according to (39) into $\mathcal{A}_a$ (in black), $\mathcal{I}_-$ (in dark grey), $\mathcal{A}_o$ (in middle grey), $\mathcal{I}_+$ (in light grey) and $\mathcal{A}_b$ (in white).

5.1. Example 3. For this example we use the differential operator $A = \nabla a(x,y)\nabla$, with $a(x,y) = y^2 + 0.05$. The remaining data are given by $y_d \equiv 0.5$, $f = 0$ and $\alpha = 0.0001$. We do not assume box constraints for the control, i.e. $a \equiv -\infty$ and $b \equiv \infty$. In Figure 3, the optimal control $\bar{u}$ and the corresponding multiplier $\bar{\mu}$ are shown for $\beta = 0.005$.

5.2. Qualitative discussion of the results.

Complementarity conditions. We start with visually verifying the complementarity conditions for Example 1, see Figure 1, upper row. First note that $\bar{u} = 0$ on a relatively large part of Ω. Further, observe that on this part $|\bar{\mu}| \leq \beta$ holds, as expected. On subdomains with $0 \leq u \leq b$, the condensed

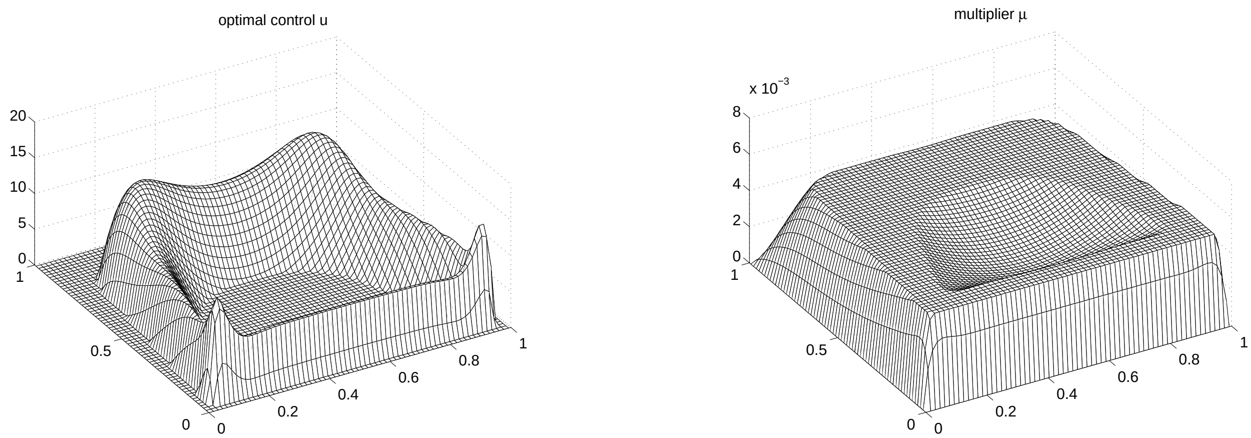

FIGURE 3. Example 3: Optimal control $\bar{u}$ (left) and corresponding multiplier $\bar{\mu}$ (right).

multiplier $\mu = \beta$; Moreover, $\bar{u} = b$ corresponds to sets where $\mu \geq \beta$, as desired. Similarly, on can verify visually that the complementarity conditions hold for $\bar{u} \leq 0$ as well.

The role of β – placement of control devices. To show the influence of the parameter β on the optimal control $\bar{u}$, we solve Example 1 for various values of β while keeping $\alpha = 0.0001$ fixed. For $\beta = 0.02$ or larger, the optimal control $\bar{u}$ is identically zero, compare with Lemma 4. As β decreases, the size of the region with $\bar{u}$ different from zero increases. In Figure 4 we depict the optimal controls for $\beta = 0.008$, 0.003, 0.0005 and 0. To realize the optimal control for $\beta = 0.008$ in an application, only two relatively small control devices are needed since $\bar{u}$ is zero on large parts of the domain. This means that no distributed control device that acts on the whole of Ω is necessary. Note that the solution for $\beta = 0$ shown in Figure 4 is the solution for the classical smooth optimal control problem $(\mathcal{P}_2)$.

Next, we turn to Example 3. In Figure 5, we visualize for different values of β those parts of Ω, where control devices need to be placed (i.e. where the optimal control is nonzero). It can be seen that the control area shrinks for increasing β. Moreover, for different β also the shape of the domain with nonzero control changes significantly.

Derivatives with respect to β. Now, we derive the right-side directional derivatives $\dot{u}_\beta$ using Theorem 6. To be precise, we use an extended version of this theorem, since we also allow for box constraints on the control. Without proof we remark that, for strongly active box constraints (i.e. where the Lagrange multiplier satisfies $|\bar{\mu}| > \beta$), $\dot{u}_\beta = 0$ holds. For $\alpha = 0.0001$ and

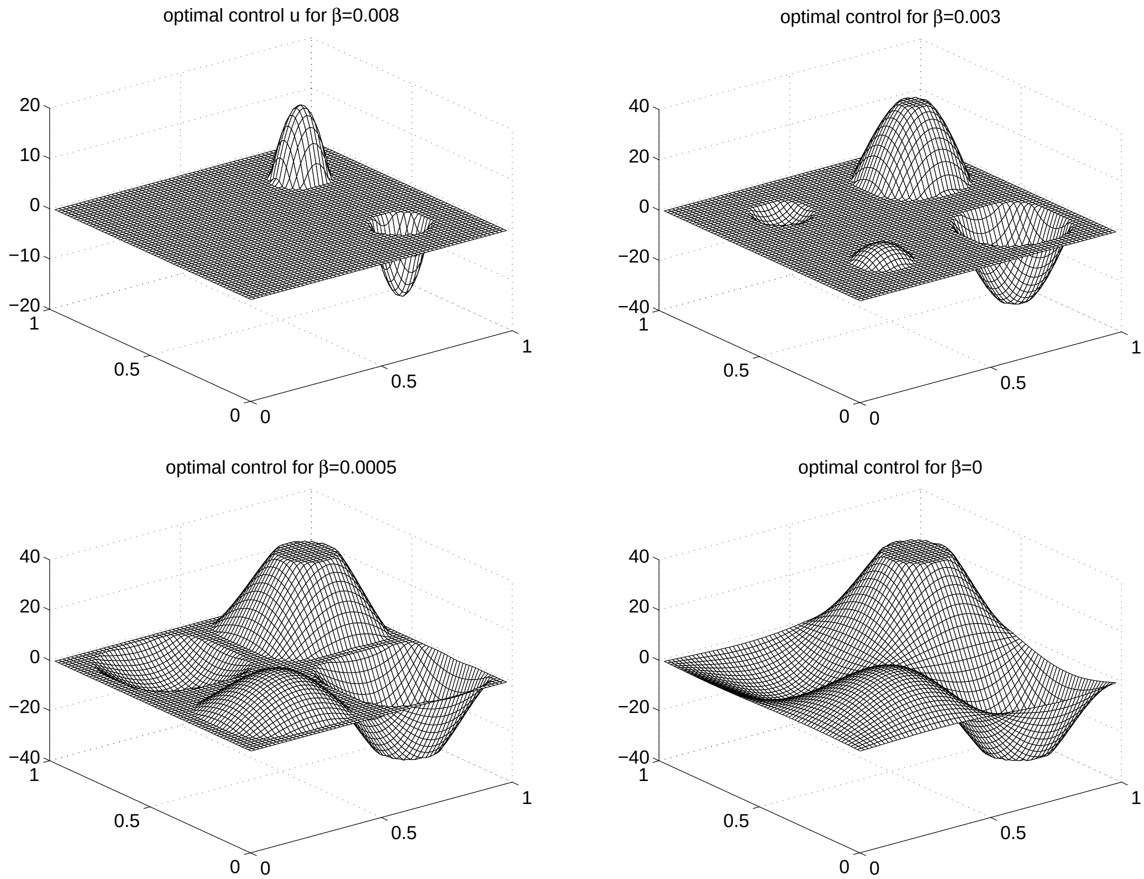

FIGURE 4. Example 1: Optimal control $\bar{u}$ for $\beta = 0.008$ (upper left), $\beta = 0.003$ (upper right), $\beta = 0.0005$ (lower right) and $\beta = 0$ (lower left).

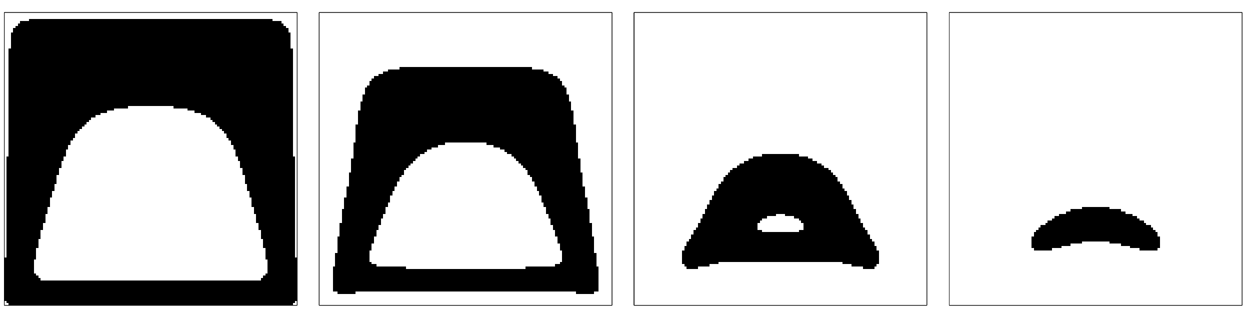

FIGURE 5. Example 3: Visualization of sets with nonzero optimal control $\bar{u}$ (black) for $\beta = 0.001, 0.01, 0.05, 0.1$ (from left to right); Only in the black parts of Ω control devices are needed.

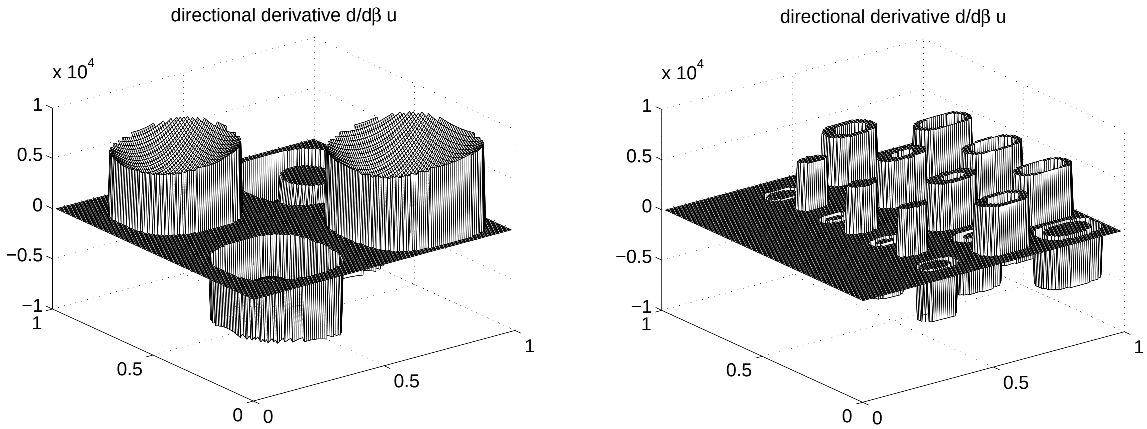

FIGURE 6. Directional derivatives $\dot{u}_\beta$ for $\beta = 0.001$ in Example 1 (left) and $\beta = 0.002$ in Example 2 (right).

$\beta = 0.001$, the result is shown on the left of Figure 6. Observe that the sensitivity for changes in β in only nonzero on $\mathcal{I}_\beta^- \cup \mathcal{I}_\beta^+$, since in this example numerical strict complementarity holds. As observed at the end of Section 3, $\dot{u}_\beta$ is noncontinuous with jumps approximate magnitude $\alpha^{-1} = 10^4$. On the right hand side of Figure 6, we depict $\dot{u}_\beta$ for Example 2, where $\beta = 0.002$ and $\alpha = 0.0002$. Again, the height of the jump along the boundaries of zones where $\dot{u}_\beta = 0$ is of magnitude $\alpha^{-1} = 5000$.

5.3. Performance of the algorithm.

Number of iterations. In Table 1 we show the number of iterations required for the solution of Example 1 for various values of β. For all mesh-sizes h and choices of β the algorithm yields an efficient behavior. Note also the stable behavior for various meshsizes h. The parameter β does not have a significant influence on the performance of the algorithm. Thus, the computational effort for solving problem $(\mathcal{P})$ is comparable to the one for the solution of $(\mathcal{P}_2)$.

Though in the second example the solution obeys a more complex structure of active sets, the algorithm detects the solution after 3 iterations for meshes with $h = 1/32, \ldots 1/256$.

Convergence rate. For Example 3, we study the convergence of the iterates u^k to the optimal control $\bar{u}$. In Table 2, we show the L^2-norm of the differences $u^k - \bar{u}$ and the quotients $\|u^k - \bar{u}\|_{L^2} / \|u^{k-1} - \bar{u}\|_{L^2}$ for $\alpha = 0.0001$ and $\beta = 0.005$. Observe that after a few iterations, the quotients decrease monotonically.

TABLE 1. Number of iterations for Example 1 with $\alpha = 0.0001$ for meshsize h and various values for β.

h	β				
	0.008	0.003	0.001	0.0005	0
1/32	5	5	4	4	3
1/64	5	5	4	4	4
1/128	5	5	5	4	4
1/256	6	6	5	5	4

This numerically verifies the local superlinear convergence rate proven in Theorem 9.

TABLE 2. Convergence of u^k in Example 3, $h = 1/128$.

k	$\|u^k - \bar{u}\|$	$\|u^k - \bar{u}\|/\|u^{k-1} - \bar{u}\|$
1	$4.588585e + 00$	$-$
2	$5.045609e + 00$	$1.099600e - 00$
3	$2.148787e - 01$	$4.258726e - 02$
4	$2.054260e - 01$	$9.560091e - 01$
5	$5.556892e - 02$	$2.705058e - 01$
6	$6.047484e - 04$	$1.088285e - 02$
7	0.000000	0.000000

Attempts to speed up the algorithm when solving ($\mathcal{P}$) *for various* β. If one is interested in placing control devices in an optimal way, one needs to solve ($\mathcal{P}$) for several values of β in order to find a control structure that is realizable with the available resources. We considered two ideas to speed up the algorithm for this case: The first one used an available solution for β as initialization for the algorithm to solve ($\mathcal{P}$) for β'. The second one used $u_\beta + (\beta' - \beta)\dot{u}_\beta$ as initialization for ($\mathcal{P}$) with β'. We tested both approaches when we solved Example 3 for various β. Unfortunately, we did not observe a significant speedup of the iteration.

Speeding up the algorithm using a nested iteration. Here, we use a prolongated solution on a rough grid as initialization on a finer grid. We start with a very rough grid and iteratively repeat the process (solve – prolongate – solve of next finer grid – ...) until the desired mesh-size is obtained. In Table 3, we give the results obtained for Example 3 with $\beta = 0.001$ and $\beta = 0.01$, where $h = 1/256$ for the finest grid. Using the nested strategy, only 3 iterations are needed on the finest grid, compared to 11 when the iteration is only done on that grid. Since the effort on the rougher meshes is small compared to the finer one, using the nested approach speeds up the solution process considerably (only 39% of CPU time is needed).

Table 3. Example 3: Number of iterations for nested iteration and for direct solution on finest grid.

h	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}	2^{-7}	2^{-8}	2^{-8}	CPU-time ratio
#iterations	3	2	2	6	4	3	3	11	0.39

6. Acknowledgement

The author would like to thank Isabel Figueiredo (University of Coimbra, Portugal), Michael Hintermüller and Karl Kunisch (both University of Graz, Austria) for helpful discussions and comments on a previous version of this paper.

References

[1] R. A. Adams. *Sobolev Spaces*. Academic Press, New York, 1975.

[2] U. M. Ascher, E. Haber, and H. Huang. On effective methods for implicit piecewise smooth surface recovery. *SIAM J. Sci. Comput.*, 28(1):339–358 (electronic), 2006.

[3] M. Bergounioux, K. Ito, and K. Kunisch. Primal-dual strategy for constrained optimal control problems. *SIAM J. Control Optim.*, 37(4):1176–1194, 1999.

[4] D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, Belmont, 1999.

[5] Tony F. Chan and Xue-Cheng Tai. Identification of discontinuous coefficients in elliptic problems using total variation regularization. *SIAM J. Sci. Comput.*, 25(3):881–904 (electronic), 2003.

[6] Xiaojun Chen, Zuhair Nashed, and Liqun Qi. Smoothing methods and semismooth methods for nondifferentiable operator equations. *SIAM J. Numer. Anal.*, 38(4):1200–1216, 2000.

[7] F. H. Clarke. *Optimization and nonsmooth analysis*. Canadian Mathematical Society Series of Monographs and Advanced Texts. John Wiley & Sons Inc., New York, 1983. A Wiley-Interscience Publication.

[8] L. Costa, I.N. Figueiredo, R. Leal, P. Oliveira, and G. Stadler. Modeling and study of sensor and actuator capablilities of a piezoelectric laminated plate. Technical report, CMUC preprint 06-21, Portugal, 2006.

[9] S. Durand and M. Nikolova. Stability of the minimizers of least squares with a non-convex regularization. I. Local behavior. *Appl. Math. Optim.*, 53(2):185–208, 2006.

[10] I. Ekeland and R. Temam. *Convex Analysis and Variational Problems*. North-Holland Publishing Company, Amsterdam, 1976.

[11] I.N. Figueiredo and C. Leal. A piezoelectric anisotropic plate model. *Asymptotic Analysis*, 44(3–4):327–346, 2005.

[12] T. Gilberg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Springer-Verlag, Berlin, 1983.

[13] R. Glowinski. *Numerical Methods for Nonlinear Variational Inequalities*. Springer-Verlag, New York, 1984.

[14] T. Grund and A. Rösch. Optimal control of a linear elliptic equation with a supremum-norm functional. *Optim. Meth. Softw.*, 15:299–329, 2001.

[15] M. Hintermüller, K. Ito, and K. Kunisch. The primal-dual active set strategy as a semi-smooth Newton method. *SIAM J. Optim.*, 13(3):865–888, 2003.
[16] M. Hintermüller and K. Kunisch. Path-following methods for a class of constrained minimization problems in function space. *SIAM J. Optim.*, 17(1):159–187 (electronic), 2006.
[17] M. Hintermüller and M. Ulbrich. A mesh-independence result for semismooth Newton methods. *Math. Prog., Ser. B*, 101(1):151–184, 2004.
[18] K. Ito and K. Kunisch. Augmented Lagrangian methods for nonsmooth convex optimization in Hilbert spaces. *Nonlinear Analysis, TMA*, 41:591–616, 2000.
[19] K. Ito and K. Kunisch. The primal-dual active set method for nonlinear optimal control problems with bilateral constraints. *SIAM J. Control Optim.*, 43(1):357–376 (electronic), 2004.
[20] M. Nikolova. Local strong homogeneity of a regularized estimator. *SIAM J. Appl. Math.*, 61(2):633–658 (electronic), 2000.
[21] M. Nikolova. Minimizers of cost-functions involving nonsmooth data-fidelity terms. Application to the processing of outliers. *SIAM J. Numer. Anal.*, 40(3):965–994 (electronic), 2002.
[22] L. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Phys. D*, 60(1-4):259–268, 1992.
[23] G. Stadler. Semismooth Newton and augmented Lagrangian methods for a simplified friction problem. *SIAM J. Optim.*, 15(1):39–62 (electronic), 2004.
[24] F. Tröltzsch. *Optimale Steuerung partieller Differentialgleichungen*. Vieweg, Wiesbaden, Germany, 2005.
[25] M. Ulbrich. Semismooth Newton methods for operator equations in function spaces. *SIAM J. Optim.*, 13(3):805–842, 2003.
[26] C. R. Vogel. *Computational methods for inverse problems*. Frontiers in Applied Mathematics. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.
[27] G. Vossen and H. Maurer. On L^1-minimization in optimal control and applications to robotics. *preprint*, 2006.
[28] M. Weiser. Interior point methods in function space. *SIAM J. Control Optim.*, 44(5):1766–1786 (electronic), 2005.
[29] M. Weiser, T. Gänzler, and A. Schiela. A control reduced primal interior point method for PDE constrained optimization. Report 04–38, ZIB, Berlin, 2004.

Georg Stadler
Center of Mathematics, University of Coimbra, Apartado 3008, 3001-454 Coimbra, Portugal
E-mail address: georgst@mat.uc.pt