Available online at www.sciencedirect.com

ScienceDirect

Ann. I. H. Poincaré – AN ●●● (●●●●) ●●●–●●●

ANNALES
DE L'INSTITUT
HENRI
POINCARÉ
ANALYSE
NON LINÉAIRE

www.elsevier.com/locate/anihpc

Multi-bang control of elliptic systems

Christian Clason [*], Karl Kunisch

Institute of Mathematics and Scientific Computing, University of Graz, Heinrichstrasse 36, 8010 Graz, Austria

Received 29 March 2013; received in revised form 8 August 2013; accepted 31 August 2013

Abstract

Multi-bang control refers to optimal control problems for partial differential equations where a distributed control should only take on values from a discrete set of allowed states. This property can be promoted by a combination of L^2 and L^0-type control costs. Although the resulting functional is nonconvex and lacks weak lower-semicontinuity, application of Fenchel duality yields a formal primal-dual optimality system that admits a unique solution. This solution is in general only suboptimal, but the optimality gap can be characterized and shown to be zero under appropriate conditions. Furthermore, in certain situations it is possible to derive a generalized multi-bang principle, i.e., to prove that the control almost everywhere takes on allowed values except on sets where the corresponding state reaches the target. A regularized semismooth Newton method allows the numerical computation of (sub)optimal controls. Numerical examples illustrate the effectiveness of the proposed approach as well as the structural properties of multi-bang controls.

Keywords: Optimal control; Elliptic partial differential equations; Nonconvex relaxation; Fenchel duality; Newton methods

1. Introduction

This work is concerned with the problem

$$\begin{cases} \displaystyle\min_{u,y} \frac{1}{2}\|y-z\|_{L^2}^2 + \frac{\alpha}{2}\|u\|_{L^2}^2 + \beta \int_\Omega \prod_{i=1}^d |u(x)-u_i|_0 \, dx \\ \text{s.t.} \quad Ay = u, \quad u_1 \leqslant u(x) \leqslant u_d \text{ for almost every } x \in \Omega \end{cases} \tag{1.1}$$

for given $\alpha > 0$, $\beta > 0$, real numbers $u_1 < \cdots < u_d$, $d \geqslant 2$, and a target $z \in L^2(\Omega)$. We assume that $A : V \to V^*$ is an isomorphism for a Hilbert space V with continuous, compact and dense embeddings $V \hookrightarrow L^2(\Omega) \hookrightarrow V^*$ (typically, an elliptic partial differential operator). The *binary* term

$$|t|_0 := \begin{cases} 0 & \text{if } t = 0, \\ 1 & \text{if } t \neq 0, \end{cases}$$

[*] Corresponding author.
E-mail addresses: christian.clason@uni-graz.at (C. Clason), karl.kunisch@uni-graz.at (K. Kunisch).

0294-1449/$ – see front matter © 2013 Elsevier Masson SAS. All rights reserved.
http://dx.doi.org/10.1016/j.anihpc.2013.08.005

is related to Donoho's counting measure. Problem (1.1) is motivated by optimal control problems where it is only possible or desired for the control to take on values from a discrete set of given *control states* u_i (e.g., velocities or voltages), preferably those of smallest possible magnitude. In analogy to bang-bang controls, which (under suitable conditions) attain their control constraints almost everywhere, we refer to such controls as *multi-bang controls*.

Let us remark on some related control problems. For $d = 1$, $u_1 = 0$, and no control constraints, problem (1.1) was investigated in [1], where the choice of the cost was motivated by obtaining sparsity in the structure of the optimal controls. Sparsity can also be promoted by L^1-type and measure-space functionals; see, e.g. [2–4] and the references given there. We point out that although the desired controls are piecewise constant, problem (1.1) differs fundamentally from control problems with a total-variation-type penalty as considered in [4], since here the constants are fixed a priori. For $d = 2$ and $\alpha = \beta = 0$, problem (1.1) is a classical bang-bang control problem, where optimal controls satisfy a *generalized bang-bang-principle*, i.e., the control constraints u_1 and u_2 are attained almost everywhere outside a set where the optimal state reaches the target; see, e.g., [5–8]. The case $d = 3$ and $u_1 < u_2 = 0 < u_3$ has been treated as a "bang-sparse-bang" control problem in [1, Section 4]. In the context of time-dependent systems, controls taken pointwise in time from a discrete set of states are referred to as switching controls and have been treated in the literature mainly with respect to feedback control for ordinary differential equations and exact controllability. Regarding the former we refer to [9–11], where feedback controls and compensators are constructed that switch between a discrete set of gain operators; typically with the goal of stability of the closed loop system. In [12,13], controllability of ordinary differential equations and of the heat equation is analyzed for control actuators with switching structure.

Problem (1.1) is challenging since the penalty term is neither convex nor lower semicontinuous. We thus cannot apply the standard approach in optimal control, which consists in arguing existence of a solution via limits of a minimizing sequence and deriving necessary optimality conditions using separation theorems from convex analysis. Recall that for Fréchet-differentiable $\mathcal{F}$ and convex $\mathcal{G}$, a minimizer $\bar{u}$ of

$$\min_u \mathcal{F}(u) + \mathcal{G}(u) \tag{1.2}$$

satisfies the following necessary optimality conditions: there exists a $\bar{p} = -\mathcal{F}'(\bar{u})$ such that $\bar{p} \in \partial\mathcal{G}(\bar{u})$, which holds if and only if $\bar{u} \in \partial\mathcal{G}^*(\bar{p})$; see, e.g., [14, Proposition 4.4.4]. Here, $\mathcal{G}^*$ denotes the Fenchel conjugate of the convex functional $\mathcal{G}$, and $\partial\mathcal{G}^*$ denotes its convex subdifferential. We thus obtain the primal-dual optimality system

$$\begin{cases} -\bar{p} = \mathcal{F}'(\bar{u}), \\ \bar{u} \in \partial\mathcal{G}^*(\bar{p}). \end{cases} \tag{1.3}$$

Note that since Fenchel conjugates are always convex, this system is well-defined even for nonconvex $\mathcal{G}$, although one cannot derive it as a necessary optimality condition for minimizers of (1.2).[1] We thus follow the approach from [1], in that we show existence of a solution to (1.3) and verify that (under some conditions) it is a minimizer of (1.1). This approach is based on deriving an explicit, pointwise characterization of the subdifferential $\partial\mathcal{G}^*$, which also yields that under some assumptions on α, β, A and z, the solution will attain the values $u_1, \dots, u_d$ almost everywhere (i.e., it satisfies a *generalized multi-bang principle*). This characterization is also instrumental for the numerical solution of (1.3) using a semismooth Newton method.

This paper is organized as follows. The next section is concerned with the formal optimality system (1.3), where an explicit form is derived in Section 2.1, existence and stability of a unique solution is shown in Section 2.2, and the structure of the resulting controls – in particular, conditions for a generalized multi-bang principle – is investigated in Section 2.3. Suboptimality of controls is characterized in Section 3, and conditions for optimality are given. Section 4 addresses the computation of solutions by introducing a regularization of (1.3) for which a semismooth Newton method is applicable. Finally, Section 5 illustrates the structure of multi-bang controls with numerical examples.

2. Formal optimality system

In this section we consider the system (1.3) with

[1] This "formal convex analysis" approach should be compared to the formal Lagrangian approach for deriving explicit optimality conditions in optimal control of partial differential equations.

$$\mathcal{F} : L^2(\Omega) \to \mathbb{R}, \qquad u \mapsto \frac{1}{2}\|A^{-1}u - z\|_{L^2}^2,$$
$$\mathcal{G} : L^2(\Omega) \to \overline{\mathbb{R}}, \qquad u \mapsto \int_\Omega \left(\frac{\alpha}{2}|u(x)|^2 + \beta \prod_{i=1}^d |u(x) - u_i|_0\right) dx + \delta_U(u),$$

where δ_U is the indicator function of the admissible set

$$U := \{u \in L^2(\Omega) \colon u_1 \leqslant u(x) \leqslant u_d \text{ for almost every } x \in \Omega\}.$$

2.1. Fenchel conjugate and subdifferential

We begin by computing the Fenchel conjugate $\mathcal{G}^*$ of $\mathcal{G}$, which is defined as

$$\mathcal{G}^*(p) = \sup_{u \in L^2(\Omega)} \langle u, p\rangle_{L^2} - \mathcal{G}(u),$$

where $\langle\cdot,\cdot\rangle_{L^2}$ denotes the inner product in $L^2(\Omega)$. Since $\mathcal{G}$ is the integral of the function

$$g : \mathbb{R} \to \overline{\mathbb{R}}, \qquad v \mapsto \frac{\alpha}{2}v^2 + \beta \prod_{i=1}^d |v - u_i|_0 + \delta_{[u_1,u_d]}(v),$$

the Fenchel conjugate can be computed pointwise as well; see, e.g., [15, Proposition IV.1.2]. Hence,

$$\mathcal{G}^*(p) = \int_\Omega g^*(p(x))\, dx,$$

where

$$g^* : \mathbb{R} \to \overline{\mathbb{R}}, \qquad q \mapsto \sup_v vq - g(v), \tag{2.1}$$

is the Fenchel conjugate of g. To compute $g^*(q)$, we assume that the supremum in (2.1) for given q is attained at $\bar v$. Then we discriminate the following cases:

(i) $\bar v = u_i$ for an $i \in \{1, \ldots, d\}$. Then,

$$g(\bar v) = \frac{\alpha}{2}u_i^2,$$

and hence

$$g^*(q) = qu_i - \frac{\alpha}{2}u_i^2.$$

(ii) $\bar v \neq u_i$ for any $i \in \{1, \ldots, d\}$. Then, $v \mapsto |v - u_i|_0 \equiv 1$ is differentiable at $\bar v$ and hence the supremum in (2.1) is attained if the necessary condition

$$\bar v - \frac{1}{\alpha}q = 0$$

is satisfied. Hence in this case,

$$g^*(q) = \frac{1}{2\alpha}q^2 - \beta.$$

It remains to decide which of these cases is attained based on the value of q. For this purpose, it will be convenient to define the functions

$$g_i^*(q) = \begin{cases} qu_i - \frac{\alpha}{2}u_i^2 & \text{if } 1 \leqslant i \leqslant d, \\ \frac{1}{2\alpha}q^2 - \beta & \text{if } i = 0. \end{cases}$$

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

Firstly, case (i) together with the box constraints on v imply that for $q < \alpha u_1$, the supremum is attained at $\bar v = u_1$, and similarly, for $q > \alpha u_d$, at $\bar v = u_d$. Hence we have that

$$g^*(q) = \begin{cases} g_1^*(q) & \text{if } q < \alpha u_1, \\ g_d^*(q) & \text{if } q > \alpha u_d. \end{cases}$$

In fact, for $q < \alpha u_1$ we have $\bar v \notin \{u_1, \dots, u_d\}$ since otherwise $\alpha \bar v = q < \alpha u_1$ and hence $\bar v < u_1$, which is impossible. Hence $\bar v \in \{u_1, \dots, u_d\}$. Since for all $j \in \{2, \dots, d\}$ we have

$$\left(\frac{\alpha(u_j + u_1)}{2} - q\right)(u_j - u_1) > (\alpha u_1 - q)(u_j - u_1) > 0,$$

it follows that

$$g_1^*(q) = q u_1 - \frac{\alpha}{2} u_1^2 > q u_j - \frac{\alpha}{2} u_j^2 = g_j^*(q).$$

We find that $g^*(q) = g_1^*(q)$ if $q < \alpha u_1$. The case $q > \alpha u_d$ is argued analogously.

We turn to the case $q \in [\alpha u_1, \alpha u_d]$. Then the pointwise supremum in (2.1) is attained at

$$g^*(q) = \max\{g_0^*(q), g_1^*(q), \dots, g_d^*(q)\}.$$

From the definition of the g_i^*, we have that $g_0^*(q) > g_j^*(q)$ for all $j \in \{1, \dots, d\}$ if and only if

$$\frac{1}{2\alpha}(q - \alpha u_j)^2 > \beta \quad \text{for all } j \in \{1, \dots, d\}. \tag{2.2}$$

Hence, for all $q \in [\alpha u_1, \alpha u_d]$ satisfying (2.2), we have that $g^*(q) = g_0^*(q)$. Next consider some q for which

$$\frac{1}{2\alpha}(q - \alpha u_j)^2 < \beta \quad \text{for some } j \in \{1, \dots, d\},$$

and determine i such that $g_i^*(q) > g_j^*(q)$ for all $j \neq i$, $j > 0$. This is the case if and only if

$$(u_i - u_j) q > \frac{\alpha}{2}\left(u_i^2 - u_j^2\right),$$

and since the u_j are distinct and ordered, this is equivalent to

$$q \gtrless \frac{\alpha}{2}(u_i + u_j) \quad \text{if } u_i \gtrless u_j.$$

This can be written explicitly as

$$\begin{aligned} &\alpha u_1 \leqslant q < \frac{\alpha}{2}(u_1 + u_2) && \text{if } i = 1, \\ &\frac{\alpha}{2}(u_i + u_{i-1}) < q < \frac{\alpha}{2}(u_i + u_{i+1}) && \text{if } 1 < i < d, \\ &\frac{\alpha}{2}(u_{d-1} + u_d) < q \leqslant \alpha u_d && \text{if } i = d. \end{aligned}$$

Making use of the fact that the u_i are ordered, we introduce the sets

$$P_i := \begin{cases} \{q\colon |q - \alpha u_j| > \sqrt{2\alpha\beta} \text{ for all } j \in \{1, \dots, d\} \text{ and } \alpha u_1 < q < \alpha u_d\} & \text{if } i = 0, \\ \{q\colon q - \alpha u_1 < \sqrt{2\alpha\beta} \text{ and } q < \frac{\alpha}{2}(u_1 + u_2)\} & \text{if } i = 1, \\ \{q\colon q - \alpha u_i| < \sqrt{2\alpha\beta} \text{ and } \frac{\alpha}{2}(u_{i-1} + u_i) < q < \frac{\alpha}{2}(u_i + u_{i+1})\} & \text{if } 1 < i < d, \\ \{q\colon q - \alpha u_d > \sqrt{2\alpha\beta} \text{ and } \frac{\alpha}{2}(u_d + u_{d-1}) < q\} & \text{if } i = d. \end{cases}$$

Summarizing the above calculation and using this definition, we then have

$$g^*(q) = \begin{cases} q u_i - \frac{\alpha}{2} u_i^2 & \text{if } q \in \overline{P}_i,\ 1 \leqslant i \leqslant d, \\ \frac{1}{2\alpha} q^2 - \beta & \text{if } q \in \overline{P}_0. \end{cases}$$

Note that the P_i are pairwise disjoint and that $\bigcup_{i=0}^d \overline{P}_i = \mathbb{R}$. Furthermore, $g_i^*(q) = g_j^*(q)$ for $q \in \overline{P}_i \cap \overline{P}_j$, since in this case equality must hold in place of the corresponding inequalities. This implies that g^* is well-defined, and in particular that g^* is (locally Lipschitz) continuous.

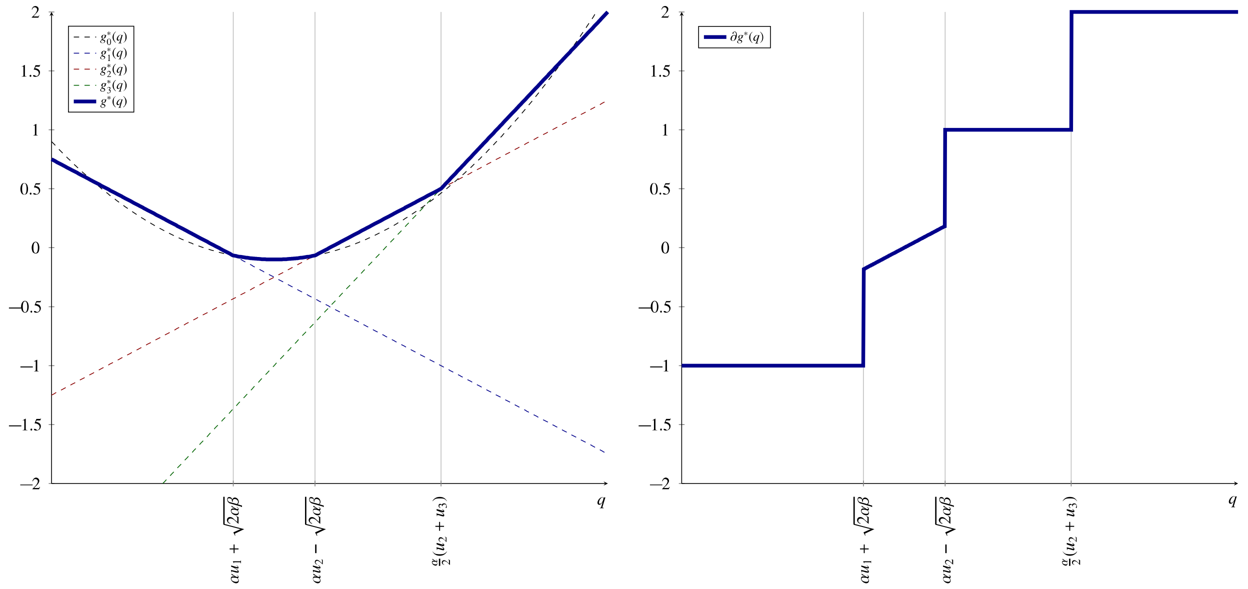


Fig. 1. Plot of $g^*(q)$ (left) and $\partial g^*(q)$ (right) for $d = 3$, $(u_1, u_2, u_3) = (-1, 1, 2)$, $\alpha = 0.5$, $\beta = 0.1$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

We also note that while the sets $\{P_i\}_{i=1}^d$ are open intervals ordered along $\mathbb{R}$, the set P_0 consists of several connected components of the form

$$(\sqrt{2\alpha\beta} + u_i, u_{i+1} - \sqrt{2\alpha\beta}) = \left(\frac{\alpha(u_{i+1} + u_i)}{2} - \rho_i, \frac{\alpha(u_{i+1} + u_i)}{2} + \rho_i\right)$$

if

$$\rho_i := \frac{\alpha(u_{i+1} - u_i)}{2} - \sqrt{2\alpha\beta} > 0$$

for $i \in \{1, \dots, d-1\}$. If $\rho_i \leqslant 0$ for all $i \in \{1, \dots, d-1\}$, the set P_0 is empty.

Since the Fenchel conjugate g^* is locally Lipschitz continuous, its convex subdifferential coincides with Clarke's generalized gradient [14, Proposition 7.3.9] and hence is given by

$$\begin{aligned}\partial g^*(q) &= \overline{\mathrm{co}}\Bigg(\bigcup_{\{i:\ g^*(q) = g_i^*(q)\}} \big\{(g_i^*)'(q)\big\}\Bigg)\\ &= \begin{cases} \{u_i\} & \text{if } q \in P_i,\ 1 \leqslant i < d,\\ \{\frac{1}{\alpha}q\} & \text{if } q \in P_0,\\ [u_i, u_{i+1}] & \text{if } q \in \overline{P}_i \cap \overline{P}_{i+1},\ 1 \leqslant i < d,\\ \big[\min\{u_i, \frac{1}{\alpha}q\}, \max\{u_i, \frac{1}{\alpha}q\}\big] & \text{if } q \in \overline{P}_i \cap \overline{P}_0,\ 1 \leqslant i \leqslant d, \end{cases}\end{aligned} \tag{2.3}$$

where $\overline{\mathrm{co}}$ denotes the closed convex hull. Fig. 1 gives an example of $g^*(q)$ and $\partial g^*(q)$, where P_0 only consists of a single connected component between P_1 and P_2.

2.2. Existence and stability

We now verify existence of a solution $(\bar{u}, \bar{p})$ to the formal optimality system (1.3), which can be written as

$$\begin{cases} -\bar{p} = A^{-*}\big(A^{-1}\bar{u} - z\big),\\ \bar{u} \in \partial\mathcal{G}^*(\bar{p}), \end{cases} \tag{2.4}$$

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

where using (2.3), the subdifferential of the convex function $\mathcal{G}^*$ is defined pointwise almost everywhere by

$$\partial\mathcal{G}^*(p)(x) = \partial g^*\big(p(x)\big) = \begin{cases} \{u_i\} & \text{if } p(x) \in P_i, \\ \{\frac{1}{\alpha}p(x)\} & \text{if } p(x) \in P_0, \\ [u_i, u_{i+1}] & \text{if } p(x) \in \overline{P}_i \cap \overline{P}_{i+1}, \\ \big[\min\big(u_i, \frac{1}{\alpha}p(x)\big), \max\big\{u_i, \frac{1}{\alpha}p(x)\big\}\big] & \text{if } p(x) \in \overline{P}_i \cap \overline{P}_0. \end{cases}$$

Theorem 2.1. *There exists a unique solution $(\bar u, \bar p) \in L^2(\Omega) \times V$ to (2.4).*

Proof. We introduce $\bar y := z - A^{-1}\bar u \in L^2(\Omega)$ and eliminate $\bar u$ and $\bar p$ from (2.4) to obtain the reduced optimality condition

$$z \in \bar y + A^{-1}\partial\mathcal{G}^*\big(A^{-*}\bar y\big). \tag{2.5}$$

Since $\partial\mathcal{G}^*$ is the subdifferential of a convex function, it is maximal monotone from $L^2(\Omega)$ to $L^2(\Omega)$; see [16, Theorem 20.40]. Furthermore, A and thus A^* are isomorphisms by assumption, and hence we have that A^{-*} is a bounded operator with $\operatorname{ran}(A^{-*}) = V \hookrightarrow L^2(\Omega)$. From $\operatorname{dom}(\partial\mathcal{G}^*) = L^2(\Omega)$, it follows that

$$\bigcup_{\lambda>0} \lambda\big(\operatorname{ran}\big(A^{-*}\big) - \operatorname{dom}\big(\partial\mathcal{G}^*\big)\big) = L^2(\Omega)$$

is closed. This implies that $A^{-1}\partial\mathcal{G}^*(A^{-*}\cdot)$ is maximal monotone; see [16, Theorem 24.5]. Since the identity is clearly maximal monotone with domain $L^2(\Omega)$, we have that $B := I + A^{-1}\partial\mathcal{G}^*(A^{-*}\cdot)$ is maximal monotone from $L^2(\Omega)$ to $L^2(\Omega)$ as well; see [16, Corollary 24.4].

Now by convexity of $\mathcal{G}^*$, we have $\langle\partial\mathcal{G}^*(v) - \partial\mathcal{G}^*(0), v\rangle_{L^2} \geqslant 0$ for all $v \in L^2(\Omega)$, and hence

$$\begin{aligned} \big\langle y + A^{-1}\partial\mathcal{G}^*\big(A^{-*}y\big), y\big\rangle_{L^2} &= \|y\|_{L^2}^2 + \big\langle\partial\mathcal{G}^*\big(A^{-*}y\big), A^{-*}y\big\rangle_{L^2} \\ &\geqslant \|y\|_{L^2}^2 + \big\langle A^{-1}\partial\mathcal{G}^*(0), y\big\rangle_{L^2} \\ &\to \infty \end{aligned}$$

as $\|y\|_{L^2} \to \infty$. Hence B is coercive and maximal monotone on $L^2(\Omega)$ and thus surjective on $L^2(\Omega)$; see [16, Corollary 21.2]. From this, we obtain that for any $z \in L^2(\Omega)$ there exists a $\bar y \in L^2(\Omega)$ satisfying (2.5). Furthermore, for any such $\bar y$ we have

$$z - \bar y \in A^{-1}\big(\partial\mathcal{G}^*\big(A^{-*}\bar y\big)\big),$$

and hence $z - \bar y \in \operatorname{ran}(A^{-1}) = \operatorname{dom}(A)$. We can thus set

$$\begin{aligned} &\bar u := A(z - \bar y) \in \partial\mathcal{G}^*(\bar p) \subset L^2(\Omega), \\ &\bar p := A^{-*}\bar y = A^{-*}\big(z - A^{-1}\bar u\big) \in V, \end{aligned}$$

and obtain the desired solution of (2.4).

To show uniqueness, assume that $\bar y_1, \bar y_2 \in L^2(\Omega)$ are two solutions. By inserting both into (2.5) and subtracting, we obtain

$$\begin{aligned} 0 &= \big\langle \bar y_1 - \bar y_2 + A^{-1}\partial\mathcal{G}^*\big(A^{-*}\bar y_1\big) - A^{-1}\partial\mathcal{G}^*\big(A^{-*}\bar y_2\big), \bar y_1 - \bar y_2\big\rangle_{L^2} \\ &= \|\bar y_1 - \bar y_2\|_{L^2}^2 + \big\langle\partial\mathcal{G}^*\big(A^{-*}\bar y_1\big) - \partial\mathcal{G}^*\big(A^{-*}\bar y_2\big), A^{-*}\bar y_1 - A^{-*}\bar y_2\big\rangle_{L^2} \\ &\geqslant \|\bar y_1 - \bar y_2\|_{L^2}^2 \end{aligned}$$

by monotonicity of $\partial\mathcal{G}^*$, and hence that $\bar y_1 = \bar y_2$. Next, let $(\bar u, \bar p) \in L^2(\Omega) \times V$ solve (2.4). Then, $\bar y := z - A^{-1}\bar u = A^*\bar p \in L^2(\Omega)$ satisfies (2.5). Since this solution is unique and A is an isomorphism, the pair $(\bar u, \bar p)$ must be unique as well. □

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

For later reference, we recall that if $(\bar u, \bar p)$ satisfies $\bar u \in \partial\mathcal{G}^*(\bar p)$, we have that pointwise almost everywhere

$$\bar u(x) \in \begin{cases} \{u_1\} & \text{if } \bar p(x) < \min\{\alpha u_1 + \sqrt{2\alpha\beta}, \frac{\alpha}{2}(u_1+u_2)\}, \\ \{u_i\} & \text{if } \max\{\alpha u_i - \sqrt{2\alpha\beta}, \frac{\alpha}{2}(u_{i-1}+u_i)\} < \bar p(x) \\ & \qquad < \min\{\alpha u_{i+1} + \sqrt{2\alpha\beta}, \frac{\alpha}{2}(u_i+u_{i+1})\},\ 1<i<d, \\ \{u_d\} & \text{if } \bar p(x) > \max\{\alpha u_d - \sqrt{2\alpha\beta}, \frac{\alpha}{2}(u_{d-1}+u_d)\}, \\ \{\frac{1}{\alpha}\bar p(x)\} & \text{if } \bar p(x) \in (\alpha u_1, \alpha u_d) \text{ and } |\bar p(x) - \alpha u_j| > \sqrt{2\alpha\beta} \text{ for all } 1 \leqslant j \leqslant d, \\ [u_i, u_{i+1}] & \text{if } \bar p(x) = \frac{\alpha}{2}(u_i + u_{i+1}),\ 1 \leqslant i < d, \\ [\frac{1}{\alpha}\bar p(x), u_i] & \text{if } \bar p(x) = \alpha u_i - \sqrt{2\alpha\beta},\ 1 < i \leqslant d, \\ [u_i, \frac{1}{\alpha}\bar p(x)] & \text{if } \bar p(x) = \alpha u_i + \sqrt{2\alpha\beta},\ 1 \leqslant i < d. \end{cases} \tag{2.6}$$

Continuous dependence of the solution $(\bar u, \bar y, \bar p)$ on the data z is considered next.

Proposition 2.2. *Let $(\bar u_z, \bar y_z, \bar p_z)$ denote the solution to (2.4) for given $z \in L^2(\Omega)$. Then, the following statements hold.*

(a) *There exists a constant $K > 0$ such that*

$$\|\bar y_{z_1} - \bar y_{z_2}\|_{L^2} + \|\bar p_{z_1} - \bar p_{z_2}\|_V \leqslant K\|z_1 - z_2\|_{L^2},$$

for all $z_1 \in L^2(\Omega)$, $z_2 \in L^2(\Omega)$.

(b) *For $z_n \to z$ in $L^2(\Omega)$ we have that $(u_{z_n}, y_{z_n}) \rightharpoonup (u_z, y_z)$ weakly in $V^* \times V$ and $p_{z_n} \to p_z$ strongly in V.*

(c) *For $z \in L^2(\Omega)$, let S be a compact subset of $\bigcup_{i=0}^d P_i$. If A is an isomorphism from $H^2(\Omega) \cap H_0^1(\Omega)$ to $L^2(\Omega)$ and $\Omega \subset \mathbb{R}^n$, $n \leqslant 3$, then there exist a neighborhood $U(z)$ in $L^2(\Omega)$ and a constant K_S such that*

$$\|u_{\tilde z} - u_z\|_{H^2(\Omega_S)} \leqslant K_S\|\tilde z - z\|_{L^2} \quad \textit{for all } \tilde z \in U(z),$$

where $\Omega_S = \{x\colon\ \bar p(x) \in S\}$.

Proof.

(a) Proceeding as in the second part of the proof of Theorem 2.1 we find that

$$\langle z_1 - z_2, \bar y_{z_1} - \bar y_{z_2}\rangle_{L^2} \geqslant \|\bar y_{z_1} - \bar y_{z_2}\|_{L^2},$$

and hence that $\|\bar y_{z_1} - \bar y_{z_2}\|_{L^2} \leqslant \|z_1 - z_2\|_{L^2}^2$. The estimate now follows using the first equation of (2.4).

(b) To simplify notation, let $(u_n, y_n, p_n) = (\bar u_{z_n}, \bar y_{z_n}, \bar p_{z_n})$. By (a) we have that $\{p_n\}_{n\in\mathbb{N}}$ is bounded in V, and hence by the compact embedding $V \hookrightarrow L^2(\Omega)$, there exists a subsequence (denoted by the same symbol) such that $p_n \to \bar p := \bar p_z$ almost everywhere in Ω; see, e.g., [17, Theorem 4.3]. Furthermore, the box constraints on u_n imply that $\{u_n\}_{n\in\mathbb{N}}$ is bounded in $L^2(\Omega)$ and hence that there exists a subsequence of $\{u_n\}_{n\in\mathbb{N}}$ (also denoted by the same symbol) such that $u_n \rightharpoonup \tilde u$ weakly in $L^2(\Omega)$ for some $\tilde u \in L^2(\Omega)$. From (2.6), we deduce that

$$u_n \to \bar u \quad \text{almost everywhere on } \left\{x \in \Omega\colon\ \bar p(x) \in \bigcup_{i=0}^d P_i\right\},$$

where we use that the sets P_i are open and $p_n \to \bar p$ almost everywhere. Since $u_n \rightharpoonup \tilde u$ and $u_n = \frac{1}{\alpha}p_n$ on P_0, with $p_n \to \bar p$ in $L^2(\Omega)$, we conclude that $u_n \to \tilde u = \bar u$ almost everywhere on $\{x \in \Omega\colon\ \bar p(x) \in \bigcup_{i=0}^d P_i\}$.
We still need to consider the sets $S_{i,i+1} = \{x \in \Omega\colon\ \bar p(x) \in \overline{P}_i \cap \overline{P}_{i+1}\}$ for $i \in \{1, \dots, d-1\}$ and $S_{i,0} = \{x \in \Omega\colon\ \bar p(x) \in \overline{P}_i \cap \overline{P}_0\}$ for $i \in \{1, \dots, d\}$; note that these sets can be empty. Since $u_n \rightharpoonup \tilde u$ weakly in $L^2(\Omega)$, from Mazur's theorem (see, e.g., [18, Theorem V.1.2]) we obtain existence of a convex combination of u_n that converges strongly in $L^2(\Omega)$ to a $\tilde u \in L^2(\Omega)$. Specifically, there exist coefficients $\{\gamma_k^n\}_{k,n\in\mathbb{N}}$ and summation bounds $\{l_n\}_{n\in\mathbb{N}}$ with $\sum_{k=1}^{l(n)} \gamma_k^n = 1$, and indices $n_k \in \{n, n+1, \dots\}$ such that

$$\tilde u_n := \sum_{k=1}^{l_n} \gamma_k^n u_{n_k} \to \tilde u \quad \text{strongly in } L^2(\Omega),$$

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

see [17, Exercise 3.4]. Taking another subsequence, we have $\tilde u_n \to \tilde u$ almost everywhere in $L^2(\Omega)$. To argue that $\tilde u \in \partial\mathcal{G}^*(\bar p)$ on $\bigcup_{i=1}^{d-1} S_{i,i+1} \cup \bigcup_{i=1}^{d} S_{i,0}$, we first consider

$$S_{1,0} = \left\{x \in \Omega:\ \bar p(x) \in \overline{P}_1 \cap \overline{P}_0\right\} = \left\{x \in \overline{\Omega}:\ \bar p(x) = \alpha u_1 + \sqrt{2\alpha\beta}\right\}$$

(in case it is nonempty). Recall that $p_n(x) \to \bar p(x)$ for almost every $x \in \Omega$. Let $x \in S_{1,0}$ be such that $p_n(x) \to \bar p(x)$ and $\tilde u_n(x) \to \tilde u(x)$. Then $u_n(x) \in [u_1, \frac{1}{\alpha} p_n(x)]$ for all n sufficiently large. Consequently,

$$\tilde u_n(x) = \sum_{k=1}^{l_n} \gamma_k^n u_{n_k}(x) \in \Big[u_1, \sup_{k \geqslant n} p_k(x)\Big].$$

Taking the limit as $n \to \infty$, we obtain that $\tilde u(x) \in [u_1, \frac{1}{\alpha}\bar p(x)] = [u_1, \alpha u_1 + \sqrt{2\alpha\beta}\,]$, and hence that $\tilde u(x) \in \partial\mathcal{G}^*(\bar p)(x)$ for almost every $x \in S_{d,0}$. Analogous arguments can be used for the sets $S_{i,i+1}$ and $S_{i,0}$ with $i \in \{2, \dots, d\}$ (if they are nonempty). Altogether we have that

$$\tilde u \in \partial\mathcal{G}^*(\bar p) \quad \text{almost everywhere on } \Omega.$$

Thus $(\tilde u, \bar y, \bar p)$ satisfies (2.4). Uniqueness of the solution to (2.4) implies that $\tilde u = \bar u$.

(c) Next consider the sets P_i associated to the solution $(\bar u_z, \bar y_z, \bar p_z)$ of (2.4). By the assumptions on Ω and A we have that $p_{\tilde z} \to p_z$ in $H^2(\Omega) \hookrightarrow C(\overline{\Omega})$ as $\tilde z \to z$ in $L^2(\Omega)$. Hence there exists a neighborhood U_z such that $\{x \in \Omega:\ \bar p_z(x) \in S \cap P_i\} \subset \{x \in \Omega:\ p_{\tilde z}(x) \in P_i\}$ for $i \in \{0, \dots, d\}$. Consequently,

$$\bar u_z = \bar u_{\tilde z} = u_i \quad \text{on } \left\{x \in \Omega:\ \bar p_z(x) \in S \cap P_i\right\} \text{ for } 1 \leqslant i \leqslant d, \text{ and}$$
$$\bar u_z - \bar u_{\tilde z} = \frac{1}{\alpha}(\bar p_z - \bar p_{\tilde z}) \quad \text{on } \left\{x \in \Omega:\ \bar p_z(x) \in S \cap P_0\right\}.$$

These equalities imply the claim. □

2.3. *Structure of solution*

We now discuss the structure of the solution $\bar u$ to (2.4), and in particular, conditions under which $\bar u$ only takes on the values $u_1, \dots, u_d$ almost everywhere. First, observe that

$$\begin{aligned}\Omega &= \bigcup_{i=1}^{d}\{x \in \Omega:\ \bar u(x) = u_i\} \cup \left\{x \in \Omega:\ \bar u(x) = \frac{1}{\alpha}\bar p(x) \text{ and } \bar u(x) \notin \{u_1, \dots, u_d\}\right\}\\ &\quad \cup \left\{x \in \Omega:\ \bar u(x) \notin \left\{u_1, \dots, u_d, \frac{1}{\alpha}\bar p(x)\right\}\right\}\\ &=: \bigcup_{i=1}^{d}\mathcal{A}_i \cup \mathcal{F} \cup \mathcal{S}.\end{aligned}$$

In analogy to bang-bang control problems, we refer to $\mathcal{A} := \bigcup_{i=1}^{d}\mathcal{A}_i$ as the *multi-bang arc*, to $\mathcal{F}$ as the *free arc*, and to $\mathcal{S}$ as the *singular arc*.

Consider first the free arc. From (2.6), we can deduce that

$$\mathcal{F} \subset \left\{x \in \Omega:\ \alpha u_i + \sqrt{2\alpha\beta} < \bar p(x) < \alpha u_{i+1} - \sqrt{2\alpha\beta} \text{ for all } 1 \leqslant i < d\right\} \cap (\alpha u_1, \alpha u_d).$$

Hence, if α and β are chosen such that

$$\sqrt{2\beta/\alpha} \geqslant \frac{1}{2}(u_{i+1} - u_i) \quad \text{for all } 1 \leqslant i < d, \tag{2.7}$$

then $P_0 = \emptyset$, and thus there is no free arc. Note that as long as (2.7) is satisfied, the value of β does not appear in (2.6), and thus it has no further influence on the structure of $\bar u$. Rather than the values themselves, it is therefore the relations between β and α and between α and $u_{i+1} - u_i$, $1 \leqslant i < d$, that determine the structural properties of $\bar u$.

Similarly, we have

$$\begin{aligned}\mathcal{S} &= \left\{x \in \Omega\colon\ \bar p(x) \in \bigcup_{i=1}^{d-1}(\overline{P}_i \cap \overline{P}_{i+1}) \cup \bigcup_{i=1}^{d}(\overline{P}_i \cap \overline{P}_0)\right\}\\ &\subset \left\{x \in \Omega\colon\ \bar p(x) \in \bigcup_{i=1}^{d-1}\left\{\frac{\alpha}{2}(u_i + u_{i+1})\right\} \cup \bigcup_{i=1}^{d}\{\alpha u_i - \sqrt{2\alpha\beta}, \alpha u_i + \sqrt{2\alpha\beta}\}\right\}. \qquad (2.8)\end{aligned}$$

From this representation it is possible to derive *generalized multi-bang principles* for $\bar u$. A specific case is given in the following result, where we assume that A is a second order elliptic partial differential operator of the form

$$Ay = -\sum_{i,j=1}^{n} \partial_{x_i}(a_{i,j}\partial_{x_j} y) + \sum_{i=1}^{n} \partial_{x_i}(b_i y)$$

with $a_{i,j} \in W^{1,\infty}(\Omega)$ and $b_i \in L^\infty(\Omega)$.

Proposition 2.3. *If α and β are chosen according to (2.7), and A satisfies $A^{-*}(L^2(\Omega)) \subset W^{2,1}(\Omega)$ in addition to the assumptions above, then*

$$\Omega = \bigcup_{i=1}^{d}\{x \in \Omega\colon\ \bar u(x) = u_i\} \cup \{x \in \Omega\colon\ \bar y(x) = z(x)\}. \qquad (2.9)$$

Proof. The choice of α and β ensures that there is no free arc. Assume now that the singular arc $\mathcal{S}$ has positive Lebesgue measure (otherwise we are finished). To show that $\bar y = z$ almost everywhere on $\mathcal{S}$, we make use of the following result from [19, Lemma II.A.4]: Let $w \in W^{1,p}(\Omega)$ for some $1 \leqslant p \leqslant \infty$. Then $\nabla w = 0$ almost everywhere on $\{x \in \Omega\colon\ w(x) = t\}$ for every $t \in \mathbb{R}$. Applying this result for $w = \bar p$ and $\mathcal{S}_1 := \{x \in \Omega\colon\ \bar p(x) = \frac{\alpha}{2}(u_1 + u_2)\}$, we deduce that $\nabla \bar p = 0$ almost everywhere on $\mathcal{S}_1$. Utilizing this fact, the assumed regularity $\bar p \in W^{2,1}(\Omega)$ of the adjoint state, and the regularity of the coefficients, we can proceed as in [8, Theorem 5.2] to argue that $A^*\bar p = 0$ almost everywhere on $\mathcal{S}_1$. This argument can be repeated for all components of $\mathcal{S}$. Hence $\bar y - z = A^*\bar p = 0$ on $\mathcal{S}$, and the representation (2.9) follows. □

We refer to $\bar u$ satisfying (2.9) as a *generalized multi-bang control*. In particular, if α and β are chosen according to (2.7) and $\bar y(x) \neq z(x)$ for almost every $x \in \Omega$, $\bar u$ will only attain values among the desired control states $u_1, \dots, u_d$, and thus it will be a *true multi-bang control*. Furthermore, for fixed α, any choice of β satisfying (2.7) leads to the same control.

As noted in the Introduction, problem (1.1) for $d = 2$ represents a bang-bang control problem, since in this case, if the parameters are chosen to satisfy (2.7),

$$g^*(q) = \begin{cases} qu_1 & \text{if } q < \frac{\alpha}{2}(u_1 + u_2),\\ qu_2 & \text{if } q \geqslant \frac{\alpha}{2}(u_1 + u_2).\end{cases}$$

Under the assumptions of Proposition 2.3, $\bar u$ is a *generalized bang-bang control*, i.e., in the almost everywhere sense, it will only take on the values of the control constraints u_1 and u_2 except on singular arcs where $\bar y = z$. Compared to standard bang-bang problems, however, the solution $\bar u$ will be biased towards either u_1 or u_2 based on the value of α (unless $u_1 = -u_2$). By setting $\alpha = 0$ in (2.6) we (formally) recover the standard optimality conditions for bang-bang controls.

Remark 1. The quadratic term $\frac{\alpha}{2}|u(x)|^2$ should not be interpreted as a regularization of the binary term $|u(x) - u_i|_0$, but as an equally important part of the "multi-bang penalty" $\mathcal{G}$. Let us first turn our attention to the fact that the control constraints alone are sufficient to ensure existence of a solution to (2.4) even for $\alpha = 0$. Following the approach in this section, we find then that the supremum in the Fenchel conjugate (2.1) is attained at the maximal or minimal bound: For any $d \geqslant 2$, we have that

$$\tilde g^*(q) = \max\{qu_1, qu_2, \dots, qu_{d-1}, qu_d\} = \begin{cases} qu_1 & \text{if } q < 0,\\ qu_d & \text{if } q \geqslant 0.\end{cases}$$

Hence, under the assumptions of Proposition 2.3, $\bar{u}$ will always be a generalized bang-bang control, independent of the value of β and the choice of u_i, $1 < i < d$. The presence of the quadratic term in $\mathcal{G}$ is therefore essential for being able to select from all possible desired control states $u_1, \dots, u_d$. Furthermore, from (2.6) we see that smaller values of $\alpha > 0$ will allow the control to attain control states of larger magnitudes. In particular, if 0 is among the control states, the value of α controls the *sparsity* of the control $\bar{u}$.

We finish this section by remarking on some alternative formulations of the multi-bang penalty in (1.1).

(i) Instead of the product, we can penalize the sum of the binary terms, i.e., consider

$$\tilde{g}(v) = \frac{\alpha}{2}|v|^2 + \beta \sum_{i=1}^{d} |v - u_i|_0 + \delta_{[u_1,u_d]}(v).$$

We then obtain

$$\tilde{g}^*(q) = \begin{cases} qu_i - \frac{\alpha}{2}u_i^2 - (n-1)\beta & \text{if } q \in \overline{P}_i,\ 1 \leqslant i \leqslant d, \\ \frac{1}{2\alpha}q^2 - n\beta & \text{if } q \in \overline{P}_0, \end{cases}$$

i.e., the Fenchel conjugate is the same up to the additive constant $-(n-1)\beta$, and hence its subdifferential coincides with ∂g^*.

(ii) The L^2 penalty is sufficient to ensure existence of a solution to problem (1.1) even without control constraints. In this case, we obtain that $\tilde{g}^*(q) = g_0^*(q)$ if

$$\frac{1}{2\alpha}(q - \alpha u_j)^2 > \beta \quad \text{for all } j \in \{1, \dots, d\},$$

and hence

$$\widetilde{P}_0 = \bigl\{q\colon |q - \alpha u_j| > \sqrt{2\alpha\beta} \text{ for all } j \in \{1, \dots, d\}\bigr\}.$$

The free arc will thus always contain the two components $(-\infty, \alpha u_1 - \sqrt{2\alpha\beta})$ and $(\alpha u_d + \sqrt{2\alpha\beta}, \infty)$. In general, this will prevent the generalized multi-bang principle in Theorem 2.3 to hold even with the choice (2.7).

(iii) Finally, if the control constraints take the form $a \leqslant u(x) \leqslant b$ with $a < u_1$ and $b > u_d$, and $\alpha = 0$, a similar argument as above shows that

$$\tilde{g}^*(q) = \max\{qa - \beta, qu_1, \dots, qu_d, qb - \beta\} = \begin{cases} qa - \beta & \text{if } q < -\frac{\beta}{u_1 - a}, \\ qu_1 & \text{if } -\frac{\beta}{u_1 - a} \leqslant q < 0, \\ qu_d & \text{if } 0 \leqslant q < \frac{\beta}{b - u_d}, \\ qb - \beta & \text{if } \frac{\beta}{b - u_1} \leqslant q. \end{cases}$$

Although $\bar{u}$ may now take on the two bounds a and b besides the states u_1 and u_d, the multi-bang arc will not contain the control states $u_2, \dots, u_{d-1}$.

3. Duality gap and (sub)optimality

We now investigate in which cases the system (1.3) represents sufficient optimality conditions for problem (1.1). Let $(\bar{u}, \bar{p})$ satisfy (1.3). Since $\mathcal{F}$ is convex and Fréchet differentiable, the first relation of (1.3) implies that for any u,

$$\mathcal{F}(u) - \mathcal{F}(\bar{u}) - \langle -\bar{p}, u - \bar{u}\rangle \geqslant 0.$$

If we could similarly argue that

$$\mathcal{G}(u) - \mathcal{G}(\bar{u}) - \langle \bar{p}, u - \bar{u}\rangle \geqslant 0, \tag{3.1}$$

we would obtain that $\bar{u}$ is a minimizer of $J(u) := \mathcal{F}(u) + \mathcal{G}(u)$. For convex functionals, this inequality follows from the characterization of the subdifferential together with the Fenchel–Young inequality. However, since $\mathcal{G}$ is not convex, inequality (3.1) cannot hold in general and hence we can only expect that $(\bar u, \bar p)$ is suboptimal. This manifests itself in a duality gap in the Fenchel extremality relations, i.e., there is an $\varepsilon \geqslant 0$ such that

$$\mathcal{G}(\bar u) + \mathcal{G}^*(\bar p) \leqslant \langle \bar p, \bar u\rangle + \varepsilon,$$

which in our case arises due to the set-valued nature of the subdifferential $\partial\mathcal{G}^*$. To estimate the gap ε, we therefore introduce the *critical set*

$$\mathcal{C} := \bigl\{x \in \Omega\colon\ \bar p(x) \in \overline{P}_i \cap \overline{P}_j \text{ and } \bar u(x) \notin \bigl\{\bigl(g_i^*\bigr)'(x), \bigl(g_j^*\bigr)'(x)\bigr\} \text{ for some } i < j \in \{0, \dots, d\}\bigr\}.$$

The geometrical interpretation of the condition $x \in \mathcal{C}$ is that $\bar u(x)$ is not an extremal point of $[u_i, u_{i+1}]$ if $\bar p(x) \in \overline{P}_i \cap \overline{P}_{i+1}$ for some $i \in \{1, \dots, d-1\}$, respectively $\bar u(x)$ is not an extremal point of $[\min\{u_i, \frac{1}{\alpha}\bar p(x)\}, \max\{u_i, \frac{1}{\alpha}\bar p(x)\}]$ if $\bar p(x) \in \overline{P}_i \cap \overline{P}_0$ for some $i \in \{1, \dots, d\}$.

Lemma 3.1. *Let $(\bar u, \bar p)$ satisfy $\bar u \in \partial\mathcal{G}^*(\bar p)$, i.e., (2.6). Then we have*

$$\mathcal{G}(\bar u) + \mathcal{G}^*(\bar p) - \langle \bar p, \bar u\rangle \leqslant \beta|\mathcal{C}|,$$

where $|\mathcal{C}|$ denotes the Lebesgue measure of the critical set.

Proof. We discriminate pointwise based on the value of $\bar p(x)$ for almost every $x \in \Omega$.

(i) $\bar p(x) \in P_i$ for some $i \in \{1, \dots, d\}$. In this case, the relation (2.6) yields $\bar u(x) = u_i$ and thus

$$g\bigl(\bar u(x)\bigr) + g^*\bigl(\bar p(x)\bigr) - \bar p(x)\bar u(x) = \frac{\alpha}{2}u_i^2 + \bar p(x)u_i - \frac{\alpha}{2}u_i^2 - \bar p(x)u_i = 0.$$

(ii) $\bar p(x) \in P_0$ and thus $\bar u(x) = \frac{1}{\alpha}\bar p(x)$, yielding

$$g\bigl(\bar u(x)\bigr) + g^*\bigl(\bar p(x)\bigr) - \bar p(x)\bar u(x) = \frac{\alpha}{2}\left(\frac{1}{\alpha}\bar p(x)\right)^2 + \beta + \frac{1}{2\alpha}\bar p(x)^2 - \beta - \frac{1}{\alpha}\bar p(x)^2 = 0.$$

(iii) $\bar p(x) \in \overline{P}_i \cap \overline{P}_{i+1} = \{\frac{\alpha}{2}(u_i + u_{i+1})\}$ for some $i \in \{1, \dots, d-1\}$ and hence $\bar u(x) \in [u_i, u_{i+1}]$. Assume first that $u_i < \bar u(x) < u_{i+1}$. Then we have

$$\begin{aligned} g\bigl(\bar u(x)\bigr) + g^*\bigl(\bar p(x)\bigr) - \bar p(x)\bar u(x) &= \frac{\alpha}{2}\bar u(x)^2 + \beta + \frac{\alpha}{2}(u_i + u_{i+1})u_i - \frac{\alpha}{2}u_i^2 - \frac{\alpha}{2}(u_i + u_{i+1})\bar u(x) \\ &= \frac{\alpha}{2}\bigl(\bar u(x) - u_i\bigr)\bigl(\bar u(x) - u_{i+1}\bigr) + \beta \\ &< \beta \end{aligned}$$

for all $\bar u(x) \in (u_i, u_{i+1})$ since the first term is negative there.
For $\bar u(x) \in \{u_i, u_{i+1}\}$, we argue as in case (i) to find that

$$g\bigl(\bar u(x)\bigr) + g^*\bigl(\bar p(x)\bigr) - \bar p(x)\bar u(x) = 0.$$

(iv) $\bar p(x) \in \overline{P}_0 \cap \overline{P}_i$. This implies

$$\bar p(x) \in \begin{cases} \{\alpha u_1 + \sqrt{2\alpha\beta}\} & \text{if } i = 1, \\ \{\alpha u_i - \sqrt{2\alpha\beta}, \alpha u_i + \sqrt{2\alpha\beta}\} & \text{if } 1 < i < d, \\ \{\alpha u_d - \sqrt{2\alpha\beta}\} & \text{if } i = d. \end{cases}$$

Consider first the case $\bar p(x) = \alpha u_i + \sqrt{2\alpha\beta}$ for some $i \in \{1, \dots, d-1\}$, which implies that $\bar u(x) \in [u_i, \frac{1}{\alpha}\bar p(x)]$. Assume that $u_i < \bar u(x) < \frac{1}{\alpha}\bar p(x)$ (otherwise argue as in case (i) or (ii)). Then

$$g\bigl(\bar u(x)\bigr) + g^*\bigl(\bar p(x)\bigr) - \bar p(x)\bar u(x) = \frac{\alpha}{2}\bar u(x)^2 + \beta + \bar p(x)u_i - \frac{\alpha}{2}u_i^2 - \bar p(x)\bar u(x).$$

A simple calculus argument shows that the right-hand side is a monotonically decreasing function of $\bar u(x)$ on $(u_i, \frac{1}{\alpha}\bar p(x))$ and hence attains its supremum for $\bar u(x) = u_i$, which implies that

$$g\bigl(\bar u(x)\bigr) + g^*\bigl(\bar p(x)\bigr) - \bar p(x)\bar u(x) < \beta$$

for all $\bar u(x) \in (u_i, \frac{1}{\alpha}\bar p(x))$.

We argue similarly for $\bar p(x) = \alpha u_i - \sqrt{2\alpha\beta}$ for some $i \in \{2, \dots, d\}$.

Integrating over Ω yields the claim. □

Having computed the duality gap, we can now characterize (sub)optimality of solutions to the formal optimality system (1.3) by similar arguments as in the convex case.

Theorem 3.2. *Let $(\bar u, \bar p)$ satisfy (1.3). Then for any $u \in L^2(\Omega)$,*

$$J(\bar u) \leqslant J(u) + \beta|\mathcal{C}|.$$

In particular, if $\mathcal{C}$ is a set of Lebesgue measure zero, $\bar u$ is a solution to (1.1).

Proof. Assume that $(\bar u, \bar p)$ is a solution to (1.3) and let $u \in L^2(\Omega)$ be arbitrary. Recall that the first relation of (1.3) then implies that

$$\mathcal{F}(u) - \mathcal{F}(\bar u) - \langle -\bar p, u - \bar u\rangle \geqslant 0.$$

Furthermore, Lemma 3.1 and the Fenchel–Young inequality (which holds for any proper $\mathcal{G}$) imply that

$$\begin{aligned}\mathcal{G}(u) - \mathcal{G}(\bar u) - \langle \bar p, u - \bar u\rangle &\geqslant \mathcal{G}(u) - \langle \bar p, u\rangle + \mathcal{G}^*(\bar p) - \beta|\mathcal{C}| \\ &\geqslant -\beta|\mathcal{C}|.\end{aligned}$$

Hence,

$$\begin{aligned}J(u) - J(\bar u) &= \big(\mathcal{F}(u) + \mathcal{G}(u)\big) - \big(\mathcal{F}(\bar u) + \mathcal{G}(\bar u)\big) \\ &= \big(\mathcal{F}(u) - \mathcal{F}(\bar u) - \langle -\bar p, u - \bar u\rangle\big) + \big(\mathcal{G}(u) - \mathcal{G}(\bar u) - \langle \bar p, u - \bar u\rangle\big) \\ &\geqslant -\beta|\mathcal{C}|\end{aligned}$$

as claimed. □

Since the critical set $\mathcal{C}$ is contained in the singular arc $\mathcal{S}$, see (2.8), we immediately obtain the following optimality result.

Corollary 3.3. *If the solution $\bar u$ to (2.4) is a true multi-bang control (in particular, if $|\mathcal{S}| = 0$), then $\bar u$ is a solution to (1.1).*

4. Solution of optimality system

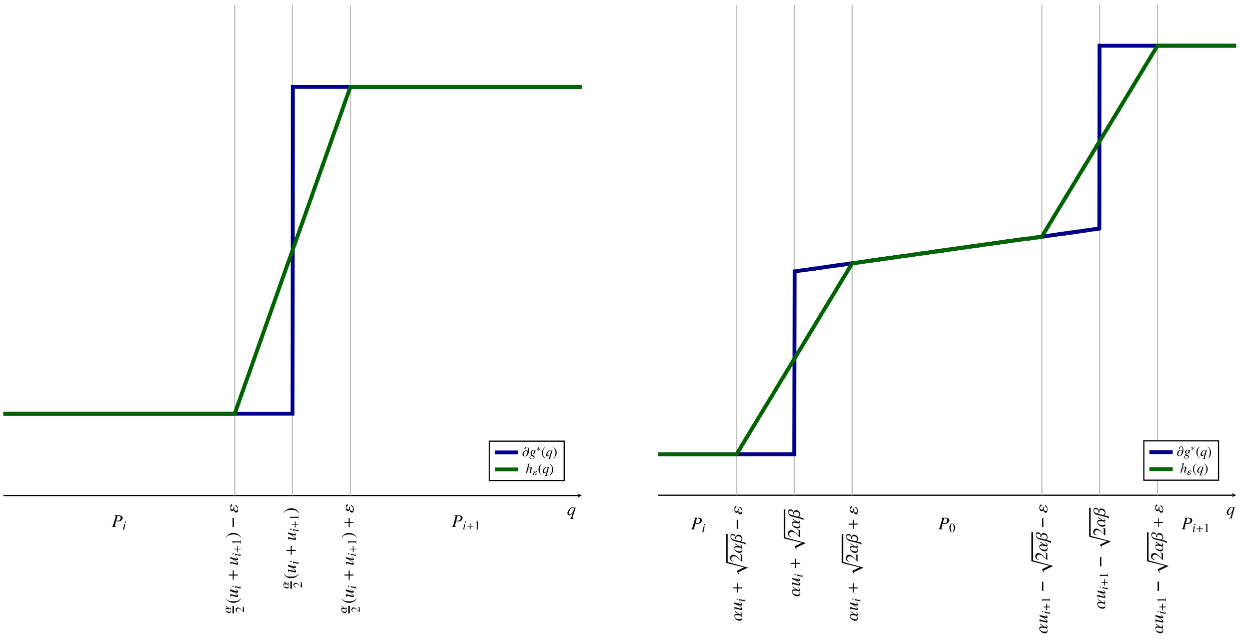
We now address the computation of solutions to the formal optimality system (1.3), which after introduction of the optimal state $\bar y := A^{-1}\bar u$ can be written as

$$\begin{cases} A\bar y = \bar u, \\ A^*\bar p = z - \bar y, \\ \bar u \in \partial\mathcal{G}^*(\bar p). \end{cases} \tag{4.1}$$

The main difficulty here lies in the set-valued nature of the subdifferential in (4.1). We therefore introduce a single-valued regularization of (4.1) for which a semismooth Newton method can be applied.

4.1. Regularization

We consider a continuous, piecewise linear regularization of $\partial\mathcal{G}^*$. Constructing this regularization is complicated by the possible presence of free arcs; see Section 2.3 and Fig. 2. We thus need to introduce the following sets:



(a) no free arc

(b) free arc

Fig. 2. Possible situations for $\partial g^*(q)$ and its regularization $h_\varepsilon(q)$. Left: $\overline{P}_i \cap \overline{P}_{i+1} \neq \emptyset$; right: $\overline{P}_i \cap \overline{P}_{i+1} = \emptyset$ and hence $\overline{P}_i \cap \overline{P}_{i,0+} \neq \emptyset$, $\overline{P}_{i+1} \cap \overline{P}_{i+1,0-} \neq \emptyset$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

$$P_1^\varepsilon = \left\{q\colon\ q \leqslant \alpha u_1 + \sqrt{2\alpha\beta} - \varepsilon \text{ and } q \leqslant \frac{\alpha}{2}(u_1+u_2) - \varepsilon\right\},$$

$$P_i^\varepsilon = \left\{q\colon\ |q-\alpha u_i| \leqslant \sqrt{2\alpha\beta} - \varepsilon \text{ and } \frac{\alpha}{2}(u_{i-1}+u_i) + \varepsilon \leqslant q \leqslant \frac{\alpha}{2}(u_i+u_{i+1}) - \varepsilon\right\},$$

$$P_d^\varepsilon = \left\{q\colon\ q \geqslant \alpha u_d - \sqrt{2\alpha\beta} + \varepsilon \text{ and } q \geqslant \frac{\alpha}{2}(u_{d-1}+u_d) + \varepsilon\right\},$$

$$P_0^\varepsilon = \left\{q\colon\ |q-\alpha u_j| \geqslant \sqrt{2\alpha\beta} + \varepsilon \ \text{ for all } \ j \in \{1,\dots,d\} \text{ and } q \in (\alpha u_1, \alpha u_d)\right\},$$

and if $\frac{\alpha}{2}(u_{i+1}-u_i) \leqslant \sqrt{2\alpha\beta}$ (i.e., there is no free arc),

$$P_{i,i+1}^\varepsilon = \left\{q\colon\ \left|q - \frac{\alpha}{2}(u_i+u_{i+1})\right| < \varepsilon\right\}, \qquad P_{i,0-}^\varepsilon = P_{i,0+}^\varepsilon = \emptyset,$$

else

$$P_{i,0-}^\varepsilon = \left\{q\colon\ \left|q - (\alpha u_i - \sqrt{2\alpha\beta})\right| < \varepsilon\right\},$$

$$P_{i,0+}^\varepsilon = \left\{q\colon\ \left|q - (\alpha u_i + \sqrt{2\alpha\beta})\right| < \varepsilon\right\}, \qquad P_{i,i+1}^\varepsilon = \emptyset.$$

In both cases, $P_{1,0-}^\varepsilon$ and $P_{d,0+}^\varepsilon$ are always defined as empty. To guarantee that the sets P_i^ε are well-defined, we need to assume that

$$\varepsilon < \min\left\{\sqrt{2\alpha\beta}, \frac{\alpha}{4}(u_{i+1}-u_{i-1})\right\}. \tag{4.2}$$

We then define

$$h_\varepsilon(q)=\begin{cases} u_i & \text{if } q\in P_i^\varepsilon,\\ \frac{1}{\alpha}q & \text{if } q\in P_0^\varepsilon,\\ \frac{1}{2\varepsilon}\big[u_i\big(\frac{\alpha}{2}(u_i+u_{i+1})+\varepsilon-q\big)+u_{i+1}\big(q-\frac{\alpha}{2}(u_i+u_{i+1})+\varepsilon\big)\big] & \text{if } q\in P_{i,i+1}^\varepsilon,\\ \frac{1}{2\varepsilon}\big[\frac{1}{\alpha}(\alpha u_i-\sqrt{2\alpha\beta}-\varepsilon)(\alpha u_i-\sqrt{2\alpha\beta}+\varepsilon-q) & \\ \quad+u_i(q-\alpha u_i+\sqrt{2\alpha\beta}+\varepsilon)\big] & \text{if } q\in P_{i,0-}^\varepsilon,\\ \frac{1}{2\varepsilon}\big[u_i(\alpha u_i+\sqrt{2\alpha\beta}+\varepsilon-q) & \\ \quad+\frac{1}{\alpha}(\alpha u_i+\sqrt{2\alpha\beta}+\varepsilon)(q-\alpha u_i-\sqrt{2\alpha\beta}+\varepsilon)\big] & \text{if } q\in P_{i,0+}^\varepsilon. \end{cases}$$

The regularization H_ε of the subdifferential $\partial\mathcal{G}^*$ is then defined pointwise almost everywhere as

$$H_\varepsilon(p)(x)=h_\varepsilon\big(p(x)\big).$$

Since $h_\varepsilon:\mathbb{R}\to\mathbb{R}$ is (by construction) a continuous and monotone function, the corresponding superposition operator $H_\varepsilon:L^2(\Omega)\to L^2(\Omega)$ is maximal monotone; see [20, Exemple 2.3.3]. The regularized system

$$\begin{cases} Ay_\varepsilon=u_\varepsilon,\\ A^*p_\varepsilon=z-y_\varepsilon,\\ u_\varepsilon=H_\varepsilon(p_\varepsilon), \end{cases} \tag{4.3}$$

thus has a unique solution $(u_\varepsilon,y_\varepsilon,p_\varepsilon)$ by the same arguments as in the proof of Theorem 2.1. Note that P_i^ε is strictly contained in P_i for all $i\in\{0,\dots,d\}$. Therefore, if $p_\varepsilon(x)\in P_i^\varepsilon$ for some $i\in\{0,\dots,d\}$ and almost all $x\in\Omega$, we obtain by comparing the definitions of H_ε and $\partial\mathcal{G}^*$ that

$$u_\varepsilon=H_\varepsilon(p_\varepsilon)\in\partial\mathcal{G}^*(p_\varepsilon),$$

and hence that $(u_\varepsilon,y_\varepsilon,p_\varepsilon)$ satisfies (4.1). Furthermore, there is no singular arc in this case, and so u_ε is a true multi-bang control and thus optimal. Otherwise, solutions to (4.3) converge to a solution to (4.1) in the following sense.

Theorem 4.1. *As $\varepsilon\to0$, the sequence $\{(u_\varepsilon,y_\varepsilon,p_\varepsilon)\}_{\varepsilon>0}$ converges weakly in $L^2(\Omega)\times V\times V$ to the solution $(\bar u,\bar y,\bar p)$ to (4.1). If the critical set $\mathcal{C}$ has Lebesgue measure zero, $u_\varepsilon\to\bar u$ pointwise almost everywhere.*

Proof. Using (4.3) and eliminating u_ε and p_ε, we have

$$y_\varepsilon+A^{-1}H_\varepsilon\big(A^{-*}y_\varepsilon\big)=z,$$

and hence

$$y_\varepsilon+A^{-1}H_\varepsilon\big(A^{-*}y_\varepsilon\big)-A^{-1}H_\varepsilon(0)=z-A^{-1}H_\varepsilon(0).$$

Since H_ε is maximal monotone for every $\varepsilon>0$, we obtain by taking the inner product with y_ε that

$$\begin{aligned} \|y_\varepsilon\|_{L^2}^2 &\leqslant \|z\|_{L^2}\|y_\varepsilon\|_{L^2}+\big\langle H_\varepsilon(0),A^{-*}y_\varepsilon\big\rangle_{L^2}\\ &\leqslant \|z\|_{L^2}\|y_\varepsilon\|_{L^2}+\big\|H_\varepsilon(0)\big\|_{L^2}\big\|A^{-*}y_\varepsilon\big\|_{L^2}\\ &\leqslant C\|y_\varepsilon\|_{L^2}, \end{aligned}$$

where we have used that A^{-*} is an isomorphism and that $H_\varepsilon(p)$ is bounded pointwise almost everywhere independently of ε by $\max\{|u_1|,|u_d|\}$. Together with (4.3), this implies that $\{(u_\varepsilon,y_\varepsilon,p_\varepsilon)\}_{\varepsilon>0}$ is bounded in $L^2(\Omega)\times V\times V$. Therefore, there exists a sequence $\{\varepsilon_n\}_{n\in\mathbb{N}}$ with $\lim_{n\to\infty}\varepsilon_n=0$ such that $(u_{\varepsilon_n},y_{\varepsilon_n},p_{\varepsilon_n})\rightharpoonup(\hat u,\hat y,\hat p)$ weakly in $L^2(\Omega)\times V\times V$. Since $V\hookrightarrow L^2(\Omega)$ compactly and after extracting a subsequence, we have in addition that $p_{\varepsilon_n}\to\hat p$ pointwise almost everywhere in Ω. We can thus pass to the limit in (the weak formulation of) the first two equations in (4.3) to obtain

$$A\hat y=\hat u,\qquad A^*\hat p=z-\hat y.$$

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

Arguing as in the proof of Proposition 2.2, we find coefficients $\{\gamma_k^n\}_{k,n\in\mathbb{N}}$ and summation bounds $\{l_n\}_{n\in\mathbb{N}}$ with $\sum_{k=1}^{l_n}\gamma_k^n = 1$, and indices $\varepsilon_{n_k}\in\{\varepsilon_n,\varepsilon_{n+1},\dots\}$ such that

$$\hat{u}_{\varepsilon_n} := \sum_{k=1}^{l_n}\gamma_k^n u_{\varepsilon_{n_k}} \to \hat{u} \quad \text{strongly in } L^2(\Omega).$$

Hence, after extracting a further subsequence, $\hat{u}_{\varepsilon_n}\to\hat{u}$ almost everywhere in Ω. We then discriminate pointwise almost everywhere for $\hat{p}$:

(i) $\hat{p}(x)\in P_i$ for some $i\in\{0,\dots,d\}$: Since the P_i are open sets, by definition of the regularization we have $h_{\varepsilon_n}(\hat{p}(x)) = (g_i^*)'(\hat{p}(x))$ for all n sufficiently large. Pointwise convergence of p_{ε_n} thus implies that $\lim_{n\to\infty} h_{\varepsilon_n}(p_{\varepsilon_n}(x)) = (g_i^*)'(\hat{p}(x))$, and since $u_\varepsilon(x) = h_\varepsilon(p_\varepsilon(x))$ we have

$$\lim_{n\to\infty}\hat{u}_{\varepsilon_n}(x) = \hat{u}(x) = \left(g_i^*\right)'\left(\hat{p}(x)\right).$$

(ii) $\hat{p}(x)\in\overline{P}_i\cap\overline{P}_{i+1}$ for some $1\leqslant i<d$: Again by definition, we have

$$u_i \leqslant h_\varepsilon\left(\hat{p}(x)\right) \leqslant u_{i+1}.$$

Since $\lim_{n\to\infty} p_{\varepsilon_n}(x) = \hat{p}(x)$, we have using the definition of h_ε that

$$u_i \leqslant h_{\varepsilon_n}\left(p_{\varepsilon_n}(x)\right) \leqslant u_{i+1}$$

for all n sufficiently large. Since

$$\hat{u}_{\varepsilon_n}(x) = \sum_{k=1}^{l_n}\gamma_k^n u_{\varepsilon_{n_k}} = \sum_{k=1}^{l_n}\gamma_k^n h_{\varepsilon_{n_k}}\left(p_{\varepsilon_{n_k}}(x)\right),$$

we obtain that

$$u_i \leqslant \lim_{n\to\infty}\hat{u}_{\varepsilon_n}(x) = \hat{u}(x) \leqslant u_{i+1}.$$

(iii) $\hat{p}(x)\in\overline{P}_i\cap\overline{P}_0$ for some $1\leqslant i\leqslant d$. We argue similarly as in (ii) that

$$\lim_{n\to\infty}\hat{u}_{\varepsilon_n}(x) = \hat{u}(x) \in \left[\min\left\{u_i, \frac{1}{\alpha}p(x)\right\}, \max\left\{u_i, \frac{1}{\alpha}p(x)\right\}\right].$$

These calculations imply that pointwise almost everywhere,

$$\hat{u}(x)\in\partial g^*\left(\hat{p}(x)\right),$$

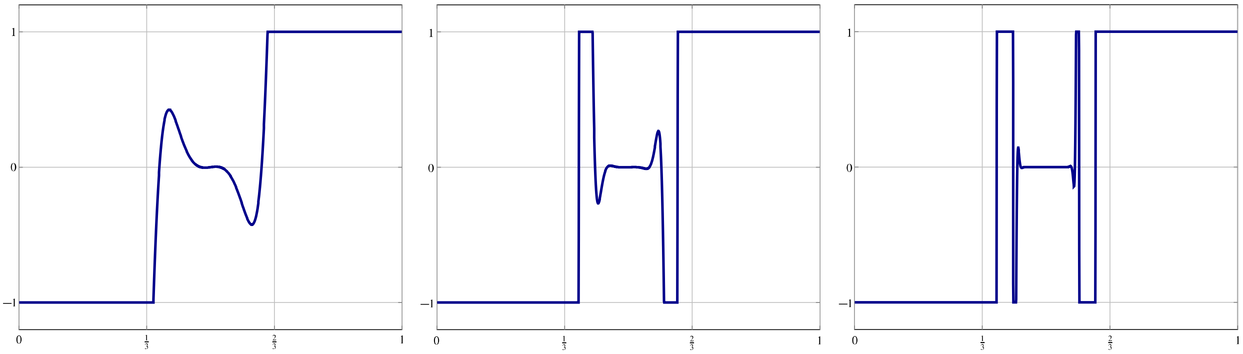
from which we deduce that $(\hat{u},\hat{y},\hat{p})$ satisfies (4.1). Since this solution is unique, the full sequence converges weakly in $L^2(\Omega)\times V\times V$ to $(\bar{u},\bar{y},\bar{p})$.

Finally, if $|\mathcal{C}| = 0$, the convergence $u_\varepsilon\to\bar{u}$ is pointwise almost everywhere since only case (i) needs to be considered. □

Remark 2. In the case that the singular arc $\mathcal{S}$ is not a set of measure zero, the weak but not pointwise convergence can be observed numerically. An example for $\Omega=(0,1)$, $A=-\partial_{xx}$ with homogeneous Dirichlet conditions, and $(u_1,u_2)=(-1,1)$ can be constructed as follows: Choose

$$\bar{p} = \begin{cases} -x\left(x-\frac{1}{3}\right)^2 & \text{if } 0<x<\frac{1}{3},\\ 0 & \text{if } \frac{1}{3}<x<\frac{2}{3},\\ (1-x)\left(x-\frac{2}{3}\right)^2 & \text{if } \frac{2}{3}<x<1,\end{cases}$$

$$\bar{u} = \begin{cases} -1 & \text{if } 0<x<\frac{1}{3},\\ 0 & \text{if } \frac{1}{3}<x<\frac{2}{3},\\ 1 & \text{if } \frac{2}{3}<x<1,\end{cases}$$



(a) $\varepsilon \approx 2.7 \cdot 10^{-7}$ (b) $\varepsilon \approx 2.3 \cdot 10^{-9}$ (c) $\varepsilon \approx 1.9 \cdot 10^{-11}$

Fig. 3. Controls u_ε showing weak but not pointwise convergence.

which satisfy $\bar u \in \partial\mathcal{G}^*(\bar p)$. The corresponding state is

$$\bar y = \begin{cases} \frac{1}{2}x^2 - \frac{2}{9}x & \text{if } 0 < x < \frac{1}{3}, \\ \frac{1}{9}x - \frac{1}{18} & \text{if } \frac{1}{3} < x < \frac{2}{3}, \\ -\frac{1}{2}x^2 + \frac{7}{9}x - \frac{5}{18} & \text{if } \frac{2}{3} < x < 1. \end{cases}$$

Setting $z = \bar y + (-\partial_{xx}\bar p)$, this choice of $(\bar u, \bar y, \bar p)$ is a solution to (4.1). However, the sequence $\{(u_\varepsilon, y_\varepsilon, p_\varepsilon)\}_{\varepsilon>0}$ does not converge pointwise on the singular arc $[\frac{1}{3}, \frac{2}{3}]$ but for smaller ε begins to oscillate between -1 and 1 with increasing frequency; see Fig. 3 (where $\alpha = \beta = 5 \cdot 10^{-1}$). This phenomenon is also known in the context of bang-bang optimal control of ordinary differential equations; see, e.g., [21].

4.2. Semismooth Newton method

We now wish to apply a semismooth Newton method to (4.3). Since h_ε is Lipschitz continuous and piecewise differentiable, the corresponding superposition operator H_ε is semismooth from $V \hookrightarrow L^r(\Omega)$ to $L^2(\Omega)$ for any $r > 2$; see, e.g., [22]. Its Newton derivative at p in direction δp is defined pointwise almost everywhere by

$$\big[D_N H_\varepsilon(p)\delta p\big](x) = \begin{cases} 0 & \text{if } p(x) \in P_i^\varepsilon, \\ \frac{1}{\alpha}\delta p(x) & \text{if } p(x) \in P_0^\varepsilon, \\ \frac{u_{i+1}-u_i}{2\varepsilon}\delta p(x) & \text{if } p(x) \in P_{i,i+1}^\varepsilon, \\ \frac{\sqrt{2\alpha\beta}+\varepsilon}{2\varepsilon\alpha}\delta p(x) & \text{if } p(x) \in P_{i,0-}^\varepsilon, \\ \frac{\sqrt{2\alpha\beta}+\varepsilon}{2\varepsilon\alpha}\delta p(x) & \text{if } p(x) \in P_{i,0+}^\varepsilon. \end{cases} \tag{4.4}$$

After eliminating u_ε, a semismooth Newton step then amounts to solving for $(\delta y, \delta p)$ in

$$\begin{cases} \delta y + A^*\delta p = z - y^k - A^* p^k, \\ A\delta y - D_N H_\varepsilon\big(p^k\big)\delta p = -Ay^k + H_\varepsilon\big(p^k\big), \end{cases} \tag{4.5}$$

and setting $y^{k+1} = y^k + \delta y$ and $p^{k+1} = p^k + \delta p$.

To show local superlinear convergence of the semismooth Newton method, it remains to show uniform boundedness of the inverse Newton matrix in (4.5), which we take as an equation from $V \times V \to V^* \times V^*$.

Proposition 4.2. *For any $(w_1, w_2) \in V^* \times V^*$ and any $p \in V$, there exists a unique solution $(\delta y, \delta p) \in V \times V$ to*

$$\begin{cases} \delta y + A^*\delta p = w_1, \\ A\delta y - D_N H_\varepsilon(p)\delta p = w_2, \end{cases}$$

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

satisfying

$$\|\delta y\|_V + \|\delta p\|_V \leqslant C\bigl(\|w_1\|_{V^*} + \|w_2\|_{V^*}\bigr),$$

where the constant $C > 0$ *depends on* α, ε, *and* u_i, *but not on* p.

Proof. First note that definition (4.4) implies that $[D_N H_\varepsilon(p)\delta p](x) \leqslant C\delta p(x)$ almost everywhere, with a constant $C > 0$ depending only on the values stated above. Now, eliminating δp using the second equation and applying A^{-1} yields

$$\delta y + A^{-1} D_N H_\varepsilon(p)\bigl(A^{-*}\delta y\bigr) = A^{-1}w_1 + A^{-1} D_N H_\varepsilon(p) A^{-*} w_2.$$

Taking the inner product with δy and using that A^{-1} and A^{-*} are isomorphisms from V^* to V as well as the continuous embedding $V \hookrightarrow L^2(\Omega) \hookrightarrow V^*$, we obtain that

$$\begin{aligned}
\|\delta y\|_{L^2}^2 &\leqslant \|\delta y\|_{L^2}^2 + \bigl\langle D_N H_\varepsilon(p) A^{-*}\delta y, A^{-*}\delta y\bigr\rangle_{L^2} \\
&\leqslant \bigl\|A^{-1}w_1\bigr\|_{L^2}\|\delta y\|_{L^2} + C\bigl\|D_N H_\varepsilon(p) A^{-*} w_2\bigr\|_{L^2}\bigl\|A^{-*}\delta y\bigr\|_{L^2} \\
&\leqslant \bigl\|A^{-1}w_1\bigr\|_{L^2}\|\delta y\|_{L^2} + C\bigl\|A^{-*} w_2\bigr\|_{L^2}\bigl\|A^{-*}\delta y\bigr\|_{L^2} \\
&\leqslant C\bigl(\|w_1\|_{V^*} + \|w_2\|_{V^*}\bigr)\|\delta y\|_{L^2},
\end{aligned} \tag{4.6}$$

the second term in the first line being nonnegative almost everywhere.

The assumption that $A^* : V \to V^*$ is an isomorphism implies coercivity of A^* and hence, using the first equation, that

$$\|\delta p\|_V \leqslant C\bigl(\|w_1\|_{V^*} + \|\delta y\|_{L^2}\bigr). \tag{4.7}$$

Similarly, the first equation and the pointwise almost everywhere bound on $D_N H_\varepsilon(p)$ implies that

$$\|\delta y\|_V \leqslant C\bigl(\|w_2\|_{V^*} + \|\delta p\|_{L^2}\bigr). \tag{4.8}$$

Combining (4.6), (4.7), and (4.8) yields the claimed estimate. □

As a consequence of Newton differentiability of H_ε and Proposition 4.2, we obtain the following result; see, e.g., [23,22].

Theorem 4.3. *The semismooth Newton iteration* (4.5) *converges locally superlinearly in* $V \times V$.

Since the right-hand side of the Newton system (4.5) is linear apart from the term $H_\varepsilon(p^k)$, we have the following termination criterion for the Newton iteration: If all active sets

$$\begin{aligned}
&\mathcal{A}_i^\varepsilon(p) = \bigl\{x \in \Omega:\ p(x) \in P_i^\varepsilon\bigr\}, \quad 0 \leqslant i \leqslant d, \\
&\mathcal{A}_{i,i+1}^\varepsilon(p) = \bigl\{x \in \Omega:\ p(x) \in P_{i,i+1}^\varepsilon\bigr\}, \quad 1 \leqslant i < d, \\
&\mathcal{A}_{i,0-}^\varepsilon(p) = \bigl\{x \in \Omega:\ p(x) \in P_{i,0-}^\varepsilon\bigr\}, \quad 1 < i \leqslant d, \\
&\mathcal{A}_{i,0+}^\varepsilon(p) = \bigl\{x \in \Omega:\ p(x) \in P_{i,0+}^\varepsilon\bigr\}, \quad 1 \leqslant i < d,
\end{aligned}$$

coincide for p^k and p^{k+1}, and the regularized control is computed as $u^{k+1} = H_\varepsilon(p^{k+1})$, then $(u^{k+1}, y^{k+1}, p^{k+1})$ satisfies (4.3); see, e.g., [23, Remark 7.1.1].

This can be used as part of a continuation strategy to compute a (possibly) optimal control: Starting with an ε^0 satisfying (4.2) and starting values $(y^0, p^0) = (0, 0)$, we solve the regularized optimality system (4.3) using the semismooth Newton iteration (4.5). If the iteration converged for some ε^m (in the sense that all active sets coincide) and the active sets $\mathcal{A}_{i,i+i}^{\varepsilon^m}(p_{\varepsilon^m})$, $\mathcal{A}_{i,0-}^{\varepsilon^m}(p_{\varepsilon^m})$ and $\mathcal{A}_{i,0+}^{\varepsilon^m}(p_{\varepsilon^m})$ are all empty, we assume that the corresponding control u_{ε^m} is optimal and stop the continuation. If they are not empty, we reduce $\varepsilon^{m+1} = \frac{1}{10}\varepsilon^m$ and solve (4.3) again with the solution for ε^m as the starting point. If the Newton iteration fails to converge within 10 steps for ε^m, we terminate the continuation and return the last iterate $u_{\varepsilon^{m-1}}$. In any case, the continuation is stopped when $\varepsilon^m < 10^{-12}$ is reached.

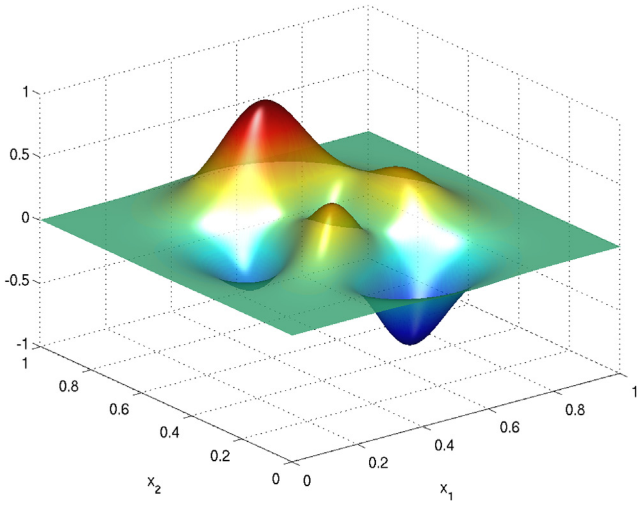


Fig. 4. Target z for numerical examples.

5. Numerical examples

We now illustrate the structure of the (sub)optimal controls with numerical examples for $\Omega = [0,1]^2$ and $A = -\Delta$ together with homogeneous Dirichlet conditions. The target for all examples is

$$z(x_1,x_2) = \frac{3}{10}(4-6x_1)^2 e^{-(6x_1-3)^2-(6x_2-2)^2} - \left(\frac{1}{5}(6x_1-3) - (6x_1-3)^3 - (6x_2-1)^5\right) e^{-(6x_1-3)^2-(6x_2-3)^2} \\ - \frac{1}{30} e^{-(6x_1-2)^2-(6x_2-3)^2},$$

i.e., a scaled version of Matlab's `peaks` function; see Fig. 4. The state y and adjoint p are discretized using piecewise linear finite elements based on a uniform triangulation $\mathcal{T}_h$ of the domain Ω with $N_h = 256 \times 256$ nodes. Integration over the piecewise defined functions $H_\varepsilon(p_h)$ and $D_N H_\varepsilon(p_h)\delta p_h$ in the weak formulation of (4.5) is approximated by applying the mass matrix to the vector of nodal values. Specifically, let V_h denote the space of piecewise linear finite elements based on the interior nodes $\{x_j\}_{j=1}^{N_h}$ of $\mathcal{T}_h$, and let $\vec{v} \in \mathbb{R}^{N_h}$ denote the vector of expansion coefficients of $v_h \in V_h$ with respect to the nodal basis of V_h, i.e., $\vec{v}_j = v_h(x_j)$ for $1 \leqslant i \leqslant N_h$. Then we define $\vec{H}_\varepsilon(\vec{p}) \in \mathbb{R}^{N_h}$ and $D\vec{H}_\varepsilon(\vec{p}) \in \mathbb{R}^{N_h}$ via

$$\big[\vec{H}_\varepsilon(\vec{p})\big]_j = h_\varepsilon\big(p(x_j)\big), \qquad \big[D\vec{H}_\varepsilon(\vec{p})\big]_j = D_N H_\varepsilon(p_h)(x_j), \qquad 1 \leqslant j \leqslant N_h.$$

The variational equation

$$\langle \nabla \delta y_h, \nabla v_h \rangle_{L^2} - \big\langle D_N H_\varepsilon\big(p_h^k\big)\delta p_h, v_h\big\rangle_{L^2} = -\big\langle \nabla y_h^k, \nabla v_h\big\rangle_{L^2} + \big\langle H_\varepsilon\big(p_h^k\big), v_h\big\rangle_{L^2} \quad \text{for all } v_h \in V_h$$

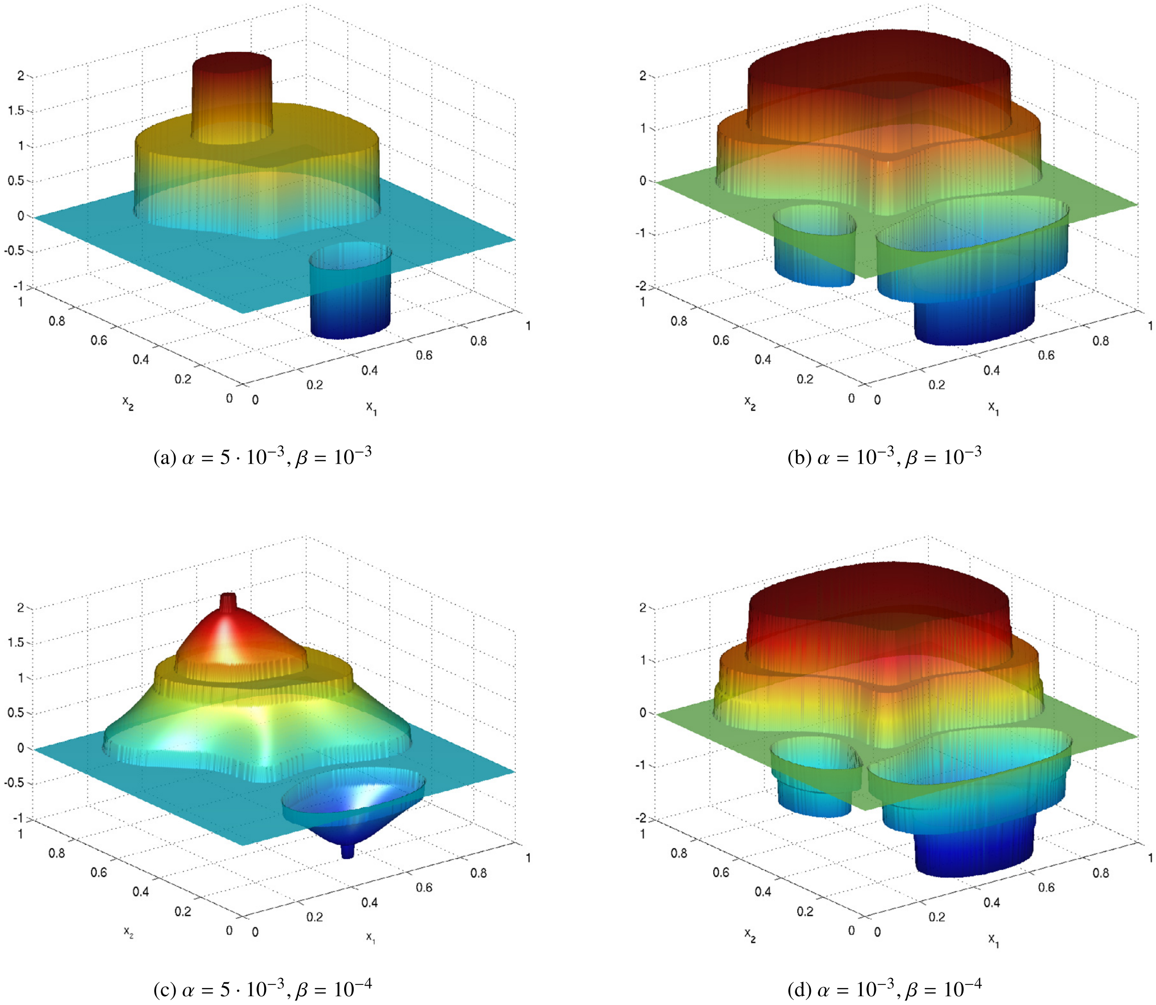
in (4.5) is then approximated by

$$A_h \vec{\delta y} - M_h\big(D\vec{H}_\varepsilon(\vec{p}) \odot \vec{\delta p}\big) = -A_h \vec{y}^k + M_h \vec{H}_\varepsilon\big(\vec{p}^k\big),$$

where A_h and M_h are, respectively, the stiffness and mass matrices corresponding to V_h, and $\odot$ denotes the Hadamard (i.e., componentwise) product of two vectors. Implementations of the described algorithm in Matlab and Python (using the `DOLFIN` module from the finite element package `FEniCS` [24,25]) can be downloaded from http://www.uni-graz.at/~clason/publications.html; the results presented in this section were obtained using the former.

We begin by illustrating the effects of the values of α and β on the structure of the resulting controls. We fix the $d = 5$ control states $(u_1, \dots, u_5) = (-2, -1, 0, 1, 2)$ and first choose $\alpha = 5 \cdot 10^{-3}$ and $\beta = 10^{-3}$. Here, the continuation terminated at $\varepsilon = 1.25 \cdot 10^{-8}$ with all nodes having values in one of the sets P_i^ε, $0 \leqslant i \leqslant d$, and the resulting (optimal) control $\bar{u} = u_\varepsilon$ is shown in Fig. 5(a); since in this case $\sqrt{2\beta/\alpha} = \sqrt{10} > 1 = \max_{1 \leqslant i < 5}(u_{i+1} - u_i)$, there are no free arcs, and $\bar{u}$ is a true multi-bang control. For the choice $\alpha = 10^{-3}$ and $\beta = 10^{-3}$, the continuation terminated

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005



(a) $\alpha = 5 \cdot 10^{-3}, \beta = 10^{-3}$

(b) $\alpha = 10^{-3}, \beta = 10^{-3}$

(c) $\alpha = 5 \cdot 10^{-3}, \beta = 10^{-4}$

(d) $\alpha = 10^{-3}, \beta = 10^{-4}$

Fig. 5. Effect of α, β on structure of control u_ε (top: no free arcs, bottom: free arcs).

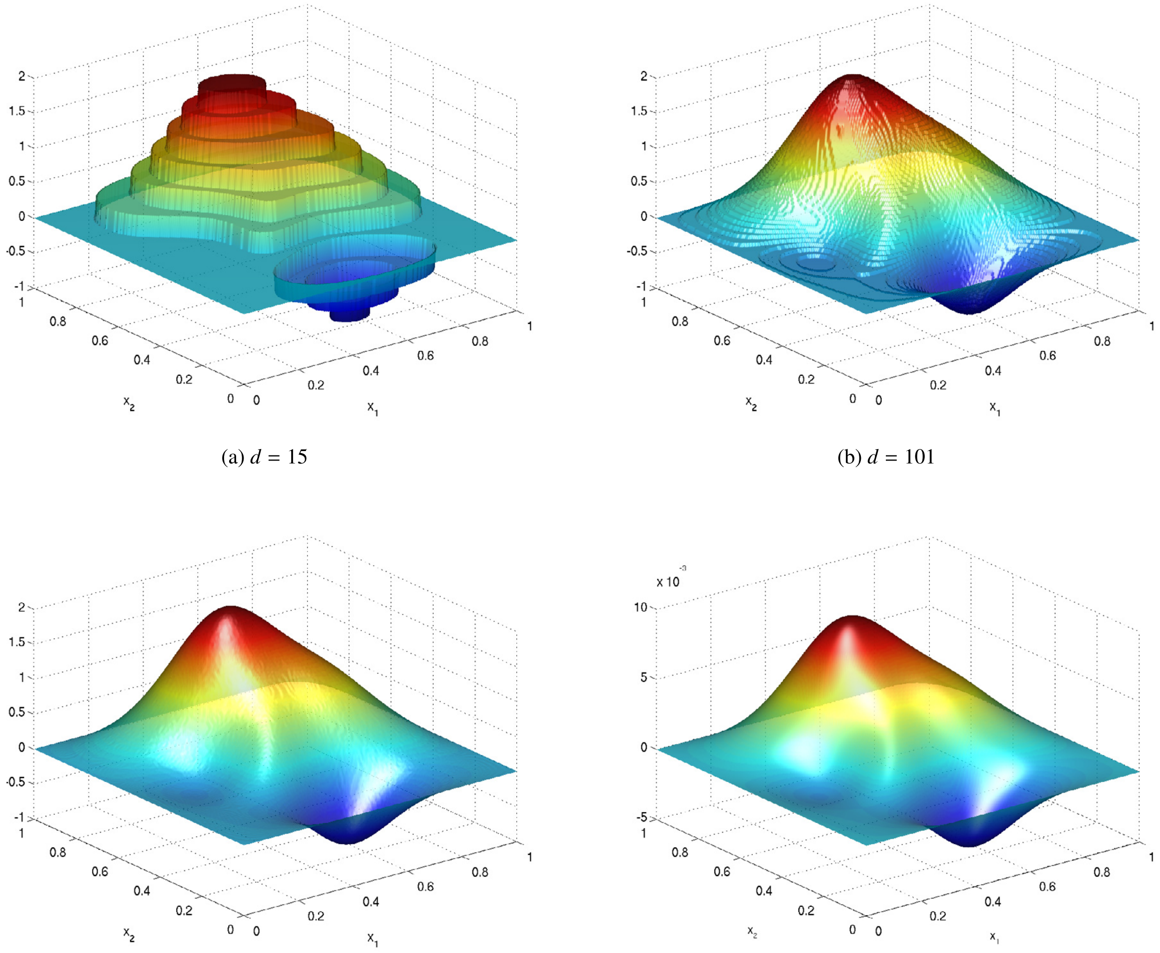
at $\varepsilon = 2.25 \cdot 10^{-7}$ with 21 out of $N_h = 65\,536$ nodes having values in one of the “regularized” sets $P^\varepsilon_{i,i+1}$, $1 \leqslant i < d$. Since in this case $\sqrt{2\beta/\alpha} = \sqrt{2} > 1$ as well, there are again no free arcs, but the smaller value of α now leads to control states of larger magnitude. In particular, $u_1 = -2$ is now attained on a subdomain; see Fig. 5(b). We repeat these tests for the smaller value $\beta = 10^{-4}$. Fig. 5(c) shows the (optimal) control $\bar{u} = u_\varepsilon$ for $\alpha = 5 \cdot 10^{-3}$ (where the continuation terminated at $\varepsilon = 9 \cdot 10^{-10}$ with no nodes having values in regularized sets), and Fig. 5(d) shows the control u_ε for $\alpha = 1 \cdot 10^{-3}$ (where the continuation terminated at $\varepsilon = 2.25 \cdot 10^{-8}$ with 7 nodes having values in one of the regularized sets $P^\varepsilon_{i,0-}$, $P^\varepsilon_{i,0+}$, $1 \leqslant i \leqslant d$). In both cases, condition (2.7) is violated, and there are free arcs around $u(x) \in \{-1.5, -0.5, 0.5, 1.5\}$ whose size decreases with decreasing α.

Table 1 shows the convergence history for the example in Fig. 5(a) by giving the total number of nodes that changed in one of the active sets after each step k (i.e., the number of indices j for which, e.g., $[\vec{\chi}_{\mathcal{A}_i^\varepsilon(\vec{p}^{k+1})}]_j \neq [\vec{\chi}_{\mathcal{A}_i^\varepsilon(\vec{p}^{k})}]_j$; nodes changing between active sets are counted separately for each set). The residual norm in (4.3) behaves very similarly and is thus not shown. The first iteration with $\varepsilon^0 = 1.2510^{-3}$ and starting from $(y^0, p^0) = (0, 0)$ demonstrates the typical behavior of a converging semismooth Newton method: After some initial steps of relatively constant decrease, the iteration enters a superlinear phase (here, after step 3) in which convergence is achieved within a few steps. Due to the continuation, the following iterations already start in the superlinear phase and require successively fewer steps. The shown behavior is representative of the other examples as well, as long as the Newton iterations converged. If an iteration failed to converge, it entered a cycle where a small number of nodes (typically 2–6) alternated between two active sets.

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005

Table 1
Convergence history for example in Fig. 5(a) (shown are the number of nodes $n(k)$ that changed in the active sets after step k).

ε	1.25e-3					1.25e-4			1.25e-5			1.25e-6			1.25e-7			1.25e-8	
k	1	2	3	4	5	1	2	3	1	2	3	1	2	3	1	2	3	1	2
$n(k)$	65 353	65 386	35 218	130	0	36014	88	0	3396	2	0	324	4	0	36	2	0	2	0



(a) d = 15

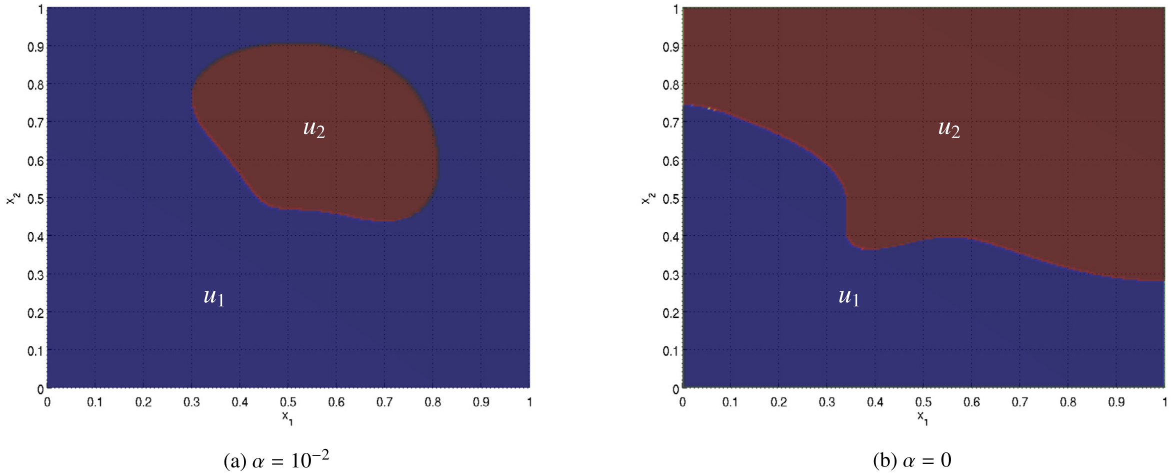
(b) d = 101

(c) d = 1001

(d) L^2 control (β = 0)

Fig. 6. Control u_ε for d control states uniformly distributed in $[-2, 2]$.

Let us next address the feasibility of the approach with respect to the number of control states. We again take d uniformly spaced control states between -2 and 2, set $\alpha = 5 \cdot 10^{-3}$ and $\beta = 10^{-3}$ to avoid free arcs, and compute the resulting controls for $d = 15$ (Fig. 6(a), for $\varepsilon = 3.21 \cdot 10^{-9}$ with no nodes remaining in regularized active sets), $d = 101$ (Fig. 6(b), for $\varepsilon = 4.5 \cdot 10^{-9}$ with 2 nodes remaining in regularized active sets) and $d = 1001$ (Fig. 6(c), for $\varepsilon = 4.5 \cdot 10^{-11}$ with 2 nodes remaining in regularized active sets). As d grows, the controls approach in shape the solution of the standard quadratic optimal control problem (i.e., with $\beta = 0$, shown for comparison in Fig. 6(d)), although they differ in magnitude due to the choice of desired control states and do not coincide with controls obtained by rounding the quadratic control appropriately. We remark that the number of steps in the semismooth Newton iterations is almost independent of d. Furthermore, the size of the system matrices does not depend on d; and since the u_i are ordered, the number of active sets grows only linearly with d. Hence, the presented approach has linear asymptotic complexity in d.



(a) $\alpha = 10^{-2}$

(b) $\alpha = 0$

Fig. 7. Comparison of multi-bang ($\alpha > 0$) and bang-bang ($\alpha = 0$) controls for $(u_1, u_2) = (-1, 2)$ (shown are level-sets $\{u_\varepsilon(x) = u_i\}$).

Finally, we illustrate the relation to (generalized) bang-bang controls. We take the $d = 2$ control states $(u_1, u_2) = (-1, 2)$ and compare the control for $\alpha = 10^{-2}$ to the control for $\alpha = 0$, choosing $\beta = 2 \cdot 10^{-2}$ sufficiently large in both cases to prevent free arcs. The control for $\alpha > 0$ is shown in Fig. 7(a) (for $\varepsilon = 6.75 \cdot 10^{-8}$ with 1 node in regularized active sets), and the control for $\alpha = 0$ is shown in Fig. 7(b) (for $\varepsilon = 10^{-7}$ with 1024 nodes in regularized active sets). It can be observed that the control for $\alpha = 10^{-2}$ is biased towards the control state $u_1 = -1$ with smaller magnitude, as opposed to the control for $\alpha = 0$ which shows no such bias.

6. Conclusion

Control problems where the control is desired to take values only from a discrete set of control states can be formulated using a combination of L^2 and L^0-type penalties. Although the resulting problem (1.1) is nonconvex and lacks weak lower-semicontinuity, Fenchel duality allows the derivation of a primal-dual optimality system that admits a unique solution and can be solved numerically using a regularized semismooth Newton method. The formulation (1.1) also has the potential to be an effective approach for inverse problems where the true solution can be assumed to take values only from a known discrete set of parameters (e.g., tissue types). It can be an attractive alternative to methods based on, e.g., topological derivatives, since the computational effort grows only linearly with the number d of possible parameter values. We point out in this context that the presented approach does not rely fundamentally on linearity of the state equation; all arguments can be extended (under suitable assumptions) to nonlinear control-to-state (or parameter-to-observation) mappings. In addition, the control states u_i are not required to be constant; since the derivations involving the multi-bang penalty $\mathcal{G}$ are pointwise in nature, they apply in a straightforward manner to distributed $u_i(x)$ as well. Finally, problem (1.1) can be seen as a prototype for nonconvex relaxations of other hybrid discrete-continuous problems.

Acknowledgements

This work was supported by the Austrian Science Fund (FWF) under grant SFB F32 (SFB “Mathematical Optimization and Applications in Biomedical Sciences”).

References

[1] K. Ito, K. Kunisch, $L^p(\Omega)$-optimization with $p \in [0, 1)$, Technical Report 2012–19, SFB MOBIS, 2012.
[2] G. Stadler, Elliptic optimal control problems with L_1-control cost and applications for the placement of control devices, Comput. Optim. Appl. 44 (2009) 159–181.
[3] E. Casas, R. Herzog, G. Wachsmuth, Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional, SIAM J. Control Optim. 22 (2012) 795–820.

[4] C. Clason, K. Kunisch, A duality-based approach to elliptic control problems in non-reflexive Banach spaces, ESAIM Control Optim. Calc. Var. 17 (2011) 243–266.

[5] F. Tröltzsch, A minimum principle and a generalized bang-bang principle for a distributed optimal control problem with constraints on control and state, Z. Angew. Math. Mech. 59 (1979) 737–739.

[6] D. Tiba, Optimal Control of Nonsmooth Distributed Parameter Systems, Lect. Notes Math., vol. 1459, Springer-Verlag, Berlin, 1990.

[7] M. Bergounioux, D. Tiba, Some examples of optimality conditions for convex control problems with general constraints, in: Control of Partial Differential Equations and Applications, Laredo, 1994, in: Lect. Notes Pure Appl. Math., vol. 174, Dekker, New York, 1996, pp. 23–30.

[8] M. Bergounioux, F. Tröltzsch, Optimality conditions and generalized bang-bang principle for a state-constrained semilinear parabolic problem, Numer. Funct. Anal. Optim. 17 (1996) 517–536.

[9] M. Fu, B.R. Barmish, Adaptive stabilization of linear systems via switching control, IEEE Trans. Autom. Control 31 (1986) 1097–1103.

[10] D. Liberzon, Switching in Systems and Control, Systems Control Found. Appl., Birkhäuser Boston Inc., Boston, MA, 2003.

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

[12] Q. Lü, E. Zuazua, Robust null controllability for heat equations with unknown switching control mode, Discrete Contin. Dyn. Syst., Ser. B (2013), in press.

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

[14] W. Schirotzek, Nonsmooth Analysis, Universitext, Springer, Berlin, 2007.

[15] I. Ekeland, R. Témam, Convex Analysis and Variational Problems, Classics Appl. Math., vol. 28, SIAM, Philadelphia, 1999.

[16] H.H. Bauschke, P.L. Combettes, Convex Analysis and Monotone Operator Theory in Hilbert Spaces, CMS Books Math./Ouvrages Math. SMC, Springer, New York, 2011.

[17] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, Springer, New York, 2010.

[18] K. Yosida, Functional Analysis, 6th edition, Grundlehren Math. Wiss., vol. 123, Springer-Verlag, Berlin, 1980.

[19] D. Kinderlehrer, G. Stampacchia, An Introduction to Variational Inequalities and Their Applications, Classics Appl. Math., vol. 31, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2000. Reprint of the 1980 original.

[20] H. Brezis, Opérateurs Maximaux Monotones et Semi-Groupes de Contractions dans les Espaces de Hilbert, North-Holland Publishing Co., Amsterdam, 1973.

[21] C. Silva, E. Trélat, Smooth regularization of bang-bang optimal control problems, IEEE Trans. Autom. Control 55 (2010) 2488–2499.

[22] M. Ulbrich, Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces, MOS-SIAM Ser. Optim., vol. 11, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011.

[23] K. Ito, K. Kunisch, Lagrange Multiplier Approach to Variational Problems and Applications, Adv. Des. Control, vol. 15, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2008.

[24] A. Logg, G.N. Wells, DOLFIN: Automated finite element computing, ACM Trans. Math. Softw. 37 (2010) 1–28.

[25] A. Logg, K.-A. Mardal, G.N. Wells, et al., Automated Solution of Differential Equations by the Finite Element Method, Springer, 2012, software available from, http://fenicsproject.org.

Please cite this article in press as: C. Clason, K. Kunisch, Multi-bang control of elliptic systems, Ann. I. H. Poincaré – AN (2013), http://dx.doi.org/10.1016/j.anihpc.2013.08.005