Nonconvex penalization of switching control of partial differential equations

Christian Clason[*] Karl Kunisch[†] Armin Rund[†]

May 31, 2016

A standard approach to treat constraints in nonlinear optimization is penalization, in particular using L^1-type norms. Applying this approach to the pointwise switching constraint $u_1(t)u_2(t) = 0$ leads to a nonsmooth and nonconvex infinite-dimensional minimization problem which is challenging both analytically and numerically. Adding H^1 regularization or restricting to a finite-dimensional control space allows showing existence of optimal controls. First order necessary optimality conditions are then derived using tools of nonsmooth analysis. Their solution can be computed using a combination of Moreau–Yosida regularization and a semismooth Newton method. Numerical examples illustrate the properties of this approach.

1 Introduction

Switching control refers to time-dependent optimal control problems with a vector-valued control of which at most one component should be active at any point in time. We focus here on optimal tracking control for a linear evolution equation $y_t + Au = Bu$ on $\Omega_T := (0, T] \times \Omega$ together with initial conditions $y(0) = y_0$ on Ω, where A is a linear elliptic operator defined on $\Omega \subset \mathbb{R}^n$ carrying suitable boundary conditions. The linear bounded control operator $B : L^2(0, T; \mathbb{R}^N) \to L^2(\Omega_T)$ is defined by, e.g.,

$$(Bu)(t, x) = \sum_{i=1}^{N} \chi_{\omega_i}(x)u_i(t),$$

where χ_{ω_i} is the characteristic function of the given control domain $\omega_i \subset \Omega$ of positive measure. Furthermore, let $\omega_{\text{obs}} \subset \Omega$ denote the observation domain and let $y^d \in L^2(0, T; L^2(\omega_{\text{obs}}))$ denote

[*]Faculty of Mathematics, University Duisburg-Essen, 45117 Essen, Germany (`christian.clason@uni-due.de`)

[†]Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, 8010 Graz, Austria (`armin.rund@uni-graz.at`, `karl.kunisch@uni-graz.at`).

the target. We consider the standard optimal control problem

$$\begin{cases} \min_{u \in L^2(0,T;\mathbb{R}^N)} \frac{1}{2}\|y - y^d\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \frac{\alpha}{2}\int_0^T |u(t)|_2^2\, dt, \\ \text{s.t.} \quad y_t + Ay = Bu, \quad y(0) = y_0, \end{cases}$$

where $|v|_2^2 = \sum_{j=1}^N v_j^2$ denotes the squared ℓ^2-norm on $\mathbb{R}^N$. To promote the switching structure of optimal controls, we suggest adding the penalty term

$$\beta \int_0^T \sum_{\substack{i,j=1 \\ i<j}}^N |u_i(t)u_j(t)|\, dt \tag{1.1}$$

with $\beta > 0$ to the objective, which can be interpreted as an L^1-penalization of the switching constraint $u_i(t)u_j(t) = 0$ for $i \neq j$ and $t \in [0,T]$. The combination of control cost and switching penalty is convex if and only if $\beta \leq \alpha$. The case $\beta = \alpha$ was investigated in [6]; the aim of this work is to treat the case $\beta > \alpha$, which allows choosing the switching penalty parameter independently of the control cost parameter. In this case, the convex analysis approach followed in [6] is not applicable. The main difficulty stems from the fact that the integrand $g : \mathbb{R}^2 \to \mathbb{R}$, $(u_1,u_2) \mapsto |u_1u_2|$, is not convex, and hence the integral functional $G : L^2(0,T;R^2) \to \mathbb{R}$, $u \mapsto \int_0^T |u_1(t)u_2(t)|\, dt$, is not weakly lower semicontinuous, which is an obstacle for proving existence. It is therefore necessary to enforce strong convergence of minimizing sequences, which is possible by either considering piecewise constant and hence finite-dimensional controls or by introducing an additional (small) $H^1(0,T)$ penalty. Our analysis will cover both approaches.

Besides the question of existence of optimal controls, their numerical computation is also challenging due to the nonconvexity of the problem. Here we proceed as follows: Using the calculus of Clarke's generalized derivative [2], we can derive first-order necessary optimality conditions. It then suffices to apply a Moreau–Yosida regularization only to the nonsmooth but convex term in the optimality conditions in order to apply a semismooth Newton method.

This is a natural continuation of our previous works [3, 6] on convex relaxation of the switching constraint. Let us briefly remark on further related literature. On switching control of ordinary and partial differential equations, there exists a large body of work; here we only mention [1, 11, 14] in the former context and [18, 12, 17, 8, 10, 13] in the latter. A related topic is the control of switched systems, where we refer to, e.g., [9, 15, 16].

This paper is organized as follows. The current section concludes with a short discussion of a scalar example illustrating the difficulties arising in the nonconvex setting. Section 2 is concerned with existence of optimal controls and their convergence as $\beta \to \infty$ to a "hard switching constrained" problem. Optimality conditions are then derived in section 3, where the question of exact penalization is addressed as well. Section 4 discusses the numerical solution of the optimality conditions using a semismooth Newton method. Finally, section 5 presents numerical examples illustrating the properties of the nonconvex penalty approach.

1.1 A scalar example

We close this section by considering a finite-dimensional example involving only two "controls" $u = (u_1,u_2) \in \mathbb{R}^2$ and a cost functional which reflects some of the properties of the infinite-dimensional problems that we are about to study. It illustrates the fact that (1.1) indeed promotes switching as well as the type of difficulties which may arise in dependence on the choice of the parameters α and β. Thus we consider

$$\min_{u\in\mathbb{R}^2} J(u) = \frac{1}{2}|u_1 - z_1|^2 + \frac{1}{2}|u_2 - z_2|^2 + \frac{\alpha}{2}|u_1|^2 + \frac{\alpha}{2}|u_2|^2 + \beta|u_1 \cdot u_2|, \tag{1.2}$$

where $z = (z_1, z_2) \geq 0$. Clearly, (1.2) admits a solution, and every such solution $\bar{u} = (\bar{u}_1, \bar{u}_2)$ satisfies $\bar{u} \geq 0$ as a consequence of $z \geq 0$. A standard computation involving case studies reveals that

(i) for the case $\beta > 1 + \alpha$:

$$\begin{array}{lll}
\text{if } z_2 < z_1: & \left(\dfrac{z_1}{1+\alpha}, 0\right) & \text{is a global minimizer,} \\
\text{if } z_1 < z_2: & \left(0, \dfrac{z_2}{1+\alpha}\right) & \text{is a global minimizer,} \\
\text{if } z_1 = z_2: & \left(\dfrac{z_1}{1+\alpha}, 0\right) \text{ and } \left(0, \dfrac{z_2}{1+\alpha}\right) & \text{are global minimizers,} \\
\text{if } \frac{1+\alpha}{\beta} z_1 \leq z_2 \leq \frac{\beta}{1+\alpha} z_1: & \left(\dfrac{z_1}{1+\alpha}, 0\right) \text{ and } \left(0, \dfrac{z_2}{1+\alpha}\right) & \text{are local minimizers}
\end{array}$$

(note that these cases are not mutually exclusive but consistent, since global minimizers are *a fortiori* also local minimizers);

(ii) for the case $0 < \beta < 1 + \alpha$:

$$\begin{array}{lll}
\text{if } z_2 \leq \frac{\beta}{1+\alpha} z_1: & \left(\dfrac{z_1}{1+\alpha}, 0\right) & \text{is the global minimizer,} \\
\text{if } z_1 \leq \frac{\beta}{1+\alpha} z_2: & \left(0, \dfrac{z_2}{1+\alpha}\right) & \text{is the global minimizer,} \\
\text{if } \frac{\beta}{1+\alpha} z_1 < z_2 < \frac{1+\alpha}{\beta} z_1: & \left(\dfrac{\beta z_2 - (1+\alpha) z_1}{\beta^2 - (1+\alpha)^2}, \dfrac{\beta z_1 - (1+\alpha) z_2}{\beta^2 - (1+\alpha)^2}\right) & \text{is the global minimizer.}
\end{array}$$

Thus, for $0 < \beta < 1 + \alpha$ the solution is switching except in the wedge defined in the last bullet where the global solution is non-switching. For the case $\beta > 1 + \alpha$, which is the focus in this paper, all solutions are switching, but in the wedge defined in the last bullet there are two local solutions, only one of which is global.

2 Existence

As mentioned, we need to restrict the set of feasible controls in order to obtain existence. We thus introduce the control space $\mathcal{U} \subset L^2(0, T; \mathbb{R}^N)$ and consider in the following two cases:

(i) $\mathcal{U} = H^1(0, T; \mathbb{R}^N)$;

(ii) $\mathcal{U}$ is finite-dimensional (e.g., consisting of piecewise constant controls).

We assume that for every $u \in L^2(0,T;\mathbb{R}^N)$, the state equation $y_t + Ay = Bu$ admits a unique solution $y \in W(0,T) := H^1(0,T;V^*) \cap L^2(0,T;V)$ for some Hilbert space $V \subset H^1(\Omega)$ (depending on the specific boundary conditions). This defines an affine and bounded solution operator $S : L^2(0,T;\mathbb{R}^N) \to W(0,T)$. We will denote the corresponding linear solution operator (i.e., for $y_0 = 0$) with S_0.

For the sake of presentation, we restrict ourselves in the following to the case of two control components; the results remain valid for $N > 2$ components (although it should be pointed out that, in contrast to the convex approach in [6], the number of terms in (1.1) grows as $\binom{N}{2}$). We hence consider for $\varepsilon \geq 0$ the problem

$$\min_{u\in\mathcal{U}} \frac{1}{2}\|Su - y^d\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \frac{\alpha}{2}\|u\|^2_{L^2(0,T;\mathbb{R}^2)} + \frac{\varepsilon}{2}\|u_t\|^2_{L^2(0,T;\mathbb{R}^2)} + \beta\int_0^T |u_1(t)u_2(t)|\,dt. \tag{2.1}$$

If $\mathcal{U}$ is finite-dimensional, it is understood that $\varepsilon = 0$; otherwise we require $\varepsilon > 0$. Keeping $\varepsilon \geq 0$ fixed, we will denote the cost functional in (2.1) by J_β.

Theorem 2.1. *There exists a minimizer $\bar u \in \mathcal{U}$ to (2.1).*

Proof. We first consider the case of $\mathcal{U} = H^1(0,T;\mathbb{R}^2)$. Since J_β is bounded from below, there exists a minimizing sequence $\{u_n\}_{n\in\mathbb{N}}$ that is bounded in $H^1(0,T;\mathbb{R}^2)$. Hence, by coercivity of J_β, there exists a subsequence, still denoted by $\{u_n\}_{n\in\mathbb{N}}$, with $u_n \rightharpoonup \bar u$ in $H^1(0,T)$ and $u_n \to \bar u$ pointwise in $(0,T)$. This implies pointwise convergence of $|u_{n,1}(t)u_{n,2}(t)| \to |\bar u_1(t)\bar u_2(t)|$. Together with the continuity of S and the weak lower semicontinuity of norms, this implies

$$J_\beta(\bar u) \leq \liminf_{n\to\infty} J_\beta(u_n) = \inf_{u\in\mathcal{U}} J_\beta(u),$$

i.e., $\bar u$ is a minimizer.

The case of $\mathcal{U}$ finite dimensional follows similarly, since boundedness in $L^2(0,T;\mathbb{R}^N)$ then directly implies strong and hence pointwise convergence. □

We now address the convergence of solutions to (2.1) as $\beta \to \infty$ to a solution to the "hard switching" control problem

$$\begin{cases} \min\limits_{u\in\mathcal{U}} \frac{1}{2}\|Su - y^d\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \frac{\alpha}{2}\|u\|^2_{L^2(0,T;\mathbb{R}^2)} + \frac{\varepsilon}{2}\|u_t\|^2_{L^2(0,T;\mathbb{R}^2)} \\ \text{s.t.} \quad u_1(t)u_2(t) = 0, \quad t \in [0,T]. \end{cases} \tag{2.2}$$

Proposition 2.2. *The family $\{u_\beta\}_{\beta\geq 0}$ of minimizers to (2.1) contains at least one convergent sequence $\{u_{\beta_n}\}_{n\in\mathbb{N}}$ with $\beta_n \to \infty$. The limit $\bar u \in \mathcal{U}$ of every such sequence is a solution to (2.2).*

Proof. Again, we only consider the case of $\mathcal{U} = H^1(0,T;\mathbb{R}^2)$, the other case being analogous. First, let $\{u_{\beta_n}\}_{n\in\mathbb{N}}$ with $\beta_n \to \infty$ be a sequence of minimizers to (2.1). Since $J_\beta(u_\beta) < J_\beta(0) = J_0(0)$ for any $\beta > 0$, this sequence is bounded in $H^1(0,T;\mathbb{R}^2)$ and hence contains a subsequence $\{u_n\}_{n\in\mathbb{N}}$, with $u_n \rightharpoonup \bar u$ in $H^1(0,T)$ and $u_n \to \bar u$ pointwise in $(0,T)$. Furthermore, $J_\beta(u_\beta) < J_0(0)$ also implies

$$\int_0^T |u_{n,1}(t)u_{n,2}(t)|\,dt \leq \beta_n^{-1} J_0(0) \to 0.$$

Hence,

$$\bar{u}_1(t)\bar{u}_2(t) = \lim_{n\to\infty} u_{n,1}(t)u_{n,2}(t) = 0 \qquad \text{for all } t \in [0,T].$$

Now, let $\{u_n\}_{n\in\mathbb{N}}$ be any such sequence. Together with optimality of u_n, the above implies that for any $\tilde{u} \in \mathcal{U}$ with $\tilde{u}_1(t)\tilde{u}_2(t) = 0$ in $[0,T]$, we have

$$\begin{aligned}\frac{1}{2}\|S\tilde{u} - y^d\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \frac{\alpha}{2}\|\tilde{u}\|^2_{L^2(0,T;\mathbb{R}^2)} + \frac{\varepsilon}{2}\|\tilde{u}_t\|^2_{L^2(0,T;\mathbb{R}^2)} \\ &= J_{\beta_n}(\tilde{u}) \geq J_{\beta_n}(u_n) \geq J_0(u_n) \\ &= \frac{1}{2}\|Su_n - y^d\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \frac{\alpha}{2}\|u_n\|^2_{L^2(0,T;\mathbb{R}^2)} + \frac{\varepsilon}{2}\|u_{n,t}\|^2_{L^2(0,T;\mathbb{R}^2)}.\end{aligned}$$

Taking the limes inferior as $n \to \infty$ and using continuity of S and weak lower semicontinuity of the norms now yields $J(\tilde{u}) \geq J(\bar{u})$, i.e., $\bar{u}$ is a global minimizer. □

3 Optimality conditions

To derive optimality conditions, we can make use of the calculus of Clarke's generalized derivative [2]. First, we consider the functional

$$G : L^2(0,T;\mathbb{R}^2) \to \mathbb{R}, \qquad u \mapsto \int_0^T |u_1(t)u_2(t)|\,dt,$$

whose generalized gradient can be computed via a chain rule.

Proposition 3.1. *For any $u \in L^2(0,T;\mathbb{R}^2)$, the functional G is regular (in the sense of Clarke), and its generalized derivative is given pointwise by*

$$\partial_C G(u) = \operatorname{sign}(u_1u_2)\begin{pmatrix} u_2 \\ u_1 \end{pmatrix}.$$

Proof. First, $H : L^1(0,T) \to \mathbb{R}$, $v \mapsto \int_0^T |v|\,dt$, is finite-valued, locally Lipschitz and convex. Hence, H is regular at any $v \in L^1(0,T)$, and the generalized derivative coincides with the subdifferential in the sense of convex analysis; see [2, Prop. 2.2.7]. Furthermore, $T : L^2(0,T;\mathbb{R}^2) \to L^1(0,T)$, $u \mapsto u_1u_2$, is strictly differentiable. Hence, by [2, Theorem 2.3.10], $G = H \circ T$ is regular at any $u \in L^2(0,T;\mathbb{R}^2)$, and

$$\partial_C G(u) = T'(u)^*\partial H(T(u)).$$

The claim now follows from $\partial H(v) = \operatorname{sign}(v)$ and $T'(u)^*v = (vu_2, vu_1)^T$. □

Since G is regular and the remaining terms in (2.1) are continuously Fréchet-differentiable, we can apply the sum rule for generalized gradients, e.g., from [2, Prop. 2.3.3], to obtain the following first-order necessary optimality conditions.

Theorem 3.2. *Any local minimizer $\bar{u} \in \mathcal{U}$ to (2.1) satisfies*

$$0 \in S_0^*(S\bar{u} - z) + \alpha\bar{u} - \varepsilon\bar{u}_{tt} + \beta \operatorname{sign}(\bar{u}_1\bar{u}_2)\begin{pmatrix}\bar{u}_2\\ \bar{u}_1\end{pmatrix}. \tag{3.1}$$

Note that for $\mathcal{U} = H^1(0,T;\mathbb{R}^2)$, the right-hand side is to be understood as a subset of $H^1(0,T;\mathbb{R}^2)^*$.

Using directional derivatives, we can show that non-switching arcs can have a length of at most $\sqrt{\varepsilon}$. In the following, we set $\|S_0\| := \|S_0\|_{\mathcal{L}(L^2(0,T;\mathbb{R}^2),L^2(0,T;L^2(\omega_{\text{obs}}))}$ for brevity.

Theorem 3.3. *If $\beta > (\|S_0\|^2 + \alpha + \pi^2)$, then $\bar{u}_1(t)\bar{u}_2(t) = 0$ for all $t \in [0,T]$ apart from intervals of length at most $\sqrt{\varepsilon}$.*

Proof. Let $u \in L^2(0,T;\mathbb{R}^2)$ be given and assume that there exists $t_0 \in (0,T)$ and $\delta > \sqrt{\varepsilon}$ such that $u_1(t)u_2(t) \neq 0$ for all $t \in (t_0, t_0+\delta)$ and $u_1(t_0)u_2(t_0) = u_1(t_0+\delta)u_2(t_0+\delta) = 0$. Without loss of generality, we can assume that both $u_1(t) > 0$ and $u_2(t) > 0$ for all $t \in (t_0, \delta)$. Furthermore, since $\beta > (\|S_0\|^2 + \alpha + \pi^2)$, there exists a $\rho \in (0, \delta/2)$ such that

$$\beta > \left(\|S_0\|^2 + \alpha + \frac{\delta^2}{(\delta - 2\rho)^2}\pi^2\right) > (\|S_0\|^2 + \alpha + \pi^2).$$

Set

$$h_1(t) = \begin{cases} \sqrt{\frac{2}{\delta-2\rho}} \sin\left(\frac{\pi}{\delta-2\rho}(t - t_0 - \rho)\right) & t \in I := [t_0+\rho, t_0+\delta-\rho],\\ 0 & \text{else.}\end{cases}$$

Then, $h_1 \in H^1(0,T)$ with

$$\|h_1\|^2_{L^2(0,T)} = 1 \qquad \text{and} \qquad \|(h_1)_t\|^2_{L^2(0,T)} = \frac{\pi^2}{(\delta-2\rho)^2}.$$

We now consider directional derivatives in the specific direction $h = (h_1, -h_1)$. For this purpose, we first introduce

$$g : \mathbb{R} \to \mathbb{R}, \qquad s \mapsto G(u + sh),$$

and show that g is differentiable in 0. First, we have by definition of h that

$$\begin{aligned} g(s) - g(0) &= \int_0^T \left[|(u_1 + sh_1)(u_2 - sh_1)| - |u_1u_2|\right] dt\\ &= \int_I \left[|(u_1 + sh_1)(u_2 - sh_1)| - |u_1u_2|\right] dt. \end{aligned}$$

By the continuity of u and h, there exists $\tilde{s} > 0$ such that

$$u_1(t) \pm sh_1(t) \geq 0 \quad \text{and} \quad u_2(t) \pm sh_1(t) \geq 0 \qquad \text{for all } t \in I \text{ and } s \in [-\tilde{s}, \tilde{s}]. \tag{3.2}$$

Furthermore, $u_1(t)u_2(t) > 0$ for all $t \in I$ by assumption. Hence,

$$g(s) - g(0) = \int_I [(u_1 + sh_1)(u_2 - sh_1) - u_1u_2]\; dt = \int_I \left[sh_1(u_2 - u_1) - s^2h_1^2\right]\; dt \tag{3.3}$$

and therefore

$$g'(0) = \lim_{s\to 0} \frac{g(s) - g(0)}{s} = \int_I h_1(u_2 - u_1)\, dt.$$

For the second derivative $g''(0)$, we can proceed in the same way using (3.2) to obtain

$$\begin{aligned} g''(0) &= \lim_{s\to 0} \frac{g(s) - 2g(0) + g(-s)}{s^2} \\ &= \lim_{s\to 0} \frac{1}{s^2} \int_I [(u_1 + sh_1)(u_2 - sh_1) - 2u_1u_2 + (u_1 - sh_1)(u_2 + sh_1)]\; dt \\ &= \lim_{s\to 0} \int_I -2h_1^2\, dt = -\|h\|^2_{L^2(0,T;\mathbb{R}^2)}. \end{aligned} \tag{3.4}$$

Comparing (3.3) and (3.4), we obtain that

$$g(s) = g(0) + sg'(0) + \frac{s^2}{2}g''(0). \tag{3.5}$$

Since the remaining terms in the cost functional are differentiable, we have for

$$j : \mathbb{R} \to \mathbb{R}, \qquad s \mapsto J(u + sh),$$

that

$$\begin{aligned} j''(0) &= \|S_0h\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \alpha\|h\|^2_{L^2(0,T;\mathbb{R}^2)} + \varepsilon\|h_t\|^2_{L^2(0,T;\mathbb{R}^2)} - \beta\|h\|^2_{L^2(0,T;\mathbb{R}^2)} \\ &\le (\|S_0\|^2 + \alpha - \beta)\|h\|^2_{L^2(0,T;\mathbb{R}^2)} + \varepsilon\|h_t\|^2_{L^2(0,T;\mathbb{R}^2)} \\ &= 2(\|S_0\|^2 + \alpha - \beta) + 2\frac{\pi^2}{(\delta - 2\rho)^2}\varepsilon \\ &\le 2\left(\|S_0\|^2 + \alpha - \beta + \frac{\delta^2}{(\delta - 2\rho)^2}\pi^2\right) < 0 \end{aligned}$$

by the choice of δ, ρ, and β.

Now, from (3.5) and the fact that the remaining terms in J are quadratic and hence that the second-order Taylor expansion of j is exact, it follows that for all s with $sj'(0) \le 0$, we have

$$J(u + sh) = j(s) = j(0) + sj'(0) + \frac{s^2}{2}j''(0) < j(0) = J(u).$$

Hence, u cannot be a local minimizer. □

Note that the function h_1 constructed in this proof is the shifted and scaled first Dirichlet eigenfunction of the Laplacian on the interval $(0,1)$, which is the only function for which equality holds in the Poincaré inequality. Thus, the above result is likely to be sharp, i.e., perfect switching for $\varepsilon > 0$ cannot be guaranteed in general. However, it follows from Proposition 2.2 that non-switching arcs have to vanish for $\beta \to \infty$.

We now turn to the finite-dimensional case, where $\mathcal{U}$ consists of piecewise constant functions on a given grid

$$0 = t_1 < \cdots < t_M = T,$$

and we can (and must) take $\varepsilon = 0$.

Theorem 3.4. *If $\mathcal{U}$ consists of piecewise constant functions and $\beta > \|S_0\|^2 + \alpha$, then $\bar{u}_1(t)\bar{u}_2(t) = 0$ for all $t \in [0,T]$.*

Proof. We proceed as above. Let $u \in \mathcal{U}$ with $u_1(t)u_2(t) \neq 0$ be given. Then there exists an interval $I_j := (t_j, t_{j+1}]$ such that $u_1(t)u_2(t) \neq 0$ for all $t \in I_j$. As before, we can assume $u_1 > 0$ and $u_2 > 0$ on I_j. We now choose

$$h_1(t) = \begin{cases} 1 & t \in I_j, \\ 0 & \text{else,} \end{cases}$$

with $\|h_1\|^2_{L^2(0,T)} = \tau_j := t_{j+1} - t_j$ and consider again for $h = (h_1, -h_1)^T$ the function $g(s) := G(u + sh)$. We then obtain (using the fact that u is constant on I_j) that

$$g(s) - g(0) = \tau_j \left[|(u_1 + s)(u_2 - s)| - |u_1 u_2| \right].$$

Since u_1 and u_2 here are strictly positive constants, there again exists an $\tilde{s}$ such that $u_1 \pm s > 0$ and $u_2 \pm s > 0$ on for all $s \in [-\tilde{s}, \tilde{s}]$. Hence,

$$g'(0) = \tau_j(u_2 - u_1) \qquad \text{and} \qquad g''(0) = -2\tau_j = -\|h\|^2_{L^2(0,T;\mathbb{R}^2)}.$$

As above, the latter implies that

$$j''(0) \leq 2(\|S_0\|^2 + \alpha - \beta)\tau_j < 0$$

and hence that $J(u + sh) < J(u)$ for all s with $sj'(0) \leq 0$. $\square$

Remark 1. *The proof of Theorem 3.4 relies on the fact that for piecewise constant u, i.e., $u_i = \sum_{j=1}^M \xi_i^j \chi_{I_j}$ for $i \in \{1,2\}$, there holds*

$$G(u) = \sum_{j=1}^{M} \tau_j |\xi_1^j \xi_2^j|. \tag{3.6}$$

If $\mathcal{U}$ is an arbitrary finite-dimensional subspace of $L^2(0,T;\mathbb{R}^2)$, i.e., $u_i = \sum_{j=1}^M \xi_i^j e_j$ for $i \in \{1,2\}$ and some basis functions e_j, and the switching penalty G is replaced by the right-hand side of (3.6), the above proof can be modified to show that $\beta > C^{-1}(\|S_0\|^2 + \alpha)$ implies that $\xi_1^j \xi_2^j = 0$ for all j, where C is the constant of equivalence for the discrete and continuous norm on $\mathcal{U} \subset L^2(0,T;\mathbb{R}^2)$.

The relation between pointwise switching and switching of the coefficients depends on the specific choice of e_j.

4 Numerical solution

The numerical solution is based on a primal-dual reformulation of the optimality condition (3.1), which states that there exists a

$$\bar{q} \in \beta \operatorname{sign}(\bar{u}_1\bar{u}_2) = \partial(\beta\|\cdot\|_{L^1})(\bar{u}_1\bar{u}_2) \subset L^\infty(0,1) \tag{4.1}$$

such that

$$S_0^*(S\bar{u} - z) + \alpha\bar{u} - \varepsilon\bar{u}_{tt} + \bar{q}\begin{pmatrix}\bar{u}_2\\ \bar{u}_1\end{pmatrix} = 0,$$

where $\partial(\|\cdot\|_{L^1})$ denotes the convex subdifferential of the $L^1(0,T)$ norm. We now proceed as in the case of sparse control, e.g., [4, 5, 7]: Since this is a convex functional, (4.1) is equivalent to

$$\bar{u}_1\bar{u}_2 \in \partial(\mathcal{I}_{B_\infty(0,\beta)})(\bar{q}), \tag{4.2}$$

where $\mathcal{I}_{B_\infty(0,\beta)}$ denotes the indicator function (in the sense of convex analysis) of the closed ball in $L^\infty(0,T)$ around 0 with radius β. The subdifferential on the right-hand side of (4.2) is now replaced by its Hilbert space Moreau–Yosida regularization, which for any $\gamma > 0$ is given by

$$\partial(\mathcal{I}_{B_\infty(0,\beta)})_\gamma(q) = \frac{1}{\gamma}\left(q - \operatorname{prox}_{\gamma\mathcal{I}_{B_\infty(0,\beta)}}(q)\right) = \frac{1}{\gamma}\left(q - \operatorname{proj}_{B_\infty(0,\beta)}(q)\right).$$

Here, prox_f denotes the proximal mapping of a convex function f, which for indicator functions of a convex set C coincides with the metric projection proj_C onto C. The Moreau–Yosida regularization of (3.1) can thus be written as

$$\begin{cases} S_0^*(Su_\gamma - z) + \alpha u_\gamma - \varepsilon(u_\gamma)_{tt} + q_\gamma\begin{pmatrix}u_{\gamma,2}\\ u_{\gamma,1}\end{pmatrix} = 0\\[1ex] \gamma u_{\gamma,1}u_{\gamma,2} - \max(0, q_\gamma - \beta) - \min(0, q_\gamma + \beta) = 0,\end{cases} \tag{4.3}$$

where the second equation is to be understood pointwise almost everywhere in $(0,T)$.

It is known that the Moreau–Yosida regularization of the subdifferential of the indicator function is equivalent to adding an L^2 norm to the primal functional; see, e.g., [4, Remark 3.2]. In our case, this leads to the problem

$$\begin{aligned}\min_{u\in\mathcal{U}} \frac{1}{2}\|Su - y^d\|^2_{L^2(0,T;L^2(\omega_{\text{obs}}))} + \frac{\alpha}{2}\|u\|^2_{L^2(0,T;\mathbb{R}^2)} + \frac{\varepsilon}{2}\|u_t\|^2_{L^2(0,T;\mathbb{R}^2)}\\ + \beta\int_0^T |u_1(t)u_2(t)|\,dt + \frac{\gamma}{2}\int_0^T |u_1(t)u_2(t)|^2\,dt.\end{aligned} \tag{4.4}$$

A similar proof as for Theorem 2.1 yields existence of a solution u_γ. Proceeding as in the derivation of Theorem 3.2, we deduce the existence of

$$q_\gamma \in \beta\operatorname{sign}(u_{\gamma,1}u_{\gamma,2}) + \gamma u_{\gamma,1}u_{\gamma,2} = \partial\left(\beta\|\cdot\|_{L^1} + \frac{\gamma}{2}\|\cdot\|^2_{L^2}\right)(u_{\gamma,1}u_{\gamma,2}) \tag{4.5}$$

such that the first relation of (4.3) holds. As shown in [4, Remark 3.2], the subdifferential inclusion (4.5) is equivalent to

$$u_{\gamma,1}u_{\gamma,2} = \frac{1}{\gamma}\max(0, q_\gamma - \beta) + \frac{1}{\gamma}\min(0, q_\gamma + \beta),$$

which is a reformulation of the second relation of (4.3). Hence, we obtain existence of a solution (u_γ, q_γ) to (4.3). A standard argument shows weak subsequential convergence of u_γ as $\gamma \to 0$ to a minimizer $\bar{u} \in \mathcal{U}$ to (2.1).

Since $u_\gamma \in H^1(0,T;\mathbb{R}^2)$, it follows from (4.5) that $q_\gamma \in L^\infty(0,T)$. It is well-known that the pointwise max and min are Newton differentiable from $L^p(0,T)$ to $L^1(0,T)$ for any $p > 1$, with Newton derivative in direction $h \in L^p(0,T)$ given pointwise almost everywhere by

$$D_N \max(0, q-\beta)h = \begin{cases} h & \text{if } q \leq \beta, \\ 0 & \text{else,} \end{cases} \qquad D_N \min(0, q+\beta)h = \begin{cases} h & \text{if } q \geq -\beta, \\ 0 & \text{else.} \end{cases}$$

Thus, (4.3) considered as an operator equation from $H^1(0,T;\mathbb{R}^2) \times L^2(0,T)$ to $H^1(0,T;\mathbb{R}^2)^* \times L^2(0,T)$ is Newton differentiable. We therefore apply a semismooth Newton method for its solution.

In the numerical realization, we follow a homotopy approach. Note that the Moreau–Yosida regularization acts as a smoothing of the indicator function. At the same time, it exacerbates the nonconvexity of the problem; cf. (4.4). We therefore proceed as follows: Starting from an initial guess $(u^0, q^0) = (0,0)$ and $\gamma = 0$, we solve a sequence of problems for increasing $\gamma \geq 0$ and fixed $\beta = \beta_{\min}$ until a given final value $\gamma_{\max} > 0$ is reached, taking the previous solution as an initial guess. Keeping this value of $\gamma = \gamma_{\max}$ fixed, we then similarly increase β from some $\beta_{\min}$ to a value $\beta_{\max}$ which is chosen adaptively as the first β such that the maximal pointwise violation $\sigma_{\text{sw}} := \|u_1 u_2\|_\infty = \max_{t\in I} |u_1(t)u_2(t)|$ of the switching property drops below a given tolerance. This adaptive choice is advantageous since it ensures large enough β to obtain switching while avoiding too large β that would complicate the numerical solution at no additional benefit. Finally, keeping $\beta = \beta_{\max} > 0$ fixed, γ is reduced again until some specified $\gamma_{\min}$ is reached.

In our experience, the homotopy parameters $\beta_{\min}$ and $\gamma_{\min}$ can be chosen very small in general, while $\gamma_{\max}$ has to be chosen above a critical value that depends on the problem. For any such choice, the obtained controls are then robust with respect to the value of $\gamma_{\max}$.

5 Numerical examples

We illustrate the properties of solutions to (2.1) using the heat equation on the unit square $\Omega = (-1,1)^2$ with homogeneous Neumann and initial conditions and different choices control and observation configurations. The final time is always set to $T = 10$. The spatial discretization uses standard piecewise linear finite elements, while the temporal discretization is also chosen as continuous and piecewise linear to obtain a conforming finite element approximation of the H^1 seminorm in (2.1) as well as the weak (temporal) Laplacian in (3.1). Following Remark 1, the switching penalty is replaced by its discrete version (i.e., the right-hand side of (3.6)), which allows a componentwise formulation of the second relation in (4.5). We similarly replace the L^2 norm by its discrete version, which can be interpreted as a mass lumping.

The first example, based on the example from [6], illustrates the effect of different control cost parameters α. In the second example, we investigate the influence of the H^1-regularization on the solution. The third example addresses the issue of local versus global minimizers via a full enumeration of stationary points for a suitably coarse discretization. The MATLAB implementation of the proposed approach used to generate these results can be downloaded from `https://github.com/clason/nonconvexswitching`.

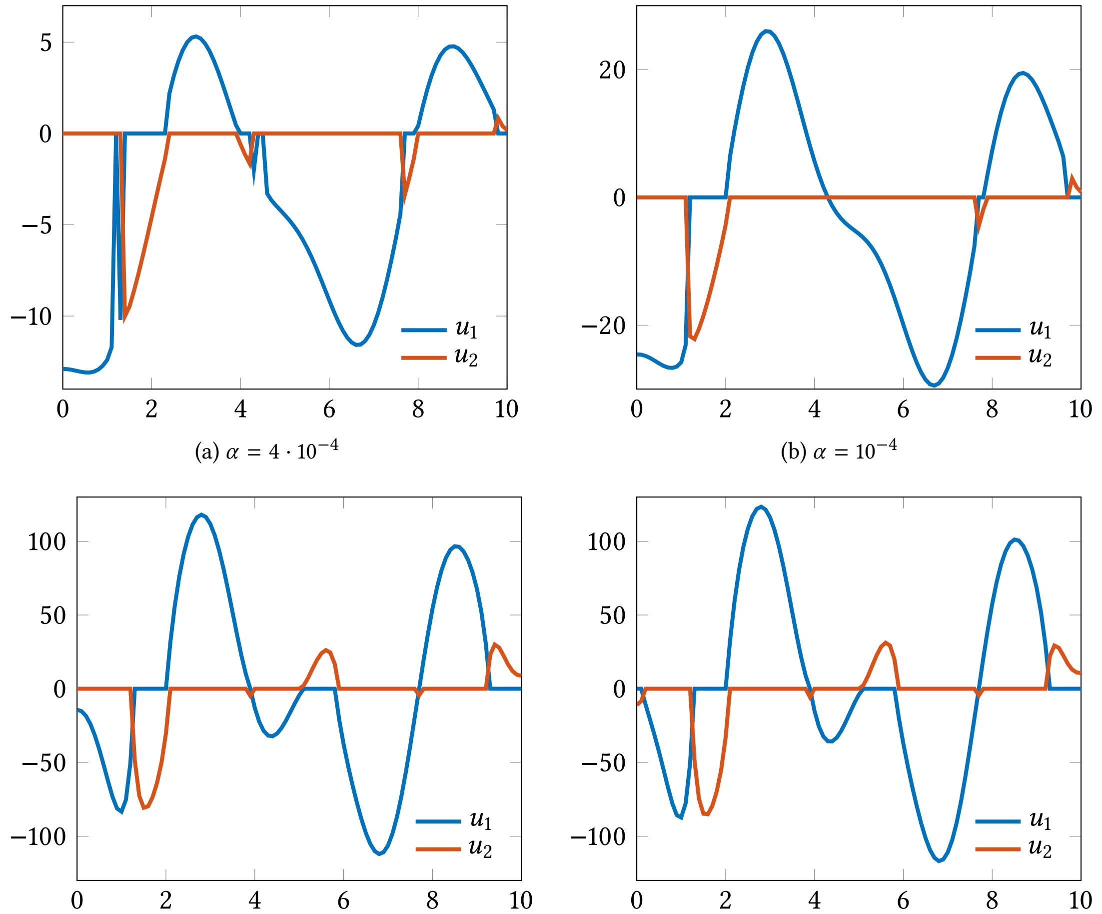
5.1 Example 1

To compare the proposed method with earlier work, we take up the example from [6] with $N = 2$ control components and 101 equidistant time points. For any $\alpha \geq 4 \cdot 10^{-4}$, the convex solution method from [6] yields switching controls, which are unique global mininimizers to the respective problems. However, smaller values of α lead to non-switching solutions. In particular, we find for $\alpha = 10^{-4}$ resp. 10^{-5} a total of 7 resp. 40 time points where both control components are essentially active.

In the following, we apply the proposed nonconvex method for different control cost parameters α and with a fixed $\varepsilon = 10^{-7}$. The homotopy parameters are set to $\beta_{\min} = 10^{-5}$, $\gamma_{\min} = 10^{-9}$, and $\gamma_{\max} = 10^2$, and incremented or decremented by a factor of 10. In this and the following examples, the final value $\beta_{\max}$ is chosen according to the relative tolerance $\sigma_{\text{sw}} \leq 10^{-10} \max\{\|u_1\|_\infty, \|u_2\|_\infty\}$. For each pair (β, γ), the maximum number of semismooth Newton iterations is set to 5. While this leads to early termination of the semismooth Newton method in the first homotopy steps (to save numerical effort), we always observe convergence with relative tolerance 10^{-6} or absolute tolerance 10^{-7} in the residual norm of the optimality system (4.3) during the later steps of the homotopy.

The solutions for different control cost parameters α are shown in Figure 1, where the solution for $\alpha = 4 \cdot 10^{-4}$ is very similar to the solution to the convex problem (differing only in the points t_j with $u_1(t_j) = u_2(t_j) = 0$, where in the convex case one of the control components is active). Since the convex solution is computed with $\varepsilon = 0$, this indicates that the influence of ε on the solution is negligible for this choice of α and ε. For smaller values of α, the solutions to the nonconvex problem maintain the switching property while increasing in amplitude until $\alpha \leq 10^{-8}$, after which any choice of α yields the same numerical solution.

This is illustrated in more detail in Table 1, which shows in particular that the tracking error (and therefore the optimal functional value $\bar{J}$) can be reduced significantly by decreasing α. At the same time, the number N_{sw} of "switching points" (i.e., points t_j such that $|u_1(t_j)| \geq |u_2(t_j)|$ but $|u_1(t_{j+1})| < |u_2(t_{j+1})|$ or vice versa) increases only slightly due to the presence of the H^1-regularization. (This relation will be investigated more closely in the next example.) The fourth column gives the switching error $\sigma_{\text{sw}} = \max_{t \in I} |u_1 u_2|$, confirming quantitatively that the switching property is satisfied up to a high accuracy for all cases. The last two columns address the convergence of the homotopy semismooth Newton method by giving the residual norm of the optimality system (4.3) for the final switching penalty $\beta = \beta_{\max}$ and $\gamma = 10^{-9}$. Note that in all cases, we have $\beta_{\max} > \alpha$ and hence a genuinely nonconvex problem.



(a) $\alpha = 4 \cdot 10^{-4}$ (b) $\alpha = 10^{-4}$

(c) $\alpha = 10^{-6}$ (d) $\alpha = 10^{-8}$

Figure 1: Optimal switching controls for different values of α

Table 1: Results for different values of α (Example 1): optimal value $\bar{J}$, number of switching points N_{sw}, switching error σ_{sw}, residual norm of the optimality system (4.3), final switching penalty $\beta_{\max}$

α	$\bar{J}$	N_{sw}	σ_{sw}	optimality	$\beta_{\max}$
$4 \cdot 10^{-4}$	0.762	11	$5 \cdot 10^{-11}$	$7 \cdot 10^{-15}$	10^{3}
10^{-4}	0.581	7	$3 \cdot 10^{-11}$	$8 \cdot 10^{-14}$	10^{2}
10^{-5}	0.209	9	$4 \cdot 10^{-11}$	$4 \cdot 10^{-14}$	10^{2}
10^{-6}	0.069	9	$3 \cdot 10^{-16}$	$1 \cdot 10^{-14}$	10^{-4}
10^{-7}	0.050	10	$2 \cdot 10^{-10}$	$2 \cdot 10^{-13}$	10^{1}
10^{-8}	0.048	10	$7 \cdot 10^{-17}$	$2 \cdot 10^{-14}$	10^{-4}
10^{-9}	0.048	10	$7 \cdot 10^{-17}$	$2 \cdot 10^{-14}$	10^{-4}

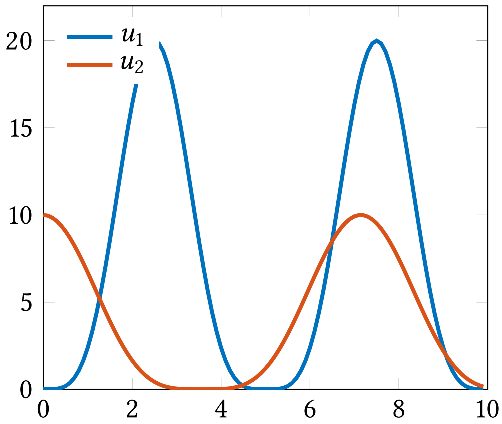


Figure 2: Generating control u_d for Example 2

5.2 Example 2

The second example addresses the influence of the H^1-regularization parameter ε on the solution. Here, we set

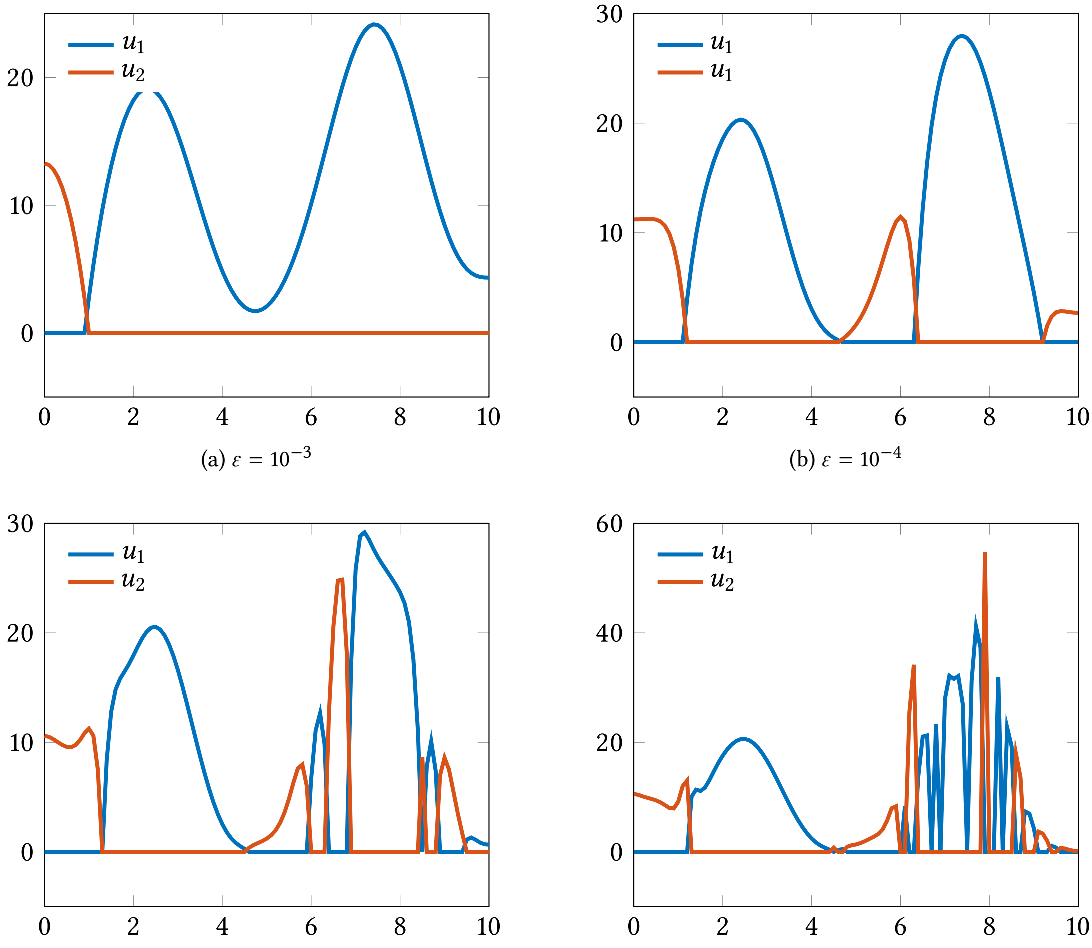
$$\omega_1 = \{(x_1, x_2) \in \Omega : x_1 \leq 0\}, \qquad \omega_2 = \{(x_1, x_2) \in \Omega : x_1 > 0\},$$

and $Bu = (\chi_{\omega_1}(x)u_1(t) + \chi_{\omega_2}(x)u_2(t))/10$. The desired state y^d is the solution corresponding to the control

$$u_d = \left(20 \sin^4(2\pi t/T),\ 10 \cos^4(1.4\pi t/T)\right),$$

see Figure 2. By construction, this desired state is expected to be difficult to attain by a pure switching control since both controls contribute significantly during the second half of the time interval. To investigate the influence of ε on the solution, we fix $\alpha = 10^{-6}$, which corresponds to comparatively small control costs. The homotopy parameters are set to $\beta_{\min} = 10^{-5}$, $\gamma_{\min} = 10^{-9}$, and $\gamma_{\max} = 10^4$, and are incremented or decremented by a factor of 10. The maximal number of semismooth Newton iterations is again set to 5; we observe the same convergence behavior as in the first example.

The solutions for different values of ε are shown in Figure 3. We see that all solutions exhibit switching. During the first half of the time interval, we observe for all $\varepsilon > 0$ the expected switching behavior from the generating control u_d. During the second half, the number of switching points increases for $\varepsilon \to 0$. A quantitative illustration of the dependence on ε is given in Table 2, which shows that the number N_{sw} decreases monotonically to zero as ε increases. At the same time, the tracking is obviously increased, resulting in a larger optimal functional value $\bar{J}$. The fourth column confirms the switching property of the optimal control for all ε. In particular, the last row demonstrates that the proposed approach is feasible for computing switching controls for very small control cost and regularization parameters ($\alpha = 10^{-6}$, $\varepsilon = 10^{-7}$, $\gamma = 10^{-9}$) compared to the switching penalty ($\beta = 10^3$). Again, in all cases $\beta_{\max} > \alpha$, and hence the problem is genuinely nonconvex.



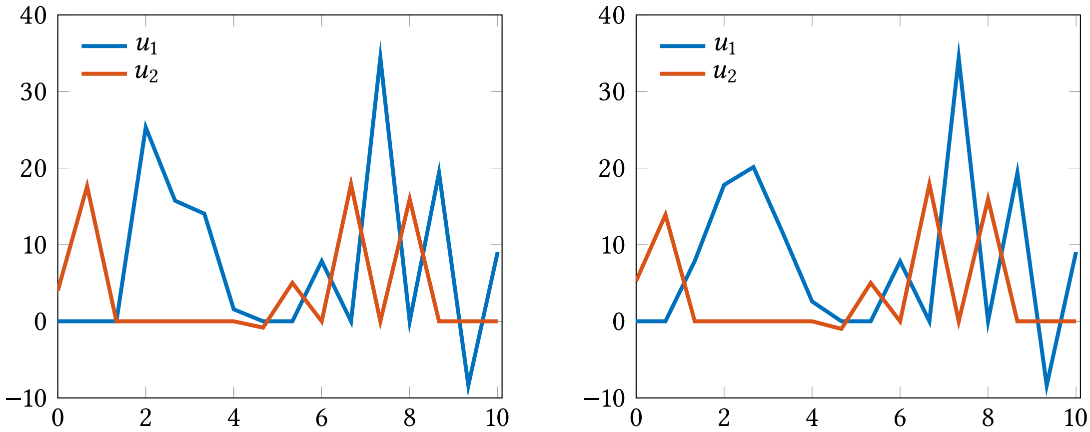
(a) $\varepsilon = 10^{-3}$ (b) $\varepsilon = 10^{-4}$

(c) $\varepsilon = 10^{-5}$ (d) $\varepsilon = 10^{-6}$

Figure 3: Optimal switching controls for different values of ε

Table 2: Results for different values of ε (Example 2): optimal functional value $\bar{J}$, number of switching points N_{sw}, switching error σ_{sw}, residual norm of the optimality system (4.3), final switching penalty $\beta_{\max}$

ε	$\bar{J}$	N_{sw}	σ_{sw}	optimality	$\beta_{\max}$
10^{-2}	2.088	0	$5 \cdot 10^{-17}$	$2 \cdot 10^{-11}$	10^{-2}
10^{-3}	1.202	1	$4 \cdot 10^{-22}$	$1 \cdot 10^{-11}$	10^{0}
10^{-4}	0.542	4	$6 \cdot 10^{-8}$	$6 \cdot 10^{-11}$	10^{3}
10^{-5}	0.318	9	$2 \cdot 10^{-10}$	$2 \cdot 10^{-11}$	10^{3}
10^{-6}	0.274	14	$3 \cdot 10^{-10}$	$2 \cdot 10^{-12}$	10^{3}
10^{-7}	0.124	22	$2 \cdot 10^{-11}$	$1 \cdot 10^{-12}$	10^{3}



(a) computed switching control

(b) globally optimal switching control

Figure 4: Comparison of computed and globally optimal switching controls

5.3 Full enumeration for a coarse discretization

The last example shows that the nonconvex problem of computing switching controls may possess many stationary points. For this purpose, we investigate the previous example with only 16 degrees of freedom per control component, where the set of all stationary points corresponding to pure switching controls can be generated by total enumeration. In this case, there are $2^{16} = 65536$ possible configurations for switching controls. For each configuration (setting $\alpha = 10^{-6}$, $\varepsilon = 10^{-5}$, $\gamma = 0$, and arbitrary $\beta > 0$), the solution to the first-order necessary conditions (3.1) can be computed directly. For $\beta = 10^3$ this system always yields a unique solution resulting in 65536 stationary points with corresponding functional values between 0.0887 and 2.44. The smallest value obviously corresponds to the global minimizer of (2.1) within the class of switching controls for this discretization, which is shown in Figure 4a. For comparison, the switching control computed by the proposed approach (with $\beta_{\max} = 10^3$, residual norm $6 \cdot 10^{-11}$ and switching error $\sigma_{\text{sw}} = 8 \cdot 10^{-11}$) is shown in Figure 4b. Note that this control differs from the global minimizer only slightly in the first half of the time interval; in particular, the switching configuration is identical apart from the third degree of freedom. The computed objective value of 0.1133 is smaller than 99.74 percent of the objective values of all stationary points.

6 Conclusion

Penalization of switching constraints leads to a nonconvex optimal control problem if the switching penalty is larger than the control costs. Under additional H^1 regularization or restriction to finite-dimensional controls, existence of solutions can be shown. Using tools from nonsmooth analysis allows deriving optimality conditions and showing (for finite-dimensional controls) exact penalization properties. These optimality conditions are amenable to numerical solution via a (still nonconvex) Moreau–Yosida regularization and a semismooth Newton method.

By virtue of the embedding of $H^1(0,T)$ into $C([0,T])$, switching of optimal controls can only occur in points where $\bar{u}_1(t) = \bar{u}_2(t) = 0$. This is not the case if H^1 regularization is replaced with regularization in the space of functions of bounded variation; however, this would introduce new difficulties due to the additional nonsmoothness and the more complicated functional-analytic setting and is therefore the subject of further work.

Acknowledgments

This work was supported in part by the Austrian Science Fund (FWF) under grant SFB F32 (SFB "Mathematical Optimization and Applications in Biomedical Sciences"), by BioTechMed-Graz, and by the ERC advanced grant 668998 (OCLOC) under the EU's H2020 research program.

References

[1] I. Capuzzo Dolcetta and L. C. Evans, *Optimal switching for ordinary differential equations*, SIAM Journal on Control and Optimization, 22 (1984), pp. 143–161.

[2] F. H. Clarke, *Optimization and Nonsmooth Analysis*, Classics Appl. Math. 5, SIAM, 2 ed., 1990.

[3] C. Clason, K. Ito, and K. Kunisch, *A convex analysis approach to optimal controls with switching structure for partial differential equations*, ESAIM: Control, Optimisation and Calculus of Variations, 22 (2016), pp. 581–609.

[4] C. Clason and K. Kunisch, *A duality-based approach to elliptic control problems in non-reflexive Banach spaces*, ESAIM: Control, Optimisation and Calculus of Variations, 17 (2011), pp. 243–266.

[5] ——, *A measure space approach to optimal source placement*, Computational Optimization and Applications, 53 (2012), pp. 155–171.

[6] C. Clason, A. Rund, K. Kunisch, and R. C. Barnard, *A convex penalty for switching control of partial differential equations*, Systems & Control Letters, 89 (2016), pp. 66–73.

[7] C. Clason and A. Schiela, *Optimal control of elliptic equations with positive measures*, ESAIM: Control, Optimisation and Calculus of Variations, forthcoming (2015).

[8] M. Gugat, *Optimal switching boundary control of a string to rest in finite time*, ZAMM, 88 (2008), pp. 283–305.

[9] F. M. Hante, G. Leugering, and T. I. Seidman, *Modeling and analysis of modal switching in networked transport systems*, Applied Mathematics and Optimization, 59 (2009), pp. 275–292.

[10] F. M. Hante and S. Sager, *Relaxation methods for mixed-integer optimal control of partial differential equations*, Computational Optimization and Applications, 55 (2013), pp. 197–225.

[11] D. Liberzon, *Switching in systems and control*, Systems & Control: Foundations & Applications, Birkhäuser Boston, Inc., Boston, MA, 2003.

[12] Q. Lü and E. Zuazua, *Robust null controllability for heat equations with unknown switching control mode*, Discrete Contin. Dyn. Syst., B, 34 (2014), pp. 4183–4210.

[13] T. I. Seidman, *Optimal control of a diffusion/reaction/switching system*, Evol. Equ. Control Theory, 2 (2013), pp. 723–731.

[14] R. Shorten, F. Wirth, O. Mason, K. Wulff, and C. King, *Stability criteria for switched and hybrid systems*, SIAM Rev., 49 (2007), pp. 545–592.

[15] S. Stojanovic and J. M. Yong, *Optimal switching for partial differential equations I*, J. Math. Anal. Appl., 138 (1989), pp. 418–438.

[16] ——, *Optimal switching for partial differential equations II*, J. Math. Anal. Appl., 138 (1989), pp. 439–460.

[17] L. Wang and Q. Yan, *Time optimal controls of semilinear heat equation with switching control*, Journal of Optimization Theory and Applications, (2014), pp. 1–16.

[18] E. Zuazua, *Switching control*, J. Eur. Math. Soc. (JEMS), 13 (2011), pp. 85–117.