SIAM J. CONTROL OPTIM.
Vol. 50, No. 2, pp. 943–963

DIRECTIONAL SPARSITY IN OPTIMAL CONTROL OF PARTIAL DIFFERENTIAL EQUATIONS*

ROLAND HERZOG†, GEORG STADLER‡, AND GERD WACHSMUTH†

Abstract. We study optimal control problems in which controls with certain sparsity patterns are preferred. For time-dependent problems the approach can be used to find locations for control devices that allow controlling the system in an optimal way over the entire time interval. The approach uses a nondifferentiable cost functional to implement the sparsity requirements; additionally, bound constraints for the optimal controls can be included. We study the resulting problem in appropriate function spaces and present two solution methods of Newton type, based on different formulations of the optimality system. Using elliptic and parabolic test problems we research the sparsity properties of the optimal controls and analyze the behavior of the proposed solution algorithms.

Key words. sparsity, optimal control, control device placement, L^1-norm minimization, semismooth Newton

AMS subject classifications. 49K20, 65K10, 49M15, 49J52

DOI. 10.1137/100815037

1. Introduction.

Optimal control problems with control costs of L^1-type are known to produce sparse solutions [8, 26], i.e., the control function is identically zero on parts of the domain. Since at these points no control needs to be applied, problems with sparsity terms naturally lend themselves to situations where the placement of control devices or actuators is not a priori given but is part of the problem.

In this paper, we analyze a general class of optimal control problems with a sparsity measure that promotes *striped sparsity patterns*. As a model problem in $L^2(\Omega)$, we consider

$$(\mathcal{P}_1)\qquad \begin{aligned} &\text{Minimize} && J(u) = \frac{1}{2}\|\mathcal{S}u - y_d\|_H^2 + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\,\|u\|_{1,2} \\ &\text{subject to} && u_a \le u \le u_b \quad \text{a.e. in } \Omega, \end{aligned}$$

where H is a Hilbert space corresponding to the state space of the controlled system, $y_d \in H$ is the desired state, and $\mathcal{S} : L^2(\Omega) \to H$ is the linear control-to-state mapping. Moreover, $-\infty \le u_a < 0 < u_b \le \infty$ are bound constraints for the optimal controls, $\alpha \ge 0$, $\beta > 0$, and $\|u\|_{1,2}$, which as specified below favors sparse solutions. To motivate the approach, in this introduction we assume that $\Omega = \Omega_1 \times (0, T)$ is a space-time cylinder, where $\Omega_1 \subset \mathbb{R}^n$, $n \ge 1$, and $T > 0$, in which case the directional sparsity term is given by

$$\|u\|_{1,2} := \int_{\Omega_1} \left(\int_0^T u^2 \,\mathrm{d}t \right)^{1/2} \mathrm{d}x. \tag{1.1}$$

*Received by the editors November 15, 2010; accepted for publication (in revised form) November 2, 2011; published electronically April 19, 2012.
http://www.siam.org/journals/sicon/50-2/81503.html
†Faculty of Mathematics, Chemnitz University of Technology, D–09107 Chemnitz, Germany (roland.herzog@mathematik.tu-chemnitz.de, gerd.wachsmuth@mathematik.tu-chemnitz.de).
‡Institute for Computational Engineering & Sciences, University of Texas at Austin, Austin, TX 78712 (georgst@ices.utexas.edu). The work of this author was partly supported by National Science Foundation grant CCF-0427985.

This term measures the L^1-norm in space of the L^2-norm in time. We will show that the term $\|u\|_{1,2}$ leads to certain, in applications often desirable, sparsity patterns for the optimal controls u. We refer to $(\mathcal{P}_1)$ as problem with a *directional sparsity* term.

To explain our interest in $(\mathcal{P}_1)$, consider a modified problem $(\mathcal{P}_0)$, in which $\|u\|_{1,2}$ is replaced by

$$\|u\|_{0,2} := \mu(\mathcal{C}) \quad \text{with } \mathcal{C} := \left\{ x \in \Omega_1 : \int_0^T u^2 \, \mathrm{d}t > 0 \right\} \subset \Omega_1, \tag{1.2}$$

where μ denotes the Lebesgue measure. The term $\|u\|_{0,2}$ accounts for costs originating from the sheer possibility to apply controls at certain locations. For example, this could be acquisition costs for control devices or actuators, in which case β is the price per control device unit. At a solution of $(\mathcal{P}_0)$, $\mathcal{C}$ is the set where placing control devices is most efficient. Terms such as (1.2) are also called "L^0-quasi-norms" since they can be found as (formal) limits of L^p-quasi-norms as $p \to 0$. Unfortunately, using the sparsity term (1.2) results in a nonconvex, highly nonlinear, and to some extent combinatorial problem since the set $\mathcal{C}$ depends on the (unknown) solution.

However, the directional sparsity problem $(\mathcal{P}_1)$ can be used as a convex relaxation of the L^0-problem. Partial theoretical justification for this approximation can, for instance, be found in [4, 11, 12]. We point out that using the "undirected" sparsity term $\beta \|u\|_{L^1(\Omega)}$ also results in a solution with a certain sparsity structure. However, this sparsity pattern changes over time. Compared to solutions based on the directional sparsity terms $\|u\|_{0,2}$ or $\|u\|_{1,2}$, where the sparsity structure is fixed over time, this is difficult to realize in practice.

Besides the practical relevance of $(\mathcal{P}_1)$ for finding optimal locations for control devices, studying $(\mathcal{P}_1)$ is also of theoretical interest. Since the functional (1.1) cannot be differentiated in a classical sense, subdifferentials have to be used and the construction of efficient solution algorithms becomes challenging. Even though choosing $\alpha > 0$ (which we mostly assume in this paper) allows us to formulate the problem in a Hilbert space framework, the problem structure and the characteristics of the solution are very different from smooth optimal control problems.

Let us put our work into perspective. Nonsmooth terms in optimization problems with PDEs are commonly being used for the purpose of regularization in inverse problems in image processing; see, for instance, [6, 23, 24, 30]. Among the few papers where L^1-regularization is used in the context of optimal control of PDEs are [5, 8, 26, 31, 32]. In [5, 8, 26, 32] elliptic optimal control problems with the nondirectional sparsity terms $\beta \|u\|_{L^1(\Omega)}$ are analyzed. In [5, 26, 32], in additional to the sparsity term, the squared L^2-norm $\alpha/2 \|u\|^2_{L^2(\Omega)}$ is part of the cost functional, which allows the problem to be analyzed and solved by a Newton-type algorithm in a Hilbert space framework. On the contrary, in [8] the problem is posed and analyzed in $L^1(\Omega)$ or $\mathrm{BV}(\Omega)$ and a Newton-type algorithm is proposed that applies to a regularized version of the problem. In [32], convergence rates as $\alpha \to 0$ are investigated. Compared to [8, 26, 32], in the problem $(\mathcal{P}_1)$ under consideration here, points in the control space are directionally coupled, which leads to a stronger nonlinearity in the optimality system. Additional challenges arise from the combination of pointwise bound constraints with the nonlocal sparsity term. In [31], the authors use L^1-regularization for a system of ODEs, which is driven toward a desired state at final time. Here, the nonsmooth objective is motivated by the physics of the problem (the control cost measures the consumption of fuel) rather than by the need for sparse optimal controls. Related research focuses on finding sparse solutions to finite-dimensional, linear

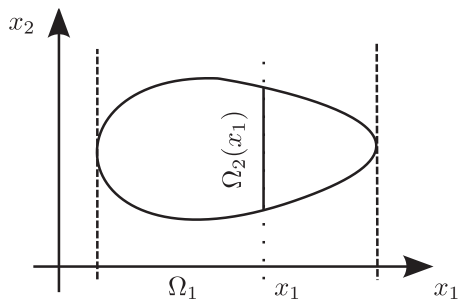

FIG. 1.1. *Two-dimensional example domain Ω, its projection Ω_1, and the cross section $\Omega_2(x_1)$.*

inverse problems. For underdetermined problems, the convex relaxed l_1-norm problem provably finds a near-optimal solution also with respect to l_0, i.e., the sparsest solution [4, 11, 12]; we refer to [29] for an overview of numerical methods employed in this context. In a different interpretation of the sparsity approach, the authors in [15] find optimal experimental designs for linear inverse problems. Here, a sparsity term is used to rule out experiments that do not contribute significant information for the inversion. In practice, this can reduce the number of experiments or measurements to be made.

The main focus of this paper is twofold. First, we analyze $(\mathcal{P}_1)$ in a function space framework, derive the optimality system, and study characteristics of the sparse optimal controls. Second, we propose two Newton-type algorithms for the numerical solution of $(\mathcal{P}_1)$ and study their convergence behavior. To the authors' knowledge, available algorithms for finite-dimensional counterparts of $(\mathcal{P}_1)$ are based only on first-order information; see, for example, [22, 25, 34].

The outline of the remainder of this paper is as follows. In section 2, we prove the existence and uniqueness of a solution for $(\mathcal{P}_1)$ and derive necessary and sufficient optimality conditions. In section 3, we present a particular form of the optimality system that applies in the absence of pointwise control constraints. Local superlinear convergence of a semismooth Newton method based on this form is proved. Section 4 is devoted to an alternative formulation of the optimality system, which allows taking into account pointwise control constraints. We also derive a semismooth Newton method based on this formulation and study its properties in finite dimensions. Finally, in section 5 properties of solutions of $(\mathcal{P}_1)$ are studied using numerical examples, and the efficiency of the algorithms proposed in this paper is analyzed.

Notation. While a time-dependent problem has been used in the introduction for illustration purposes, in the remainder of this paper we allow for a more general split of coordinate directions, which contains the time-dependent problem as a special case (see Example 1.2). Throughout the paper, $\Omega \subset \mathbb{R}^N$ with $N \geq 2$ is a bounded measurable set. The coordinate directions are partitioned according to $\mathbb{R}^N = \mathbb{R}^n \times \mathbb{R}^{N-n}$ for some $1 \leq n < N$. This induces the sets

$$\Omega_1 = \{x_1 \in \mathbb{R}^n : \exists x_2 \in \mathbb{R}^{N-n} \text{ with } (x_1, x_2) \in \Omega\}, \tag{1.3a}$$

$$\Omega_2(x_1) = \{x_2 \in \mathbb{R}^{N-n} : (x_1, x_2) \in \Omega\} \text{ for } x_1 \in \Omega_1. \tag{1.3b}$$

Hence, Ω_1 can be interpreted as the projection of Ω onto $\mathbb{R}^n$, whereas $\Omega_2(x_1)$ is the cross section of Ω at position $x_1 \in \mathbb{R}^n$ (see Figure 1.1). The general form for the directional sparsity term $\|u\|_{1,2}$ is

$$\|u\|_{1,2} := \int_{\Omega_1} \left(\int_{\Omega_2(x_1)} u^2(x_1, x_2)\, \mathrm{d}x_2 \right)^{1/2} \mathrm{d}x_1. \tag{1.4}$$

We now give two examples of problems that fit into our general framework.

Example 1.1 (elliptic PDE). The domain $\Omega \subset \mathbb{R}^N$ is a bounded domain with Lipschitz boundary, and $\mathcal{S} : L^2(\Omega) \to H := L^2(\Omega)$ denotes the solution map $u \mapsto y$ of the elliptic PDE

$$-\Delta y = u \text{ in } \Omega, \qquad y = 0 \text{ on } \Gamma.$$

Example 1.2 (parabolic PDE). This is a particular case of the example described in the introduction. Suppose that $\Omega_1 \subset \mathbb{R}^n$ is a bounded domain with Lipschitz boundary Γ, $\Omega_2 = (0,T)$ for some $T > 0$, and $\Omega = \Omega_1 \times \Omega_2$ is the space-time cylinder. Moreover, for $\kappa > 0$, $\mathcal{S} : L^2(\Omega) \to H := L^2(\Omega)$ denotes the solution operator $u \mapsto y$ for

$$y_t - \nabla \cdot (\kappa \nabla) y = u \text{ in } \Omega, \qquad y = 0 \text{ on } \Gamma \times (0,T), \qquad y(\cdot, 0) = 0 \text{ in } \Omega_1.$$

In both examples, the restriction to homogeneous boundary and initial data can be waived by modifying y_d. Note that problems with vector-valued control variables $u : \Omega \to \mathbb{R}^n$, or equivalently $u : \Omega \times \{1, 2, \ldots, n\} \to \mathbb{R}$, are included in our setting if the inner integral in (1.4) is taken with respect to the counting measure. The regularization term then becomes $\|u\|_{1,2} := \int_\Omega \|u(x)\|_{\mathbb{R}^n}\, \mathrm{d}x$, where $\|\cdot\|_{\mathbb{R}^n}$ denotes the Euclidean norm in $\mathbb{R}^n$. Problems of this type fall into the class of joint sparsity approaches; see [13, 14].

Function spaces. We denote by $L^p(\Omega)$, $L^p(\Omega_1)$, and $L^p(\Omega_2(x_1))$ the classical Lebesgue spaces on the respective domains ($1 \le p \le \infty$). The norm in these spaces is referred to as $\|\cdot\|_p$, and the domain will be clear from the context. We write $\langle \cdot\,, \cdot \rangle$ for the scalar product in L^2. For $1 \le p, q \le \infty$, we denote by $L^{p,q}(\Omega)$ the space of (equivalence classes of) all measurable functions $f : \Omega \to \mathbb{R}$ whose norm

$$\|f\|_{p,q} := \big\| \Omega_1 \ni x_1 \mapsto \|f(x_1, \cdot)\|_q \big\|_p$$

is finite. In case $\Omega_2(x_1)$ does not depend on x_1, this is the usual Bochner space $L^p(\Omega_1, L^q(\Omega_2))$. In the general case, it is a closed subspace of the latter with $\Omega_2 = \cup_{x_1 \in \Omega_1} \Omega_2(x_1)$ being the union of all cross sections. Nevertheless, we also refer to $L^{p,q}(\Omega)$ as a Bochner-type space. Finally, for functions u, v defined on Ω we introduce the shorthand notation

$$|u|_{x_1,2} := \left(\int_{\Omega_2(x_1)} u^2(x_1, x_2)\, \mathrm{d}x_2 \right)^{1/2} = \|u(x_1, \cdot)\|_2,$$

$$\langle u,\, v \rangle_{x_1,2} := \int_{\Omega_2(x_1)} u(x_1, x_2) \cdot v(x_1, x_2)\, \mathrm{d}x_2 = \langle u(x_1, \cdot),\, v(x_1, \cdot) \rangle.$$

That is, $|\cdot|_{x_1,2}$ is the L^2-norm and $\langle \cdot\,, \cdot \rangle_{x_1,2}$ the scalar product on the stripe $\{x_1\} \times \Omega_2(x_1)$. Note that for $u, v \in L^2(\Omega)$, the relations $|u|_{\cdot,2} \in L^2(\Omega_1)$ as well as $\langle u,\, v \rangle_{\cdot,2} \in L^1(\Omega_1)$ hold.

Standing assumptions. We consider problem $(\mathcal{P}_1)$ under the following assumptions. H is a Hilbert space, and $y_d \in H$ is the desired state. We assume that

$\mathcal{S} \in \mathcal{L}(L^2(\Omega), H)$ is a bounded linear map. For the analysis of the semismooth Newton method in section 3, we require

$$\mathcal{S}^\star\mathcal{S} : L^2(\Omega) \to L^{6,2}(\Omega) \text{ to be bounded, and } \mathcal{S}^\star y_d \in L^{6,2}(\Omega). \tag{1.5}$$

Here, we denoted by $\mathcal{S}^\star$ the adjoint of $\mathcal{S}$. Moreover, let $u_a, u_b \in L^2(\Omega)$ and denote by

$$U_{ad} := \{u \in L^2(\Omega) : u_a \leq u \leq u_b \text{ a.e. in } \Omega\} \tag{1.6}$$

the set of admissible controls. It is reasonable to suppose that $u_a < 0 < u_b$ holds a.e. in Ω, which makes $u \equiv 0$ an admissible control. Finally, we assume $\alpha \geq 0$, $\beta > 0$.

Note that (1.5) is satisfied for the examples mentioned above in reasonable dimensions. First we consider Example 1.1 with $N \leq 3$. Due to the smoothing property of the control-to-state mapping, $\mathcal{S}^\star$ maps $L^2(\Omega)$ continuously into $H_0^1(\Omega)$. Sobolev's embedding theorem ensures $H_0^1(\Omega) \hookrightarrow L^6(\Omega)$, and by Holder's inequality we can infer $L^6(\Omega) \hookrightarrow L^{6,2}(\Omega)$. Altogether, $\mathcal{S}^\star$ maps $L^2(\Omega)$ continuously into $L^{6,2}(\Omega)$. Let us now consider Example 1.2 with $n \leq 3$. The solution theory for parabolic equations yields the continuity of $\mathcal{S}^\star : L^2(\Omega) \to L^2(0,T;H_0^1(\Omega_1))$. By Sobolev's embedding theorem we have $L^2(0,T;H_0^1(\Omega_1)) \hookrightarrow L^2(0,T;L^6(\Omega_1))$. For functions $f \in L^2(0,T;L^6(\Omega_1))$ the continuous Minkowski inequality (see [9, p. 499]) yields $f \in L^6(\Omega_1;L^2(0,T))$ and moreover

$$\|f\|_{L^6(\Omega_1;L^2(0,T))} \leq \|f\|_{L^2(0,T;L^6(\Omega_1))}.$$

Therefore, we have $L^2(0,T;L^6(\Omega_1)) \hookrightarrow L^6(\Omega_1;L^2(0,T)) = L^{6,2}(\Omega)$. Altogether we infer the continuity of $\mathcal{S}^\star : L^2(\Omega) \to L^{6,2}(\Omega)$.

2. Analysis of the problem. In this section, we prove the existence and uniqueness of an optimal solution for $(\mathcal{P}_1)$ and derive necessary and sufficient optimality conditions. We begin by computing the subdifferential of $\|u\|_{1,2}$.

2.1. Computation of the subdifferential. Due to the separability of $L^2(\Omega_2)$ we can use a standard result [10, p. 282] to find the dual of the Bochner-type space $L^{1,2}(\Omega)$. A direct proof can be found in [33, Satz 3.11].

LEMMA 2.1. *The dual space of $L^{1,2}(\Omega)$ can be identified with the Bochner-type space $L^{\infty,2}(\Omega)$. This is the space (of equivalence classes) of all measurable functions φ defined on Ω whose norm*

$$\|\varphi\|_{\infty,2} := \operatorname*{ess\,sup}_{x_1\in\Omega_1} \left(\int_{\Omega_2(x_1)} \varphi^2(x_1,x_2)\,\mathrm{d}x_2\right)^{1/2}$$

is finite. The dual pairing is given by

$$\langle u, \varphi\rangle_{L^{1,2}(\Omega),L^{\infty,2}(\Omega)} = \int_\Omega u\cdot\varphi\,\mathrm{d}(x_1,x_2).$$

LEMMA 2.2. *The subdifferential of $\|\cdot\|_{1,2} : L^{1,2}(\Omega) \to \mathbb{R}$ at $u \in L^{1,2}(\Omega)$ is given by*

$$\partial\|\cdot\|_{1,2}(u) = \{v \in L^{\infty,2}(\Omega) : v(x_1,\cdot) \in \partial\|\cdot\|_2(u(x_1,\cdot)) \textit{ for all } x_1 \in \Omega_1\} \tag{2.1}$$

with the subdifferential of the $L^2(\Omega_2(x_1))$-norm

$$s \in \partial\|\cdot\|_2(w) \Leftrightarrow \begin{cases} \|s\|_2 \leq 1 & \textit{if } w = 0 \textit{ a.e. on } \Omega_2(x_1),\\ s = \frac{w}{\|w\|_2} & \textit{elsewhere,}\end{cases}$$

where $w \in L^2(\Omega_2(x_1))$. An equivalent characterization of the subdifferential of $\|\cdot\|_{1,2}$ is given by

$$v \in \partial\|\cdot\|_{1,2}(u) \Leftrightarrow \textit{for a.a. } x_1 \in \Omega_1 : \begin{cases} |v|_{x_1,2} \le 1 & \textit{if } |u|_{x_1,2} = 0, \\ v(x_1,\cdot) = \frac{u(x_1,\cdot)}{|u|_{x_1,2}} & \textit{if } |u|_{x_1,2} \neq 0. \end{cases}$$

The subdifferential remains the same if $\|\cdot\|_{1,2}$ is considered a function of $L^2(\Omega) \to \mathbb{R}$.

Proof. We begin by observing that (see [20, p. 56])

$$\partial\|\cdot\|_{1,2}(u) = \{v \in L^{\infty,2}(\Omega) : \|v\|_{\infty,2} \le 1,\ \langle u,\, v\rangle_{L^{1,2}(\Omega),L^{\infty,2}(\Omega)} = \|u\|_{1,2}\}. \tag{2.2}$$

To show that the right-hand sides in (2.2) and (2.1) coincide, let $v \in \partial\|\cdot\|_{1,2}(u)$ be given. The first condition in (2.2) yields $|v|_{x_1,2} \le 1$ for almost all $x_1 \in \Omega_1$. Furthermore, we have

$$\|u\|_{1,2} = \int_{\Omega_1} \underbrace{\int_{\Omega_2(x_1)} u \cdot v \,\mathrm{d}x_2}_{=:f(x_1)} \,\mathrm{d}x_1.$$

Now, for almost all $x_1 \in \Omega_1$ it holds that $f(x_1) = \langle u,\, v\rangle_{x_1,2} \le |u|_{x_1,2}|v|_{x_1,2} \le |u|_{x_1,2}$. Therefore, we have

$$\|u\|_{1,2} = \int_{\Omega_1} f(x_1)\,\mathrm{d}x_1 \le \int_{\Omega_1} |u|_{x_1,2}\,\mathrm{d}x_2 = \|u\|_{1,2}.$$

Together, these relations imply that $f(x_1) = |u|_{x_1,2}$ holds for almost all $x_1 \in \Omega_1$. This, in turn, shows $v(x_1,\cdot) \in \partial\|\cdot\|_2(u(x_1,\cdot))$. The converse inclusion can be proved easily.

If $\|\cdot\|_{1,2}$ is considered as a function from $L^2(\Omega)$ into $\mathbb{R}$, the subdifferential may become larger. However, it remains a subset of $L^{\infty,2}(\Omega)$, and thus the density of $L^2(\Omega)$ in $L^{1,2}(\Omega)$ shows that the subdifferentials coincide. Indeed, suppose $v_k \to v$ in $L^{1,2}(\Omega)$ with $v_k \in L^2(\Omega)$. Let g be a subgradient (w.r.t. $L^2(\Omega)$) of $\|\cdot\|_{1,2}$ at $u \in L^2(\Omega)$, i.e.,

$$\|w\|_{1,2} \ge \|u\|_{1,2} + \langle g,\, w - u\rangle \quad \text{for all } w \in L^2(\Omega).$$

Using $w = v_k$ and passing to the limit shows the same inequality with w replaced by $v \in L^{1,2}(\Omega)$. □

2.2. Existence, uniqueness, and optimality conditions.

LEMMA 2.3. *Problem $(\mathcal{P}_1)$ has a unique solution for $\beta \ge 0$ and $\alpha > 0$. The same holds for $\alpha = 0$ in case $\mathcal{S}$ is injective.*

Proof. For $\alpha > 0$ the objective $J(u)$ is strictly convex and continuous. The same holds true in case $\alpha = 0$, provided that $\mathcal{S}$ is injective. Furthermore, the set $U_{\mathrm{ad}} \subset L^2(\Omega)$ is convex and weakly compact and therefore the existence and uniqueness of the optimal control follows from standard arguments; see, for instance, [28, Thm. 2.14]. □

The variational inequality

$$\langle u - \bar{u},\ \partial J(\bar{u})\rangle \ge 0 \quad \text{for all } u \in U_{\mathrm{ad}} \tag{2.3}$$

is a necessary and sufficient optimality condition for $(\mathcal{P}_1)$. Hence the Moreau–Rockafellar theorem [20, p. 57] implies the following lemma.

LEMMA 2.4. *The function $\bar{u} \in L^2(\Omega)$ solves $(\mathcal{P}_1)$ if and only if*

$$\langle u - \bar{u},\, -\bar{p} + \beta\bar{\lambda} + \alpha\bar{u}\rangle \ge 0 \tag{2.4}$$

holds for all $u \in U_{\mathrm{ad}}$, where $\bar{p} = \mathcal{S}^\star(y_d - \mathcal{S}\bar{u})$ is the adjoint state and $\bar{\lambda} \in \partial\|\cdot\|_{1,2}(\bar{u})$ is a subgradient as defined in (2.1).

With the introduction of a Lagrange multiplier $\bar{\mu}$ associated with the condition $\bar{u} \in U_{\mathrm{ad}}$, the variational inequality (2.4) becomes

$$-\bar{p} + \beta\bar{\lambda} + \alpha\bar{u} + \bar{\mu} = 0, \tag{2.5a}$$

where $\bar{\lambda}$ and $\bar{\mu}$ satisfy

$$\bar{\lambda} \in \partial\|\cdot\|_{1,2}(\bar{u}), \qquad \bar{\mu}(x_1,x_2) \begin{cases} \le 0 & \text{if } \bar{u}(x_1,x_2) = u_a(x_1,x_2), \\ \ge 0 & \text{if } \bar{u}(x_1,x_2) = u_b(x_1,x_2), \\ = 0 & \text{else} \end{cases} \tag{2.5b}$$

in an almost everywhere sense. Sections 3 and 4 present two different formulations of this optimality system, which lead to different solution algorithms.

3. Semismooth Newton method I: The unconstrained case. Semismooth Newton methods are well studied in both finite-dimensional and function spaces, and they are commonly used for optimal control problems; see [2, 16, 21]. Under certain conditions, a local superlinear convergence rate and even global convergence can be proved. The aim of this section is to present a semismooth Newton method for the case that no bound constraints for the controls are present in $(\mathcal{P}_1)$, i.e., $U_{\mathrm{ad}} = L^2(\Omega)$. For this case we can show local superlinear convergence of the method in function space. To this end we state an alternative form of the optimality system (2.5) that is tailored to the unconstrained case. In this section, $\alpha > 0$ is assumed throughout.

LEMMA 3.1. *In the case $U_{\mathrm{ad}} = L^2(\Omega)$, $\bar{u} \in L^2(\Omega)$ is the optimal solution of $(\mathcal{P}_1)$ if and only if*

$$\bar{u} = \max\left(0, 1 - \frac{\beta}{|\bar{p}|_{\cdot,2}}\right)\frac{\bar{p}}{\alpha} \quad \textit{a.e. in } \Omega \tag{3.1}$$

holds, where $\bar{p} = \mathcal{S}^\star(y_d - \mathcal{S}\bar{u})$ is the adjoint state.

Proof. Let us show the equivalence of (3.1) and (2.5). Note that $U_{\mathrm{ad}} = L^2(\Omega)$ implies $\bar{\mu} = 0$ and let $x_1 \in \Omega_1$ be given. If $|\bar{u}|_{x_1,2} = 0$, both systems are equivalent to $|\bar{p}|_{x_1,2} \le \beta$. Suppose now that $|\bar{u}|_{x_1,2} > 0$. Then both systems imply that $\bar{u}(x_1,\cdot)$ is a multiple of $\bar{p}(x_1,\cdot)$. Now we have the equivalences

$$\begin{aligned} (2.5) \;\Leftrightarrow\; \bar{p}(x_1,\cdot) = (\alpha + \beta\,|u|^{-1}_{x_1,2})\,u(x_1,\cdot) \;\Leftrightarrow\; |\bar{p}|_{x_1,2} = \alpha\,|u|_{x_1,2} + \beta \\ \Leftrightarrow\; |\bar{p}|_{x_1,2} - \beta = \alpha\,|u|_{x_1,2} \;\Leftrightarrow\; (3.1), \end{aligned}$$

which shows the equivalence. $\square$

This lemma gives rise to the definition of the nonlinear operator $F : L^2(\Omega) \to L^2(\Omega)$

$$F(u) := u - \max\left(0, 1 - \frac{\beta}{|Pu|_{\cdot,2}}\right)\frac{Pu}{\alpha},$$

where $Pu := \mathcal{S}^\star(y_d - \mathcal{S}u)$ is the adjoint state belonging to u. As a direct consequence of Lemma 3.1, we have that $F(u) = 0$ if and only if u solves $(\mathcal{P}_1)$.

In the remainder of this section we prove the Newton (or slant) differentiability [7, 16] of F and the invertibility of the derivative $F'(u)$.

LEMMA 3.2. *The function $G : L^{6,2}(\Omega) \to L^2(\Omega)$ defined by*

$$p \mapsto \max\left(0, 1 - \frac{\beta}{|p|_{\cdot,2}}\right)p$$

is Newton differentiable on $L^{6,2}(\Omega)$ and a generalized derivative is given by

$$G'(p)\,h = \max\left(0, 1 - \frac{\beta}{|p|_{\cdot,2}}\right) h + \chi_p \frac{\beta\,\langle h,\, p\rangle_{\cdot,2}}{|p|_{\cdot,2}^3}\, p \quad \textit{with} \quad \chi_p = \begin{cases} 1 & \textit{if } |p|_{\cdot,2} > \beta, \\ 0 & \textit{else.} \end{cases}$$

Proof. Let $p,\, h \in L^{6,2}(\Omega)$ be given. We define the remainder $r(h)$ by

$$\begin{aligned} r(h) &= G(p+h) - G(p) - G'(p+h)\,h \\ &= \left[\max\left(0, 1 - \frac{\beta}{|p+h|_{\cdot,2}}\right) - \max\left(0, 1 - \frac{\beta}{|p|_{\cdot,2}}\right) - \chi_{p+h}\,\frac{\beta\,\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3}\right] p \\ &\qquad - \left[\chi_{p+h}\,\frac{\beta\,\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3}\right] h \\ &= (r_1 + r_2 + r_3)\,p - r_4\,h, \end{aligned}$$

where

$$\begin{aligned} r_1 &= \chi_{\mathcal{N}_1}\left(1 - \frac{\beta}{|p+h|_{\cdot,2}} - \left(1 - \frac{\beta}{|p|_{\cdot,2}}\right) - \beta\,\frac{\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3}\right), \\ r_3 &= \chi_{\mathcal{N}_3}\left(1 - \frac{\beta}{|p|_{\cdot,2}}\right), \\ r_2 &= \chi_{\mathcal{N}_2}\left(1 - \frac{\beta}{|p+h|_{\cdot,2}} - \beta\,\frac{\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3}\right), \\ r_4 &= \chi_{p+h}\,\frac{\beta\,\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3} \end{aligned}$$

with

$$\begin{aligned} \mathcal{N}_1 &:= \{x \in \Omega : |p+h|_{\cdot,2} > \beta,\ |p|_{\cdot,2} \geq \beta\}, \\ \mathcal{N}_2 &:= \{x \in \Omega : |p+h|_{\cdot,2} > \beta,\ |p|_{\cdot,2} < \beta\}, \\ \mathcal{N}_3 &:= \{x \in \Omega : |p+h|_{\cdot,2} \leq \beta,\ |p|_{\cdot,2} > \beta\}. \end{aligned}$$

Now we estimate the four parts of the sum for $r(h)$ separately.

First part. On $\mathcal{N}_1$ we have

$$\begin{aligned} r_1 &= \beta\left(-\frac{1}{|p+h|_{\cdot,2}} + \frac{1}{|p|_{\cdot,2}} - \frac{\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3}\right) \\ &= \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}}\left(|p+h|_{\cdot,2}^2\left(|p+h|_{\cdot,2} - |p|_{\cdot,2}\right) - |p|_{\cdot,2}\,\langle h,\, p+h\rangle_{\cdot,2}\right) \\ &\leq \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}}\left(|p+h|_{\cdot,2}\,\langle h,\, p+h\rangle_{\cdot,2} - |p|_{\cdot,2}\,\langle h,\, p+h\rangle_{\cdot,2}\right) \quad \text{by Cauchy–Schwarz} \\ &= \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}}\left(|p+h|_{\cdot,2} - |p|_{\cdot,2}\right)\langle h,\, p+h\rangle_{\cdot,2} \leq \frac{\beta}{|p+h|_{\cdot,2}^2\,|p|_{\cdot,2}}\,|h|_{\cdot,2}^2 \leq \frac{|h|_{\cdot,2}^2}{\beta^2}. \end{aligned}$$

Similarly, we obtain the lower bound

$$
\begin{aligned}
r_1 &= \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}} \left(|p+h|_{\cdot,2}^2 \left(|p+h|_{\cdot,2} - |p|_{\cdot,2}\right) - |p|_{\cdot,2}\,\langle h,\, p+h\rangle_{\cdot,2}\right) \\
&\geq \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}} \left(\frac{|p+h|_{\cdot,2}^2}{|p|_{\cdot,2}}\,\langle p,\, h\rangle_{\cdot,2} - |p|_{\cdot,2}\,\langle h,\, p+h\rangle_{\cdot,2}\right) \quad \text{by Cauchy–Schwarz} \\
&= \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}^2} \left(\left(|p+h|_{\cdot,2}^2 - |p|_{\cdot,2}^2\right)\langle p,\, h\rangle_{\cdot,2} - |p|_{\cdot,2}^2\,|h|_{\cdot,2}^2\right) \\
&= \frac{\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}^2} \left(\left(2\,\langle p,\, h\rangle_{\cdot,2} + |h|_{\cdot,2}^2\right)\langle p,\, h\rangle_{\cdot,2} - |p|_{\cdot,2}^2\,|h|_{\cdot,2}^2\right) \\
&\geq \frac{-\beta}{|p+h|_{\cdot,2}^3\,|p|_{\cdot,2}^2} \left(|p|_{\cdot,2}^2\,|h|_{\cdot,2}^2 + |h|_{\cdot,2}^3\,|p|_{\cdot,2}\right) \geq -\frac{|h|_{\cdot,2}^2}{\beta^2} - \frac{|h|_{\cdot,2}^3}{\beta\,|p|_{\cdot,2}}.
\end{aligned}
$$

This implies $\|p\,r_1\|_{L^2(\Omega)}^2 \leq 2\,\beta^{-4}\,(\|p\|_{L^{6,2}(\Omega)}^2\,\|h\|_{L^{6,2}(\Omega)}^4 + \beta^2\|h\|_{L^{6,2}(\Omega)}^6)$.

Second part. On $\mathcal{N}_2$ we have

$$
0 \leq 1 - \frac{\beta}{|p+h|_{\cdot,2}} = \frac{|p+h|_{\cdot,2} - \beta}{|p+h|_{\cdot,2}} \leq \frac{|h|_{\cdot,2}}{\beta} \quad \text{and} \quad \beta\,\frac{|\langle h,\, p+h\rangle_{\cdot,2}|}{|p+h|_{\cdot,2}^3} \leq \frac{|h|_{\cdot,2}}{\beta},
$$

and hence $|r_2| \leq 2\,\beta^{-1}\,|h|_{\cdot,2}$. Due to $\beta > |p|_{\cdot,2} > \beta - |h|_{\cdot,2}$ we have $\mu(\mathcal{N}_2) \to 0$ as $\|h\| \to 0$. This implies $\|p\,r_2\|_{L^2(\Omega)}^2 \leq 4\,\beta^{-2}\,\|p\|_{L^{6,2}(\Omega)}^2\,\|h\|_{L^{6,2}(\Omega)}^2\,\|\chi_{\mathcal{N}_2}\|_{L^{6,2}(\Omega)}^2$.

Third part. On $\mathcal{N}_3$ we have

$$
0 \leq r_3 = 1 - \frac{\beta}{|p|_{\cdot,2}} = \frac{|p|_{\cdot,2} - \beta}{|p|_{\cdot,2}} \leq \frac{|h|_{\cdot,2}}{|p|_{\cdot,2}} \leq \frac{|h|_{\cdot,2}}{\beta}.
$$

Due to $\beta < |p|_{\cdot,2} \leq \beta + |h|_{\cdot,2}$ we have $\mu(\mathcal{N}_3) \to 0$ as $\|h\| \to 0$. This implies $\|p\,r_3\|_{L^2(\Omega)}^2 \leq \beta^{-2}\,\|p\|_{L^{6,2}(\Omega)}^2\,\|h\|_{L^{6,2}(\Omega)}^2\,\|\chi_{\mathcal{N}_3}\|_{L^{6,2}(\Omega)}^2$.

Fourth part. We have

$$
|r_4| = \chi_{p+h}\,\left|\frac{\beta\,\langle h,\, p+h\rangle_{\cdot,2}}{|p+h|_{\cdot,2}^3}\right| \leq \beta^{-1}\,|h|_{\cdot,2}.
$$

This implies $\|h\,r_4\|_{L^2(\Omega)}^2 \leq \beta^{-2}\,\|h\|_{L^{4,2}(\Omega)}^4$.

Combining the estimates. Due to the above estimates we obtain

$$
\begin{aligned}
\|r(h)\|_{L^2(\Omega)} &\leq \|p\,r_1\|_{L^2(\Omega)} + \|p\,r_2\|_{L^2(\Omega)} + \|p\,r_3\|_{L^2(\Omega)} + \|h\,r_4\|_{L^2(\Omega)} \\
&= o\left(\|h\|_{L^{6,2}(\Omega)}\right) \quad \text{as } \|h\|_{L^{6,2}(\Omega)} \to 0,
\end{aligned}
$$

which proves that G is Newton differentiable with G' as a derivative. ∎

Due to $F = I - \alpha^{-1} G \circ P$ and the assumption (1.5) that the affine operator P maps $L^2(\Omega)$ continuously to $L^{6,2}(\Omega)$, we obtain the Newton differentiability of F.

COROLLARY 3.3. *The function $F : L^2(\Omega) \to L^2(\Omega)$ is Newton differentiable and a generalized derivative is given by*

$$
F'(u) = I + \frac{1}{\alpha}(G' \circ P)(u) \circ \mathcal{S}^\star \mathcal{S}.
$$

This corollary shows that $F'(u)$ is the sum of the identity operator and the composition of two operators. To show the bounded invertibility of $F'(u)$, we investigate its spectrum. To this end, we cite the following lemma from [18, Prop. 1] and derive a corollary. Recall that an operator A on a Hilbert space is called positive if $\langle Av,\, v\rangle \geq 0$ holds for all v.

LEMMA 3.4. *Let A and B be operators on a Hilbert space, let B be positive, and denote by $\sqrt{B}$ the positive square root of B. Then*

$$\sigma(AB) = \sigma(BA) = \sigma(\sqrt{B}A\sqrt{B}).$$

COROLLARY 3.5. *Let A and B both be positive operators on a Hilbert space with one of them self-adjoint. Then $\sigma(AB) \subset [0,\infty)$.*

Proof. We first assume that B is self-adjoint. Therefore, $\sqrt{B}$ is self-adjoint as well. This implies that $\sqrt{B}A\sqrt{B} = (\sqrt{B})^\star A\sqrt{B}$ is positive, and thus $\sigma(AB) = \sigma(\sqrt{B}A\sqrt{B}) \subset [0,\infty)$. The result for self-adjoint A follows analogously. ∎

We shall apply this corollary to infer the invertibility of $F'(u)$, using the settings

$$A = G'(P(u)), \quad B = \mathcal{S}^\star\mathcal{S}, \tag{3.2}$$

both considered as operators from $L^2(\Omega)$ into itself. Clearly, B is positive and self-adjoint. The positivity of A is shown as part of the proof of the following lemma.

LEMMA 3.6. *For every $u \in L^2(\Omega)$, the bounded linear mapping $F'(u) : L^2(\Omega) \to L^2(\Omega)$ is invertible and $\|F'(u)^{-1}\| \leq 1$ holds.*

Proof. In order to get a better understanding of the structure of $F'(u)$ we define for $p = Pu$ the linear maps

$$\chi : L^2(\Omega) \to L^2(\Omega), \qquad (\chi v)(x_1,\cdot) = \begin{cases} 0 & \text{if } |p|_{x_1,2} < \beta, \\ v(x_1,\cdot) & \text{otherwise,} \end{cases}$$

$$Q : L^2(\Omega) \to L^2(\Omega), \qquad (Qv)(x_1,\cdot) = \begin{cases} 0 & \text{if } |p|_{x_1,2} < \beta, \\ \beta\, |p|_{x_1,2}^{-1}\, v(x_1,\cdot) & \text{otherwise,} \end{cases}$$

$$R : L^2(\Omega) \to L^2(\Omega), \qquad (Rv)(x_1,\cdot) = \begin{cases} 0 & \text{if } |p|_{x_1,2} < \beta, \\ p(x_1,\cdot)\, |p|_{x_1,2}^{-2}\, \langle v,\, p\rangle_{x_1,2} & \text{otherwise.} \end{cases}$$

Note that R is the $L^2(\Omega)$-orthogonal projector onto the subspace spanned by the stripes of p, i.e., onto the subspace

$$\big\{u \in L^2(\Omega) : \exists \mu \in L^2(\Omega_1) : u(x_1,\cdot) = 0 \text{ if } |p|_{x_1,2} < \beta \\ \text{and } u(x_1,\cdot) = \mu(x_1)\, |p|_{x_1,2}^{-1}\, p(x_1,\cdot)\big\}.$$

With these definitions we have

$$A = G'(P(u)) = \chi - Q + QR.$$

In order to prove the invertibility of $F'(u)$ we perform two steps.

First step. We show that A is a positive operator. Since R is a projector, $I - R$ is a projector as well. In particular, this implies $\|R - I\| \leq 1$. Due to $\|Q\| \leq 1$ we have $\|QR - Q\| \leq 1$. Therefore $I + QR - Q$ is positive. Using $\chi Q = Q\chi = Q$, $R\chi = R$, and the self-adjointness of χ, we see that $\chi + QR - Q = \chi\,(I + QR - Q)\,\chi$ is positive as well.

Second step. We apply Corollary 3.5 to infer the invertibility of $F'(u)$ and the bound for $F'(u)^{-1}$. Applying Corollary 3.5 with the setting (3.2) yields $\sigma(G'(P(u)) \circ$

$\mathcal{S}^\star\mathcal{S}) \subset [0,\infty)$. By Corollary 3.3 we infer $\sigma(F'(u)) \subset [1,\infty)$. This implies the invertibility of $F'(u)$ as well as $\|F'(u)^{-1}\| \leq 1$ independent of $u \in L^2(\Omega)$. ∎

Using classical arguments (see, e.g., [7, Thm. 3.4]) the Newton differentiability of F and the uniform boundedness of $F'(u)^{-1}$ imply the following result.

THEOREM 3.7. *The semismooth Newton method*

$$u_{k+1} = u_k - F'(u_k)^{-1}F(u_k)$$

converges locally to the unique solution of (3.1) *with a superlinear convergence rate.*

We remark that it is possible to include in (3.1) constraints of the form

$$|u|_{\cdot,2} \leq u_b, \tag{3.3}$$

where $u_b \in L^1(\Omega_1)$. This constraint bounds the norm of u on each stripe. For that case, the optimality system (3.1) has to be replaced by

$$\bar{u} = \min\left(u_b,\ \max\left(0, \frac{|\bar{p}|_{\cdot,2} - \beta}{\alpha}\right)\right)\frac{\bar{p}}{|\bar{p}|_{\cdot,2}}. \tag{3.4}$$

The Newton differentiability and the bounded invertibility of the generalized derivative of (3.4) can be shown as for the unconstrained case. We also mention that constraints of the type (3.3) are the weakest constraints which ensure the solvability of $(\mathcal{P}_1)$ without additional L^2-regularization, i.e., for $\alpha = 0$: note that the solution space $L^{1,2}(\Omega)$ is not reflexive, but the set of admissible controls satisfying (3.3) is weakly compact, and the existence of a solution in $L^{1,2}(\Omega)$ follows by standard arguments. The inclusion of pointwise control bounds into an optimality system of the form (3.1) is not possible because the additional pointwise complementarity conditions interfere with the stripewise structure of (3.1).

4. Semismooth Newton method II: The constrained case. In this section, we propose an alternative approach to solving $(\mathcal{P}_1)$, which is based on a different reformulation of the optimality conditions. This form allows for the incorporation of pointwise bound constraints since it uses u and λ, while (3.1) is a reduced form that involves only the adjoint state. While being more flexible, a drawback of the approach in this section is that it only applies to the discretized, finite-dimensional problem and permits only a limited analysis. Since the finite-dimensional min- and max-functions used in the reformulation are semismooth [7, 16], we can compute a slant derivative. However, the resulting semismooth Newton iteration is well defined only under certain conditions (see Lemma 4.2). In numerical practice, we modify the semismooth Newton step to enforce these conditions and observe that these modifications are necessary only in the initial steps of the iteration. The equivalent form for the optimality system is given next.

LEMMA 4.1. *The function $\bar{u} \in L^2(\Omega)$ solves $(\mathcal{P}_1)$ if and only if there exist $\bar{\lambda}, \bar{\mu} \in L^2(\Omega)$ such that the following equations are satisfied a.e. in Ω:*

$$-\bar{p} + \beta\bar{\lambda} + \alpha\bar{u} + \bar{\mu} = 0, \tag{4.1a}$$

$$\max(1, |\bar{\lambda} + c_1\,\bar{u}|_{\cdot,2})\,\bar{\lambda} - (\bar{\lambda} + c_1\,\bar{u}) = 0, \tag{4.1b}$$

$$\bar{\mu} - \max(0, \bar{\mu} + c_2\,(\bar{u} - u_b)) - \min(0, \bar{\mu} + c_2\,(\bar{u} - u_a)) = 0, \tag{4.1c}$$

where $c_1, c_2 > 0$ are arbitrary and $\bar{p} = \mathcal{S}^\star(y_d - \mathcal{S}\bar{u})$.

Proof. The reformulation of optimality conditions originating from pointwise control constraints using complementarity functions is well known (see, e.g., [16, 21, 26]). Thus, we skip the proof of the equivalence between (4.1c) and (2.5) and restrict ourselves to showing the equivalence between (4.1b) and $\bar\lambda \in \partial\|\cdot\|_{1,2}(\bar u)$, the latter meaning $|\bar\lambda|_{\cdot,2} \le 1$ and

$$\bar\lambda = \bar u/|\bar u|_{\cdot,2} \qquad \text{on } \{x \in \Omega : |\bar\lambda|_{\cdot,2} = 1\}, \tag{4.2a}$$

$$\bar u = 0 \qquad \text{on } \{x \in \Omega : |\bar\lambda|_{\cdot,2} < 1\}. \tag{4.2b}$$

First, we show that (4.2) implies (4.1b). We first consider the case $x_1 \in \Omega_1$ with $|\bar\lambda + c_1\,\bar u|_{x_1,2} \ge 1$, and thus we are necessarily in case (4.2a). Hence, $\max(1, |\bar\lambda + c_1\,\bar u|_{x_1,2})\,\bar\lambda - (\bar\lambda + c_1\,\bar u) = (|\bar u|_{x_1,2}^{-1} + c_1)\bar u - (|\bar u|_{x_1,2}^{-1} + c_1)\bar u = 0$. If, on the other hand, $|\bar\lambda + c_1\,\bar u|_{x_1,2} < 1$, then we are necessarily in case (4.2b) and (4.1b) follows immediately from $\bar u = 0$.

Conversely, let us assume that (4.1b) holds. We first study $x_1 \in \Omega_1$ with $|\bar\lambda + c_1\,\bar u|_{x_1,2} \ge 1$, in which case (4.1b) implies that $|\bar\lambda|_{x_1,2} = 1$ and that $\bar\lambda = \alpha\,\bar u$ for some α. Plugging this expression for $\bar\lambda$ into (4.1b) shows that $\alpha = |\bar u|_{x_1,2}^{-1}$, yielding (4.2a). If on the other hand $|\bar\lambda + c_1\,\bar u|_{x_1,2} < 1$, we obtain from (4.1b) that $\bar u = 0$ on $\{x_1\} \times \Omega_2(x_1)$, and thus also $|\bar\lambda|_{x_1,2} < 1$, resulting in (4.2b). ∎

The form (4.1) of the optimality conditions motivates the application of a Newton algorithm as in section 3. A formal computation of a semismooth Newton step for u, p, λ, and μ in (4.1) results in the following Newton iteration: Given current iterates $(u^k, p^k, \mu^k, \lambda^k)$, compute the new iterates $(u^{k+1}, p^{k+1}, \mu^{k+1}, \lambda^{k+1})$ by solving the system

$$p^{k+1} - \mathcal{S}^\star(y_d - \mathcal{S}u^{k+1}) = 0 \qquad \text{on } \Omega, \tag{4.3a}$$

$$-p^{k+1} + \beta\,\lambda^{k+1} + \alpha\,u^{k+1} + \mu^{k+1} = 0 \qquad \text{on } \Omega, \tag{4.3b}$$

$$u^{k+1} = 0 \qquad \text{on } \mathcal{B}^k, \tag{4.3c}$$

$$\big(I - M^k\big)\,\lambda^{k+1} - c_1 M^k u^{k+1} = F^k\left(\lambda^k + c_1\,u^k\right) \text{ on } \mathcal{C}^k := \Omega \setminus \mathcal{B}^k, \tag{4.3d}$$

$$u^{k+1} = u_a \qquad \text{on } \mathcal{A}_a^k, \tag{4.3e}$$

$$u^{k+1} = u_b \qquad \text{on } \mathcal{A}_b^k, \tag{4.3f}$$

$$\mu^{k+1} = 0 \qquad \text{on } \Omega \setminus (\mathcal{A}_a^k \cup \mathcal{A}_b^k). \tag{4.3g}$$

Above,

$$\begin{aligned}\mathcal{B}^k &= \{(x_1, x_2) \in \Omega : |\lambda^k + c_1\,u^k|_{x_1,2} \le 1\},\\ \mathcal{A}_a^k &= \{x \in \Omega : \mu^k + c_2\,(u^k - u_a) < 0\},\\ \mathcal{A}_b^k &= \{x \in \Omega : \mu^k + c_2\,(u^k - u_b) > 0\},\end{aligned} \tag{4.3h}$$

I denotes the identity, and the linear operators F^k, M^k are given by

$$\begin{aligned} F^k &: v(x_1, x_2) \mapsto \langle \gamma^k,\, v\rangle_{x_1,2} |\gamma^k|_{x_1,2}^{-1} \lambda^k(x_1, x_2),\\ M^k &: v(x_1, x_2) \mapsto |\gamma^k|_{x_1,2}^{-1}\big(v(x_1, x_2) - F^k v(x_1, x_2)\big),\end{aligned} \tag{4.4}$$

where $\gamma^k := \lambda^k + c_1\,u^k$. Next, we study various aspects of (4.3), which motivate modifications that lead to an inexact semismooth Newton method that converges globally in numerical practice. First, we restrict ourselves to the case without control constraints, which amounts to $\bar\mu = 0$ in (4.1a) and omits (4.1c). We also assume that the problem has been discretized by finite differences or nodal finite elements in a way that preserves the splitting of $\mathbb{R}^N$ into $\mathbb{R}^n$ and $\mathbb{R}^{N-n}$. The discrete approximation or coefficient vectors are denoted with the subscript "h." For simplicity, we assume that these vectors are of size $\tilde{N} = N_1 N_2$, where N_1 and N_2 correspond to a tensorial discretization of the Ω_1 and the $\Omega_2(x_1)$ direction, respectively. The individual nodes of Ω_1 are denoted by x_1^j and we follow a discretize-then-optimize approach. The matrix $\mathcal{S}_h \in \mathbb{R}^{\tilde{N}\times\tilde{N}}$ is the discretization of the operator $\mathcal{S}$. For a vector $v_h \in \mathbb{R}^{\tilde{N}}$, we denote by $v_{h_j} \in \mathbb{R}^{N_2}$ the components of v_h corresponding to $\{x_1^j\} \times \Omega_2(x_1^j)$, such that

$$\langle v_h,\, w_h\rangle_j := v_{h_j}^\top N_{h_j} w_{h_j} \in \mathbb{R}$$

corresponds to the discretization of the inner product $\langle\cdot,\,\cdot\rangle_{x_1,2}$. Here, $N_{h_j} \in \mathbb{R}^{N_2\times N_2}$ is the symmetric mass matrix for the $\Omega_2(x_1)$ direction, and the overall L^2-mass matrix $N_h \in \mathbb{R}^{\tilde{N}\times\tilde{N}}$ acts on vectors u_h, w_h by $w_h^\top N_h u_h = \sum_{j=1}^{N_1} n_j\, w_{h_j}^\top N_{h_j} u_{h_j}$ with weights $n_j > 0$. The mass matrix in H is denoted by $L_h \in \mathbb{R}^{\tilde{N}\times\tilde{N}}$. Note that for a finite difference approximation, these mass matrices are diagonal. The norm corresponding to $\langle\cdot,\,\cdot\rangle_j$ is defined by $|v_h|_j := \sqrt{\langle v_h,\, v_h\rangle_j}$. Then, with $p_h = \mathcal{S}_h^\top(y_{d,h} - \mathcal{S}_h u_h)$ we obtain the discrete optimality condition

$$\mathcal{S}_h^\top L_h \mathcal{S}_h u_h + \beta N_h \lambda_h + \alpha N_h u_h - \mathcal{S}_h^\top L_h y_{d,h} = 0, \tag{4.5a}$$

$$\max(1, |\lambda_h + c_1\, u_h|_j)\lambda_{h_j} - (\lambda_{h_j} + c_1\, u_{h_j}) = 0 \qquad \text{for } j = 1,\ldots,N_1. \tag{4.5b}$$

Note that $N_h\lambda_h \in \partial(\sum_{j=1}^{N_1} n_j(u_{h_j}^\top N_j u_{h_j})^{1/2})$ holds. Computing a semismooth Newton step (see [16, 27]) for u_h, λ_h in (4.5) results in the following iteration: Given u_h^k, λ_h^k, compute the new iterates $u_h^{k+1}, \lambda_h^{k+1}$ by solving

$$-\mathcal{S}_h^\top L_h(y_{d,h} - \mathcal{S}_h u_h^{k+1}) + \beta N_h \lambda_h^{k+1} + \alpha N_h u_h^{k+1} = 0, \tag{4.6a}$$

$$u_{h_j}^{k+1} = 0 \qquad \text{for } j \in \mathcal{B}_h^k, \tag{4.6b}$$

$$\left(I_{h_j} - M_{h_j}^k\right)\lambda_{h_j}^{k+1} - c_1 M_{h_j}^k u_{h_j}^{k+1} = g_{h_j}^k \qquad \text{for } j \in \mathcal{C}_h^k \tag{4.6c}$$

with $\mathcal{B}_h^k = \{j : |\lambda_h^k + c_1\, u_h^k|_j \le 1 \text{ for } j = 1,\ldots,N_1\}$, $\mathcal{C}_h^k := \{1,\ldots,N_1\} \setminus \mathcal{B}_h^k$, and

$$F_{h_j}^k = \frac{1}{|\gamma_h^k|_j}\lambda_{h_j}^k(\gamma_{h_j}^k)^\top N_{h_j}, \tag{4.7a}$$

$$M_{h_j}^k = \frac{1}{|\gamma_h^k|_j}\left(I_{h_j} - F_{h_j}^k\right), \tag{4.7b}$$

$$g_{h_j}^k = \frac{1}{|\gamma_h^k|_j} F_{h_j}^k \gamma_{h_j}^k = \lambda_{h_j}^k, \tag{4.7c}$$

where $\gamma_h^k := \lambda_h^k + c_1\, u_h^k$. We observe that (4.6) are the discretized form of (4.3b)–(4.3d) for $\mu^{k+1} = 0$ and after elimination of p^{k+1} using (4.3a). The following result concerns the existence of a solution for the discrete Newton step (4.6).

LEMMA 4.2. *We set* $c_1 := \alpha\,\beta^{-1}$. *Provided that for all* $j \in \mathcal{C}_h^k$ *holds* $|\lambda_h^k|_j \le 1$ *and that*

$$\langle \gamma_h^k,\, \lambda_h^k\rangle_j > \max\left(0,\ |\gamma_h^k|_j\,|\lambda_h^k|_j - 2\,|\gamma_h^k|_j^2 + 2\,|\gamma_h^k|_j\right), \tag{4.8}$$

the linear system (4.6) *admits a unique solution.*

Proof. A straightforward computation shows that for $j \in \mathcal{C}_h^k$,

$$(I_{h_j} - M_{h_j}^k)^{-1} = \frac{|\gamma_h^k|_j}{|\gamma_h^k|_j - 1} \left(I_{h_j} - \frac{1}{|\gamma_h^k|_j^2 - |\gamma_h^k|_j + \langle \gamma_h^k, \lambda_h^k \rangle_j} \lambda_{h_j}^k (\gamma_{h_j}^k)^\top N_{h_j} \right).$$

Since $|\gamma_h^k|_j > 1$ and $\langle \gamma_h^k, \lambda_h^k \rangle_j \geq 0$, the matrix $(I_{h_j} - M_{h_j}^k)^{-1}$ is well defined. From (4.6c) follows

$$\lambda_{h_j}^{k+1} = \alpha\, \beta^{-1} \left((I_{h_j} - M_{h_j}^k)^{-1} - I_h \right) u_{h_j}^{k+1} + (I_{h_j} - M_{h_j}^k)^{-1} g_{h_j}^k,$$

where the fact that $(I_{h_j} - M_{h_j}^k)^{-1} M_{h_j}^k = (I_{h_j} - M_{h_j}^k)^{-1} - I_h$ is used. Plugging this expression for $\lambda_{h_j}^{k+1}$ into (4.6a) yields

$$\mathcal{S}_h^\top L_h \mathcal{S}_h u_h^{k+1} + \alpha \sum_{j \in \mathcal{C}_h^k} n_j N_{h_j} (I_{h_j} - M_{h_j}^k)^{-1} u_{h_j}^{k+1} = \tilde{g}_h^k \tag{4.9}$$

with $\tilde{g}_h^k := \mathcal{S}_h^\top L_h y_{d,h} - \beta \sum_{j \in \mathcal{C}_h^k} n_j N_{h_j} (I_{h_j} - M_{h_j}^k)^{-1} g_{h_j}^k$. Since $\mathcal{S}_h^\top L_H \mathcal{S}_h$ is positive semidefinite, it is sufficient to show that the matrix $N_{h_j}(I_{h_j} - M_{h_j}^k)^{-1}$ is positive definite. For that purpose, we choose a discrete vector u_h with $u_{h_i} = 0$ for $i \in \mathcal{B}_h^k$ and a $j \in \mathcal{C}_h^k$ and obtain

$$\begin{aligned}
u_{h_j}^\top N_{h_j} ((I_{h_j} - M_{h_j}^k)^{-1} u_{h_j} &= \frac{|\gamma_h^k|_j}{|\gamma_h^k|_j - 1} \left(|u_h|_j^2 - \frac{\langle u_h, \lambda_h^k \rangle_j \langle u_h, \gamma_h^k \rangle_j}{|\gamma_h^k|_j^2 - |\gamma_h^k|_j + \langle \gamma_h^k, \lambda_h^k \rangle_j} \right) \\
&\geq \frac{|\gamma_h^k|_j}{|\gamma_h^k|_j - 1} \left(|u_h|_j^2 - \frac{|u_h|_j^2 \left(\langle \gamma_h^k, \lambda_h^k \rangle_j + |\lambda_h^k|_j |\gamma_h^k|_j \right)}{2 \left(|\gamma_h^k|_j^2 - |\gamma_h^k|_j + \gamma_h^k \lambda_{h_j}^k \right)} \right) \\
&= \frac{|u_h|_j^2 |\gamma_h^k|_j}{|\gamma_h^k|_j - 1} \left(\frac{2|\gamma_h^k|_j^2 - 2|\gamma_h^k|_j + \langle \gamma_h^k, \lambda_h^k \rangle_j - |\lambda_h^k|_j |\gamma_h^k|_j}{2(|\gamma_h^k|_j^2 - |\gamma_h^k|_j + \langle \gamma_h^k, \lambda_h^k \rangle_j)} \right) \\
&> 0.
\end{aligned}$$

Above, (4.8) has been used in the last estimate. This shows that the system matrix in (4.9) is positive definite. Since $u_{h_j}^{k+1} = 0$ for $j \in \mathcal{B}_h^k$, there exists a unique u_h^{k+1} that solves (4.9). The corresponding λ_h^{k+1} is found from (4.6). ∎

We continue with a series of remarks.

On the conditions of Lemma 4.2. The conditions from Lemma 4.2 are satisfied at the solution $(\bar{u}_h, \bar{\lambda}_h)$ of the discretized version of $(\mathcal{P}_1)$. First, λ_h satisfies $|\bar{\lambda}_h|_j \leq 1$ for all $j \in \{1, \ldots, N_1\}$. Additionally, $\bar{\lambda}_{h_j}$ and $\bar{\gamma}_{h_j}$ are parallel vectors, as can be seen from (4.5b). Thus, $\langle \bar{\gamma}_h, \bar{\lambda}_h \rangle_j = |\bar{\gamma}_h|_j |\bar{\lambda}_h|_j$ and, since $|\bar{\gamma}_h^k|_j > 1$, the condition (4.8) holds. However, the assumptions from Lemma 4.2 are, in general, unlikely to hold for all $j \in \mathcal{C}_h^k$. To guarantee stable convergence, we modify the linear mappings $F_{h_j}^k$ and $M_{h_j}^k$, as defined in (4.7), in the semismooth Newton step by replacing them by the "damped" forms

$$\begin{aligned}
\widetilde{F}_{h_j}^k &= \min(1, |\lambda_h^k|_j^{-1}) F_{h_j}^k, \\
\widetilde{M}_{h_j}^k &= \min \left(1, \frac{2\left(|\gamma_h^k|_j^2 - |\gamma_h^k|_j + \langle \gamma_h^k, \lambda_h^k \rangle_j \right)}{|\lambda_h^k|_j |\gamma_h^k|_j + \langle \lambda_h^k, \gamma_h^k \rangle_j} \right) M_{h_j}^k.
\end{aligned} \tag{4.10}$$

Note that $\widetilde{F}^k_{h_j}$ only differs from $F^k_{h_j}$ for $j \in \mathcal{C}^k_{h_j}$, for which $|\lambda^k_h|_j \leq 1$ is not satisfied. Similarly, the second modification in (4.10) comes into play only if the angle between $\lambda^k_{h_j}$ and $\gamma^k_{h_j}$ is large. Thus, the modifications vanish as the iterates converge and we still expect local fast convergence of the semismooth Newton iterates. This is confirmed by our numerical experiments. With the modified matrices $\widetilde{F}^k_{h_j}, \widetilde{M}^k_{h_j}$, one can follow the proof of Lemma 4.2 and show that the Newton iteration is well defined. The proof of Lemma 4.2 also shows that the reduced system (4.9), which involves u_h only, is positive definite. Similar modifications of semismooth Newton derivatives are employed in [19] for friction problems, which can be written as a nonsmooth optimization problem similar to $(\mathcal{P}_1)$. In practice, the above modifications turn out to be important for reliable convergence of the algorithm.

Adding control constraints. In Lemma 4.2, the case without bound constraints on the control is discussed. To extend the directional sparsity algorithm to the case with control constraints, we introduce the discrete active sets by $\mathcal{A}^k_{h,a}$ and $\mathcal{A}^k_{h,b}$; compare with (4.3h). Since we assume $u_a < 0 < u_b$, $\mathcal{B}^k_h$ does not overlap with the active sets for the control constraints at the solution. However, during the iteration such an overlap is possible. This can lead to an unsolvable Newton system, which is a consequence of the independent treatment of the bound constraints and of the nondifferentiable term (see [17] for a similar problem). To avoid this problem, we compute $\mathcal{A}^k_{h,a}$ and $\mathcal{A}^k_{h,b}$ as subsets of $\mathcal{C}^k_h$ in our numerical implementation.

Alternatively, one can employ a quadratic penalization to enforce the pointwise control constraints approximatively using the terms

$$\frac{1}{2\rho}\|\max(0, u - u_b)\|^2_{L^2(\Omega)} + \frac{1}{2\rho}\|\min(0, u - u_a)\|^2_{L^2(\Omega)}. \tag{4.11}$$

This also has the advantage that the Newton linearization can be reduced to a system in the control variable u only, as in the case without bound constraints.

Comparison of Newton methods for different reformulations. Using (4.1a) to eliminate $\bar{\lambda}$ in (4.1b), neglecting the bound constraints, and choosing $c_1 = \alpha\beta^{-1}$ results in

$$\max(1, \beta^{-1}|\bar{p}|_{\cdot,2})(\bar{p} - \alpha\bar{u}) - \bar{p} = 0. \tag{4.12}$$

As (3.1), this formulation only involves $\bar{u}$ and $\bar{p}$. However, the two formulations lead to different semismooth Newton methods that, in practice, also yield a different convergence behavior (see section 5). Yet another formulation is obtained if (4.12) divided by $\max(1, |\bar{p}|_{\cdot,2})$ is used as a starting point for a Newton method. However, we found (4.1) and the reduced form (4.12) and (3.1) to be preferable in numerical practice.

5. Numerical results.

In this section we analyze the structure of optimal controls arising from the directional sparsity formulation $(\mathcal{P}_1)$. We also study the performance of the algorithms proposed in sections 3 and 4. We start with a presentation of the test examples.

5.1. Test examples and settings.

Example 5.1. We consider an elliptic state equation as in Example 1.1 with $\Omega = (0,1) \times (0,1)$, i.e., $N = 2$ and $n = 1$. For the discretization we use an equidistant mesh and approximate the solution operator $\mathcal{S}$ of the Poisson equation with the standard five-point finite difference stencil. The remaining problem data are $y_d =$

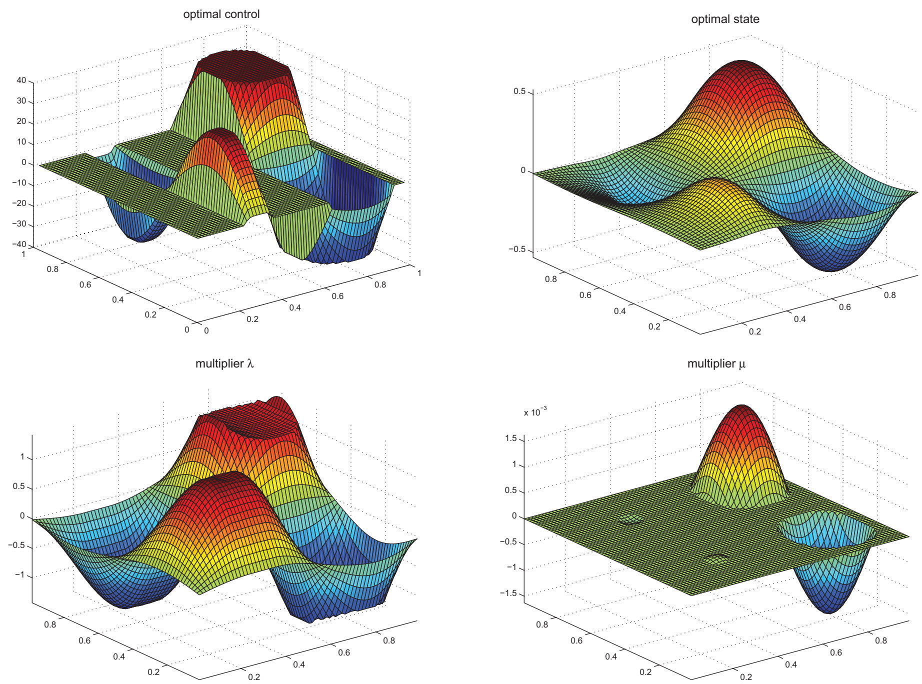

FIG. 5.1. *Solution of Example* 5.1 *with* $\alpha = 10^{-6}$, $\beta = 10^{-3}$ *on a* 64×64 *mesh: optimal control (upper left), optimal state (upper right), dual solutions/Lagrange multipliers* λ *(lower left), and* μ *(lower right).*

$\sin(2\pi x_1)\sin(2\pi x_2)\exp(2x_1)/6$ and, unless stated otherwise, the control constraints $-u_a = u_b = 35$ are used. For $\alpha = 10^{-6}$ and $\beta = 10^{-3}$, the optimal control and state, as well as the dual solutions λ and μ (i.e., the Lagrange multipliers for $(\mathcal{P}_1)$), are shown in Figure 5.1. Moreover, Figure 5.2 depicts the optimal controls for various choices of α and β.

Example 5.2. This is an example with a parabolic state equation (see Example 1.2) on the space-time cylinder $\Omega_1 \times (0, T)$ with $\Omega_1 = (0, 1) \times (0, 1)$ and $T = 1$, i.e., $N = 3$ with $n = 2$ sparse spatial directions. Linear finite elements on a triangular mesh are used for the discretization in space and the implicit Euler method for the discretization in time. The thermal conductivity is $\kappa = 0.1$, the control bounds are $-u_a = u_b \equiv 7$, and the desired state is $y_d = \sin(2\pi x_1)\sin(2\pi x_2)\sin(\pi t)\exp(2x_1)/6$. The parameters α and β are specified when needed. Snapshots of the optimal control and the difference between optimal and desired states at four time instances are shown in Figure 5.3.

Example 5.3. Here we consider the parabolic case on a three-dimensional spatial domain (see Figure 5.4), that is, $N = 4$, and we have $n = 3$ sparse directions. Further, $T = 1$, $\kappa = 2$, and $H = L^2((0.875, 1) \times \Omega_2)$, i.e., only close to the final time the state is tracked toward the desired state, which is given by $y_d \equiv 1$ for $x_2 \geq 1.2$ and $y_d \equiv 0$ for $x_2 < 1.2$. The state satisfies homogeneous Dirichlet conditions for $x_2 < 0.2$ and homogeneous Neumann conditions for $x_2 \geq 0.2$, and no control bounds are employed in this example. The sparsity structure and the state at the final time for $\alpha = 2 \cdot 10^{-6}$ and $\beta = 0.02$ can be seen in Figure 5.4.

In our implementation we follow a discretize-then-optimize approach, that is, we first discretize the problem and then derive the optimality system to obtain a con-

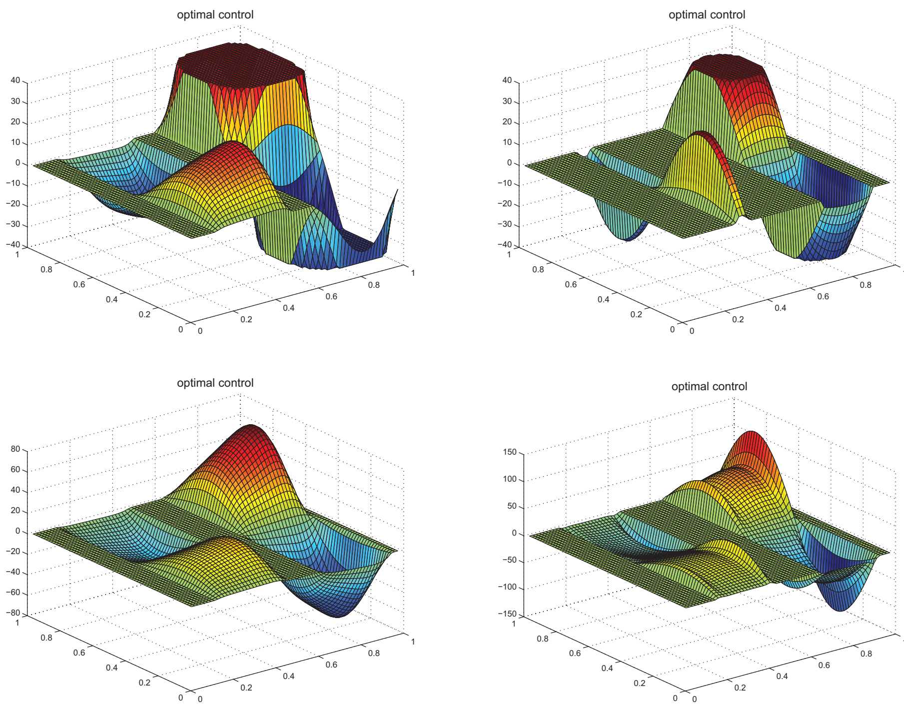

FIG. 5.2. *Optimal controls for Example* 5.1 *for different values of* α *and* β. *Upper row:* $\alpha = 10^{-6}$, *and* $\beta = 10^{-4}$ *(left) and* $\beta = 2 \cdot 10^{-3}$ *(right). The optimal control for* $\beta = 10^{-3}$ *can be seen in Figure* 5.1. *Lower row: Results without control constraints for* $\beta = 10^{-4}$, *and* $\alpha = 10^{-6}$ *(left) and* $\alpha = 10^{-8}$ *(right).*

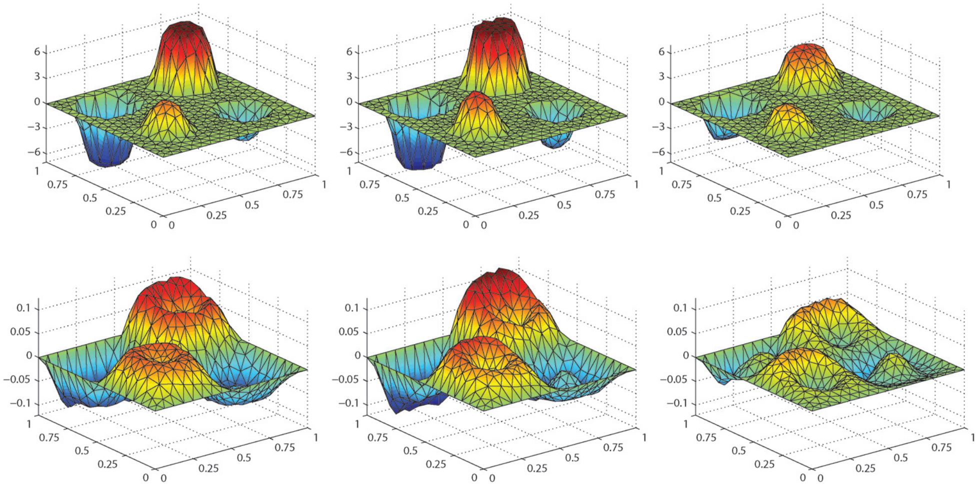

FIG. 5.3. *Optimal control* u *(top) and residual* $y_d - y$ *(bottom) for Example* 5.2 *with* $\alpha = 3 \cdots 10^{-4}$, $\beta = 5 \cdot 10^{-3}$ *at times* $t = 0.25,\ 0.5,\ 0.75$ *(from left to right).*

sistent discrete formulation. The linear systems for the elliptic problem are solved with a direct solver. The parabolic examples use the iterative generalized minimal residual method (GMRES) to solve the Newton linearizations. Since these linear

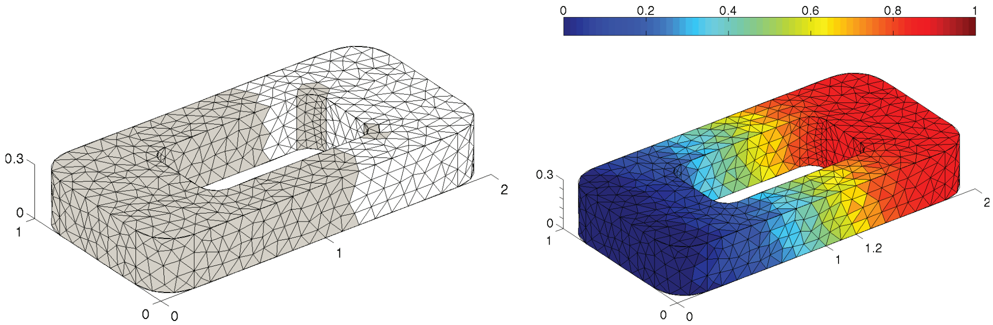

FIG. 5.4. *Example* 5.3: *Sparsity structure of optimal control (left; on gray elements the optimal control is identically zero) and temperature at final time* $t = 1$ *(right).*

TABLE 5.1

Results for Example 5.1 *for various* α *and* β *on* 64×64 *mesh with and without control bounds. The fourth and fifth columns show the sparsity of the optimal control (i.e., the percentage of* Ω *where the optimal control vanishes) and the* L^2*-norm of the difference between optimal and desired state, respectively. The number of iterations (#iter) needed by the method described in section* 4 *is shown in the sixth column, where the iteration is terminated when the nonlinear residual drops below* 10^{-8}. *For the unconstrained case we give in brackets the iteration count for the semismooth Newton method from section* 3. *The last column points to the figure showing the optimal control.*

β	α	u_a, u_b	Sparsity	$\|y - y_d\|_{L^2(\Omega)}$	#iter	Plot of $\bar{u}$ in Fig.
0	1e-6	yes	0%	5.00e-2	11	
2e-3	1e-6	yes	69%	1.42e-1	18	5.2 (upper right)
1e-3	5e-5	yes	45%	1.00e-1	7	
1e-3	1e-6	yes	55%	9.46e-2	14	5.1 (upper left)
1e-4	1e-6	yes	23%	5.22e-2	11	5.2 (upper left)
1e-4	1e-6	no	31%	1.33e-2	7(6)	5.2 (lower left)
1e-4	1e-8	no	43%	1.12e-2	16(15)	5.2 (lower right)

systems are similar to optimality systems from linear-quadratic optimal control problems, they can also be solved using advanced iterative algorithms such as multigrid methods [1, 3].

5.2. Qualitative study of solutions. First, we solve Example 5.1 for different values of the parameters α and β to study structural properties of the solutions. Figure 5.1 allows a visual verification of the complementarity conditions for u, λ, and μ. Note, in particular, that the multiplier λ satisfies $|\lambda|_{x_1,2} \leq 1$ for all $x_1 \in \Omega_1$ but does not satisfy $|\lambda| \leq 1$ in a pointwise sense. The results for Example 5.1 are summarized in Table 5.1 and can be seen in Figures 5.1 and 5.2. Also, for the parabolic problems (Examples 5.2 and 5.3) we study the influence of β on the solution. Results are shown in Figure 5.5 (sparsity patterns for Example 5.2 for different values of β) and in Figure 5.4. Recall that for time-dependent problems, our notion of sparsity implies that the optimal control vanishes for all $t \in [0, T]$ on parts of the spatial domain Ω_1. From the results we draw the following conclusions:

- As β increases, the optimal controls become sparser.
- As for problems without sparsity constraints, the parameter α controls the smoothness of the optimal control.
- Even solutions with significant sparsity are able to track the state close to the desired state. This is particularly true for problems without control constraints.

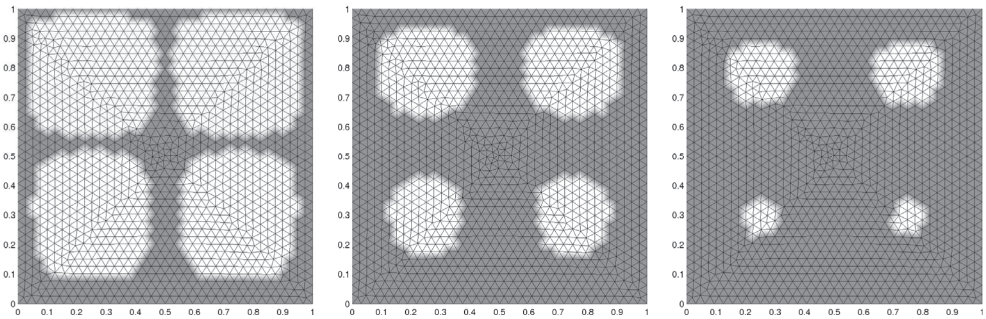

Fig. 5.5. *Mesh and sparsity structure of optimal controls for Example* 5.2 *using* 1365 *spatial degrees of freedom and* 64 *time steps for* $\alpha = 3 \cdot 10^{-4}$ *and* $\beta = 5 \cdot 10^{-4}$ *(left),* $\beta = 5 \cdot 10^{-3}$ *(middle),* $\beta = 2 \cdot 10^{-2}$ *(right). At dark gray node points, the optimal controls vanish for* all *time steps.*

Table 5.2

Example 5.2 *with* $\alpha = 10^{-4}$, $\beta = 10^{-3}$, *and* $\rho = 10^3$. *Shown are the number of nonlinear iterations #iter for different spatial resolution (#nodes denotes the number of spatial degrees of freedom of the triangulation) and numbers of time steps #tsteps. The nonlinear iteration is terminated as soon as the* L^2*-norm of the residual drops below* 10^{-4}. *With* $|\mathcal{B}|$ *we denote the number of points where the optimal control is zero for every time step and with* $|\mathcal{A}_a|$ *and* $|\mathcal{A}_a|$ *the number of points where the control constraints are active in the space-time domain.*

#nodes	#tsteps	#iter	$\lvert\mathcal{B}\rvert$	$\lvert\mathcal{A}_a\rvert$	$\lvert\mathcal{A}_b\rvert$
102	64	7	80	47	61
102	256	7	80	184	237
367	64	7	273	242	222
367	256	7	273	941	862
1365	64	8	992	864	869
1365	256	8	994	3368	3403

- The lower right plot in Figure 5.2 suggests that as $\alpha \to 0$, the optimal controls become concentrated on the boundary of the active region $\mathcal{B}$ in the absence of pointwise control constraints. A theoretical analysis of this behavior, which requires $\alpha = 0$ and, hence, a different choice of solution spaces (as, for instance, in [8]) might be worthwhile.

5.3. Performance of semismooth Newton methods. Both algorithms based on semismooth Newton methods prove to be very efficient to solve $(\mathcal{P}_1)$ for large ranges of the parameters α and β. This can be seen from the number of iterations to solve Example 5.1 as shown in Table 5.1. Note that the iteration number for the directional sparsity approach is not significantly larger than that for the solution of the problem with control constraints only (first row in Table 5.1). In the absence of pointwise control bounds, both algorithms performed equally well (see the last two rows in Table 5.1) and converged for all initializations we tested. We found that the modifications (4.10) of the Newton derivative are important for stable convergence of the method presented in section 4.

For the solution of the parabolic problems (Examples 5.2 and 5.3), a reduced space approach is used for the numerical solution of (4.3) (using the modification (4.10)), that is, y, p, and λ, are eliminated from the Newton step and the optimal control u is the only unknown. To allow for this reduction in Example 5.2, the control constraints are realized through the quadratic penalization (4.11) with penalization parameter $\rho = 10^3$. In Table 5.2, we show the number of semismooth Newton iterations required

Table 5.3
Example 5.1 with $\alpha = 10^{-6}$, $\beta = 10^{-4}$ on a 64×64-mesh, solved with the semismooth Newton method in section 3. Shown is the L^2-norm of the nonlinear residual res_k in the kth iteration.

k	0	1	2	3	4	5	6
res_k	3.72e3	2.18e1	8.28e0	3.53e0	1.04e0	6.30e-2	7.93e-13

to solve the problem for different spatial and temporal resolutions. The iteration is terminated as soon as the L^2-norm of the nonlinear residual drops below 10^{-4}. Note that the number of Newton iterations (and thus linear solves) remains approximately constant as the spatial and the temporal resolution increases.

Finally, in Table 5.3 we show the iteration behavior of the nonlinear residual during the solution of Example 5.1 with $\alpha = 10^{-6}$ and $\beta = 10^{-4}$. Note that the result shows superlinear convergence of the semismooth Newton method.

6. Conclusion and discussion. In this paper we considered a new class of nonsmooth control cost (regularization) terms for optimal control problems for PDEs. These terms promote sparsity of the optimal control in preselected directions of the underlying domain of definition. This behavior is confirmed by numerical examples. Two methods of semismooth Newton type were proposed for the solution of problems of this new class. In the absence of pointwise control bounds, a complete proof of local superlinear convergence was given for the first method. Both methods proved to be very efficient in practice.

Some questions are still open and provide opportunities for further research. A suitable formulation of the optimality system in the presence of pointwise control bounds which would allow a rigorous semismooth Newton analysis in function space is lacking. Related to this is the question of which reformulations of the optimality conditions lead to the most efficient and stable algorithm. Moreover, a priori error estimates, as given, for instance, in [5, 32] for related problem classes involving the "undirected" sparsity term $L^1(\Omega)$, are not yet available for the problems considered here. Finally, the extension to semilinear PDEs and in particular the study of second-order optimality conditions present another challenge.

REFERENCES

[1] S. S. Adavani and G. Biros, *Multigrid algorithms for inverse problems with linear parabolic PDE constraints*, SIAM J. Sci. Comput., 31 (2008), pp. 369–397.
[2] M. Bergounioux, K. Ito, and K. Kunisch, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37 (1999), pp. 1176–1194.
[3] A. Borzì and V. Schulz, *Multigrid methods for PDE optimization*, SIAM Rev., 51 (2009), p. 361.
[4] E. Candès and Y. Plan, *Near-ideal model selection by l_1 minimization*, Ann. Statist., 37 (2009), pp. 2145–2177.
[5] E. Casas, R. Herzog, and G. Wachsmuth, *Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional*, SIAM J. Optim., to appear.
[6] T. F. Chan and X.-C. Tai, *Identification of discontinuous coefficients in elliptic problems using total variation regularization*, SIAM J. Sci. Comput., 25 (2003), pp. 881–904.
[7] X. Chen, Z. Nashed, and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.
[8] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 243–266.
[9] A. Defant and K. Floret, *Tensor Norms and Operator Ideals*, North-Holland Math. Stud. 176, North-Holland, Amsterdam, 1993.
[10] N. Dinculeanu, *Vector Measures*, Deutscher Verlag der Wissenschaften, Berlin, 1966.

[11] D. Donoho, *Compressed sensing*, IEEE Trans. Inform. Theory, 52 (2006), pp. 1289–1306.

[12] D. Donoho, *For most large underdetermined systems of linear equations the minimal ℓ^1-norm solution is also the sparsest solution*, Comm. Pure Appl. Math., 59 (2006), pp. 797–829.

[13] M. Fornasier, R. Ramlau, and G. Teschke, *The application of joint sparsity and total variation minimization algorithms in a real-life art restoration problem*, Adv. Comput. Math., 31 (2009), pp. 301–329.

[14] M. Fornasier and H. Rauhut, *Recovery algorithms for vector-valued data with joint sparsity constraints*, SIAM J. Numer. Anal., 46 (2008), pp. 577–613.

[15] E. Haber, L. Horesh, and L. Tenorio, *Numerical methods for experimental design of large-scale linear ill-posed inverse problems*, Inverse Problems, 24 (2008), p. 055012.

[16] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888.

[17] M. Hintermüller, S. Volkwein, and F. Diwoky, *Fast solution techniques in constrained optimal boundary control of the semilinear heat equation*, Internat. Ser. Numer. Math., 155 (2007), pp. 119–147.

[18] M. Hladnik and M. Omladič, *Spectrum of the product of operators*, Proc. Amer. Math. Soc., 102 (1988), pp. 300–302.

[19] S. Hüeber, G. Stadler, and B. I. Wohlmuth, *A primal-dual active set algorithm for three-dimensional contact problems with Coulomb friction*, SIAM J. Sci. Comput., 30 (2008), pp. 572–596.

[20] A. D. Ioffe and V. M. Tichomirov, *Theorie der Extremalaufgaben*, VEB Deutscher Verlag der Wissenschaften, Berlin, 1979.

[21] K. Ito and K. Kunisch, *The primal-dual active set method for nonlinear optimal control problems with bilateral constraints*, SIAM J. Control Optim., 43 (2004), pp. 357–376.

[22] J. Liu, S. Ji, and J. Ye, *SLEP: Sparse Learning with Efficient Projections*, Arizona State University, Tempe, AZ, 2009.

[23] M. Nikolova, *Local strong homogeneity of a regularized estimator*, SIAM J. Appl. Math., 61 (2000), pp. 633–658.

[24] M. Nikolova, *Analysis of the recovery of edges in images and signals by minimizing nonconvex regularized least-squares*, Multiscale Model. Simul., 4 (2005), pp. 960–991.

[25] Z. Qin, K. Scheinberg, and D. Goldfarb, *Efficient Block-Coordinate Descent Algorithms for the Group Lasso*, Technical report 2806, Optimization Online, 2010.

[26] G. Stadler, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), pp. 159–181.

[27] D. Sun and J. Han, *Newton and quasi-Newton methods for a class of nonsmooth equations and related problems*, SIAM J. Optim., 7 (1997), pp. 463–480.

[28] F. Tröltzsch, *Optimal Control of Partial Differential Equations: Theory, Methods and Applications*, Grad. Stud. Math. 112, American Mathematical Society, Providence, RI, 2010.

[29] J. A. Tropp and S. J. Wright, *Computational methods for sparse solution of linear inverse problems*, Proc. IEEE, 98 (2010), pp. 948–958.

[30] C. R. Vogel, *Computational Methods for Inverse Problems*, SIAM, Philadelphia, 2002.

[31] G. Vossen and H. Maurer, *On L^1-minimization in optimal control and applications to robotics*, Optimal Control Appl. Methods, 27 (2006), pp. 301–321.

[32] G. Wachsmuth and D. Wachsmuth, *Convergence and regularization results for optimal control problems with sparsity functional*, ESAIM Control Optim. Calc. Var., 17 (2011), pp. 858–886.

[33] G. Wachsmuth, *Elliptische Optimalsteuerungsprobleme unter Sparsity-Constraints*, Diploma thesis, Technische Universität Chemnitz, 2008; also available online from http://www.tu-chemnitz.de/mathematik/part_dgl/publications.php.

[34] S. J. Wright, R. D. Nowak, and M. A. T. Figueiredo, *Sparse reconstruction by separable approximation*, IEEE Trans. Signal Process., 57 (2009), pp. 2479–2493.