Numerische Mathematik (2025) 157:143–163
https://doi.org/10.1007/s00211-024-01448-1

Numerische Mathematik

Bilinear control of semilinear elliptic PDEs: convergence of a semismooth Newton method

Eduardo Casas[1] · **Konstantinos Chrysafinos**[2,3] · **Mariano Mateos**[4]

Received: 20 September 2023 / Revised: 22 November 2024 / Accepted: 24 November 2024 /
Published online: 19 December 2024

Abstract

In this paper, we carry out the analysis of the semismooth Newton method for control-constrained bilinear control problems of semilinear elliptic PDEs. We prove existence, uniqueness and regularity for the solution of the state equation, as well as differentiability properties of the control to state mapping. Then, first and second order optimality conditions are obtained. Finally, we prove the superlinear convergence of the semismooth Newton method to local solutions satisfying no-gap second order sufficient optimality conditions as well as a strict complementarity condition.

Mathematics Subject Classification 35J61 · 49K20 · 49M15 · 49M05

The first and third authors were supported by MCIN/ AEI/10.13039/501100011033/ under research projects PID2020-114837GB-I00 and PID2023-147610NB-I00. The second author was supported by the Hellenic Foundation for Research and Innovation (H.F.R.I.) under the "First Call for H.F.R.I. Research Projects to support Faculty members and Researchers and the procurement of high-cost research equipment grant" (Project Number: 3270).

✉ Mariano Mateos
mmateos@uniovi.es

Eduardo Casas
eduardo.casas@unican.es

Konstantinos Chrysafinos
chrysafinos@math.ntua.gr

1 Departamento de Matemática Aplicada y Ciencias de la Computación, E.T.S.I. Industriales y de Telecomunicación, Universidad de Cantabria, Av. Los Castros s/n, 39005 Santander, Spain

2 Department of Mathematics, School of Applied Mathematics and Physical Sciences, National Technical University of Athens Zografou Campus, 15780 Athens, Greece

3 IACM,FORTH, 20013 Heraklion, Crete, Greece

4 Departamento de Matemáticas, Campus de Gijón, Universidad de Oviedo, 33203 Gijón, Spain

1 Introduction

In this paper, we propose a semismooth Newton method to solve the following bilinear optimal control problem:

$$\text{(P)} \min_{u \in U_{\text{ad}}} J(u) := \int_\Omega L(x, y_u(x))\,\mathrm{d}x + \frac{\nu}{2}\int_\Omega u^2(x)\,\mathrm{d}x,$$

where y_u is the state associated with the control u solution of

$$\begin{cases} Ay + a(x, y) + uy = 0 & \text{in } \Omega, \\ \partial_{n_A} y = g & \text{on } \Gamma. \end{cases} \tag{1.1}$$

Here $\Omega \subset \mathbb{R}^d$, $d = 2$ or 3, is a bounded open connected set with a Lipschitz boundary Γ. The precise assumptions on the data will be given in Sects. 2 and 3. For the moment, we underline that the parameter $\nu > 0$ and the admissible set of controls is defined as

$$U_{\text{ad}} = \{u \in L^2(\Omega) : \alpha \le u(x) \le \beta \text{ a.e. in } \Omega\}.$$

Our main goal is to prove convergence of a semi-smooth Newton method for the bilinear control problem (P). Bilinear controls have numerous applications in biology, ecology, socio-economy and engineering; see [3, 10, 15]. The key structural feature of such problems is the "bilinear" structure of the control; that is the nonlinear multiplicative coupling uy of the control variable u to its state variable y, in contrast to the classical optimal control setting (see for instance [21]) where the control typically appears in an additive way at the right hand side of the equation.

Semi-smooth Newton type methods are well known for their computational effectivity and their robust performance in a variety of optimization problems; see [11, 23] and references within. For various results related to the use of semi-smooth type methods within the context of PDE-constrained optimization we refer the reader to the books of [12, 13, 23] and references within.

Despite its wide applicability, results regarding convergence properties of semi-smooth methods associated to nonlinear PDE constrained optimization problems are very limited; see [2, 14, 18, 19, 22] for problems involving additive controls and [9] for problems involving bilinear controls. In these works strong second order assumptions are imposed, which frequently imply local convexity of the control problem around a local solution. In the recent paper [6], the superlinear convergence is proved for an additive control under no-gap second order optimality conditions and a strict complementarity assumption.

The case of bilinear controls posses additional challenges. For instance, the control enters to the PDE in a multiplicative way, and the sign of the bilinear term uy is not necessarily strictly positive. As a consequence, the derivation of suitable second order conditions substantially differs from the classical case, since various results regarding the well-posedness and differentiability properties of the control to state and adjoint-state mappings are non standard.

Our paper fills this gap in the case of bilinear controls for optimal control problems related to semi-linear elliptic PDEs. In particular, we prove the superlinear convergence of the semi-smooth Newton method, under the standard assumptions of the no-gap second order optimality conditions and a strict complementarity conditions (similar to the assumptions of [6] and to the finite dimensional case [17]). The key ingredient of the proof is the development of a suitable second order condition on an extended cone. For the later, we prove various local well posedness and differentiability results for the associated control to state and adjoint state mappings. The second order condition allows to prove the uniform boundedness of certain generalized derivatives of the solution operator equation associated to the semi-smooth Newton method, which is a necessary result in order to exploit the abstract convergence framework of [23].

The paper is organized as follows. In Sect. 2, we present the analysis of the state equation and, in particular, well-posedness results related to the control to state mapping. In Sect. 3, we study the optimal control problem, and in particular we prove first and second conditions for a local minimizer. In Sect. 4, we employ the functional framework of [23] to study the convergence of the semi-smooth Newton method while in Sect. 5 we present a numerical example that verifies our theoretical findings.

2 Analysis of the state equation

In this section we prove existence and uniqueness of solution of (1.1) as well as differentiability properties of the relation control to state. To this end, we make the following assumptions.

Assumption 2.1 The operator A is defined in Ω by the expression

$$Ay = -\sum_{i,j=1}^{d} \partial_{x_j}[a_{ij}(x)\partial_{x_i} y]$$

with $a_{ij} \in L^\infty(\Omega)$ for $1 \le i, j \le d$ satisfying for some $M_A, \Lambda_A > 0$

$$M_A|\xi|^2 \ge \sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \ge \Lambda_A|\xi|^2 \quad \text{for a.a. } x \in \Omega \text{ and } \forall \xi \in \mathbb{R}^d.$$

Assumption 2.2 We assume that $a : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function of class C^2 with respect to the second variable satisfying the following properties for a.a. $x \in \Omega$:

- $a(\cdot, 0) \in L^p(\Omega)$ for some $p > \dfrac{d}{2}$,
- $\exists a_0 \in L^\infty(\Omega)$ such that $\dfrac{\partial a}{\partial y}(x, y) \ge a_0(x)\ \forall y \in \mathbb{R}$,

- $\forall M > 0\ \exists C_{a,M}$ such that $\displaystyle\sum_{j=1}^{2}\left|\frac{\partial^j a}{\partial y^j}(x,y)\right| \le C_{a,M}\ \forall |y| \le M$,
- $\forall \varepsilon > 0$ and $\forall M > 0\ \exists \rho > 0$ such that $\left|\dfrac{\partial^2 a}{\partial y^2}(x,y_1) - \dfrac{\partial^2 a}{\partial y^2}(x,y_2)\right| \le \varepsilon$
  for all $|y_1|, |y_2| \le M$ with $|y_1 - y_2| \le \rho$.

Assumption 2.3 For the boundary data we assume that $g \in L^q(\Gamma)$ with $q > d-1$.

We observe that the normal derivative $\partial_{n_A} y$ is formally defined by

$$\partial_{n_A} y = \sum_{i,j=1}^{d} a_{ij}\partial_{x_i} y(x) n_j(x),$$

where $n(x)$ denotes the outward unit normal vector to Γ at the point x. Due to the Lipschitz regularity of Γ such a vector $n(x)$ exists for almost all $x \in \Gamma$. For a rigorous definition of the normal derivative in a trace sense the reader is referred, for instance, to [5].

Throughout this paper the following notation will be used:

$$m_u := \operatorname{ess\,inf}_{x\in\Omega} u(x),\ \mathcal{A}_0 := \{u \in L^2(\Omega) : a_0(x) + m_u \ge 0 \text{ a.e. in } \Omega \text{ and } a_0 + u \not\equiv 0\}.$$

From Assumption 2.1, for every $u \in \mathcal{A}_0$ we infer the existence of a constant $0 < \Lambda_u \le \Lambda_A$ such that

$$\int_\Omega \Big(\sum_{i,j=1}^{d} a_{i,j}\partial_{x_i} y \partial_{x_j} y + [a_0 + u]y^2\Big)\,\mathrm{d}x \ge \Lambda_u \|y\|^2_{H^1(\Omega)}\quad \forall y \in H^1(\Omega). \tag{2.1}$$

It is well known that $H^1(\Omega) \subset L^r(\Omega)$ for every $r \le \frac{2d}{d-2}$, with $r < \infty$ if $d = 2$. Hence, we have

$$\exists C_{r,\Omega} > 0 \text{ such that } \|y\|_{L^r(\Omega)} \le C_{r,\Omega}\|y\|_{H^1(\Omega)}\quad \forall y \in H^1(\Omega). \tag{2.2}$$

Analogously, since $H^{1/2}(\Gamma)$ is continuously embedded in $L^{q'}(\Gamma)$ for $q > d-1$, where $q' = \frac{q}{q-1}$ denotes the conjugate of q, we also have

$$\exists C_{q',\Gamma} > 0 \text{ such that } \|y\|_{L^{q'}(\Gamma)} \le C_{q',\Gamma}\|y\|_{H^1(\Omega)}\quad \forall y \in H^1(\Omega). \tag{2.3}$$

Theorem 2.4 *There exists $\mu \in (0, 1]$ such that for every $u \in \mathcal{A}_0$ there exists a unique solution $y_u \in H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$ of (1.1). Furthermore, the following estimates hold:*

$$\|y_u\|_{H^1(\Omega)} \leq \frac{1}{\Lambda_u} \left(C_{p',\Omega} \|a(\cdot, 0)\|_{L^p(\Omega)} + C_{q',\Gamma} \|g\|_{L^q(\Gamma)} \right), \tag{2.4}$$

$$\|y_u\|_{L^\infty(\Omega)} \leq \frac{1}{\Lambda_u} C_{p,q}, \tag{2.5}$$

$$\|y_u\|_{C^{0,\mu}(\bar{\Omega})} \leq C_{\mu,\infty} \left(\|a(\cdot, 0)\|_{L^p(\Omega)} + \|u\|_{L^2(\Omega)} + \|g\|_{L^q(\Gamma)} \right), \tag{2.6}$$

where $C_{p,q}$ depends on $\|a(\cdot, 0)\|_{L^p(\Omega)}$ and $\|g\|_{L^q(\Gamma)}$, and $C_{\mu,\infty}$ depends as well on $a(\cdot, 0)$ and g and on a monotone nondecreasing way on $\|y_u\|_{L^\infty(\Omega)}$.

Proof We define the mapping

$$b : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}, \; b(x, y) := a(x, y) - a(x, 0) - a_0(x)y. \tag{2.7}$$

Then, b satisfies $b(y, 0) = 0$ and $\frac{\partial b}{\partial y}(x, y) \geq 0$ due to Assumption 2.2. Furthermore, (1.1) can be written in the form

$$\begin{cases} Ay + (a_0 + u)y + b(x, y) = -a(x, 0) \text{ in } \Omega, \\ \partial_{n_A} y = g \text{ on } \Gamma. \end{cases} \tag{2.8}$$

From Assumption 2.1 and (2.1) and (2.2) we get that $A + (a_0 + u)I : H^1(\Omega) \longrightarrow H^1(\Omega)^*$ is a linear operator satisfying the following properties

$$\langle (A + (a_0 + u))y, y \rangle \geq \Lambda_u \|y\|^2_{H^1(\Omega)} \text{ and } \langle (A + (a_0 + u))y, \phi \rangle \leq M_u \|y\|_{H^1(\Omega)} \|\phi\|_{H^1(\Omega)},$$

where $M_u := M_A + a_0 + C^2_{4,\Omega} \|u\|_{L^2(\Omega)}$. Therefore, there exists a unique solution of (2.8) in $H^1(\Omega) \cap L^\infty(\Omega)$ (see e.g. [4]). Inequality (2.4) follows easily by testing (2.8) with y and using the established coercivity of the operator $A + (a_0 + u)I$ and the fact that $b(x, y)y \geq 0$ together with (2.2) and (2.3). To prove (2.5) we proceed similarly to [1] by introducing the function $y_k(x) := y(x) - \mathrm{Proj}_{[-k,k]}(y(x))$ for every integer $k \geq 1$. Testing (2.8) with y_k, using that $\partial_{x_i} y \partial_{x_j} y_k = \partial_{x_i} y_k \partial_{x_j} y_k$, the inequality $(a_0 + u)yy_k \geq (a_0 + u)y_k^2$, and $b(x, y)y_k \geq 0$ we infer

$$\|y_k\|^2_{H^1(\Omega)} \leq \frac{1}{\Lambda_u} \left(\int_\Omega |a(\cdot, 0)||y_k(x)| \, \mathrm{d}x + \int_\Gamma |g(x)||y_k(x)| \, \mathrm{d}x \right).$$

Following the techniques of [20] and [1] we deduce the estimate. To prove (2.6), we write (1.1) in the form

$$\begin{cases} Ay + y = (1 - u)y - a(x, y) \text{ in } \Omega, \\ \partial_{n_A} y = g \text{ on } \Gamma. \end{cases}$$

Setting $M := \|y\|_{L^\infty(\Omega)}$, with Assumption 2.2 and the mean value theorem we deduce

$$|a(x, y)| \le |a(x, 0)| + C_{a,M} M.$$

In addition, we have $\|(1-u)y\|_{L^2(\Omega)} \le (\|u\|_{L^2(\Omega)} + \sqrt{|\Omega|})M$. Combining these estimates with the results of [16] we infer the existence of $\mu \in (0, 1]$ such that $y \in C^{0,\mu}(\bar{\Omega})$ and inequality (2.6) holds. □

Next we consider the diffentiability of the mapping $u \to y_u$.

Theorem 2.5 *There exists an open set $\mathcal{A}$ in $L^2(\Omega)$ such that $\mathcal{A}_0 \subset \mathcal{A}$ and $\forall u \in \mathcal{A}$ the equation (1.1) has a unique solution $y_u \in H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$, where $\mu \in (0, 1]$ was introduced in Theorem 2.4. Further, the mapping $G : \mathcal{A} \longrightarrow H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$ defined by $G(u) := y_u$ is of class C^2 and $\forall u \in \mathcal{A}$ and $\forall v, v_1, v_2 \in L^2(\Omega)$ the functions $z = G'(u)v$ and $w = G''(u)(v_1, v_2)$ are the unique solutions of the equations:*

$$\begin{cases} Az + \dfrac{\partial a}{\partial y}(x, y_u)z + uz = -vy_u \ \ in\ \Omega, \\ \partial_{n_A} z = 0 \ \ on\ \Gamma, \end{cases} \tag{2.9}$$

$$\begin{cases} Aw + \dfrac{\partial a}{\partial y}(x, y_u)w + uw \\ \quad = -\dfrac{\partial^2 a}{\partial y^2}(x, y_u) z_{u,v_1} z_{u,v_2} - v_1 z_{u,v_2} - v_2 z_{u,v_1} \ \ in\ \Omega, \\ \partial_{n_A} w = 0 \ \ on\ \Gamma, \end{cases} \tag{2.10}$$

where $z_{u,v_i} = G'(u)v_i$, $i = 1, 2$.

Proof We define the space

$$Y_A := \{y \in H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega}) : Ay \in L^p(\Omega),\ \partial_{n_A} y \in L^q(\Gamma)\}$$

which is a Banach space when endowed with the graph norm. We also define the mapping

$$\mathcal{F} : L^2(\Omega) \times Y_A \longrightarrow L^p(\Omega) \times L^q(\Gamma), \quad \mathcal{F}(u, y) := (Ay + a(x, y) + uy, \partial_{n_A} y - g).$$

From Assumption 2.2 we deduce that $\mathcal{F}$ is of class C^2. For every $(\bar{u}, \bar{y}) \in \mathcal{A}_0 \times Y_A$ the derivative $\frac{\partial \mathcal{F}}{\partial y}(\bar{u}, \bar{y}) : Y_A \longrightarrow L^p(\Omega) \times L^q(\Gamma)$, given by

$$\frac{\partial \mathcal{F}}{\partial y}(\bar{u}, \bar{y})z = \left(Az + \frac{\partial a}{\partial y}(x, \bar{y})z + \bar{u}z, \partial_{n_A} z \right) \ \forall z \in Y_A,$$

is linear and continuous. Using Theorem 2.4, we deduce that the equation

$$\begin{cases} Az + \dfrac{\partial a}{\partial y}(x, \bar{y})z + \bar{u}z = f \ \ \text{in}\ \Omega, \\ \partial_{n_A} z = h \ \ \text{on}\ \Gamma, \end{cases} \tag{2.11}$$

has unique solution $z \in Y_A$ for all $(f, h) \in L^p(\Omega) \times L^q(\Gamma)$. The open mapping theorem implies that $\frac{\partial \mathcal{F}}{\partial y}(\bar{u}, \bar{y})$ is an isomorphism and there exists $\varepsilon_{\bar{u}} > 0$ and $\varepsilon_{\bar{y}} > 0$, such that $\forall u \in B_{\varepsilon_{\bar{u}}}(\bar{u}) \subset L^2(\Omega)$ the equation $\mathcal{F}(u, y) = 0$ has a unique solution y_u in the ball $B_{\varepsilon_{\bar{y}}}(\bar{y}) \subset Y$. Moreover the mapping $u \in B_{\varepsilon_{\bar{u}}}(\bar{u}) \to y_u \in B_{\varepsilon_{\bar{y}}}(\bar{y})$ is of class C^2. Without loss of generality, we assume $\varepsilon_{\bar{u}} < \frac{\Lambda_{\bar{u}}}{C^2_{4,\Omega}}$, where $\Lambda_{\bar{u}}$ is defined in (2.1) and $C_{4,\Omega}$ is introduced in (2.2) for $r = 4$. We prove that for every $u \in B_{\varepsilon_{\bar{u}}}$ the equation $\mathcal{F}(u, y) = 0$ has unique solution $y \in Y_A$. Indeed, suppose that y_1, y_2 are two solutions of $\mathcal{F}(u, y) = 0$. We set $y = y_1 - y_2$, subtract the corresponding equations, and apply the mean value theorem to deduce that y satisfies

$$\begin{cases} Ay + \dfrac{\partial a}{\partial y}(x, y_1 + \theta_x y)y + uy = 0 \ \text{ in } \Omega, \\ \partial_{n_A} y = 0 \ \text{ on } \Gamma, \end{cases} \tag{2.12}$$

where $\theta_x : \Omega \to [0, 1]$ is a measurable function. The equation (2.12) can be written as

$$\begin{cases} Ay + \left[\dfrac{\partial a}{\partial y}(x, y_1 + \theta_x y) + \bar{u}\right] y + (u - \bar{u})y = 0 \ \text{ in } \Omega, \\ \partial_{n_A} y = 0 \ \text{ on } \Gamma. \end{cases} \tag{2.13}$$

Testing (2.13) with y we get

$$\left(\Lambda_{\bar{u}} - C^2_{4,\Omega}\varepsilon_{\bar{u}}\right) \|y\|^2_{H^1(\Omega)} \le \Lambda_{\bar{u}}\|y\|^2_{H^1(\Omega)} - C^2_{4,\Omega}\|u - \bar{u}\|_{L^2(\Omega)}\|y\|^2_{H^1(\Omega)} \le 0.$$

Hence, $y = 0$ holds. Finally, defining in $L^2(\Omega)$ the open set $\mathcal{A} = \cup_{\bar{u}\in\mathcal{A}_0} B_{\varepsilon_{\bar{u}}}(\bar{u})$ and the mapping $G : \mathcal{A} \longrightarrow Y$ such that $G(u) = y_u$, we have that G is of class of C^2. Moreover, the equations (2.9) and (2.10) are obtained differentiating with respect to u the identity $\mathcal{F}(u, G(u)) = 0$. □

3 Analysis of the optimal control problem

In this Section we are going to prove existence of solutions of problem (P) and we analyze the first and second order optimality conditions for a local minimizer. Along this paper a local minimizer is understood in $L^2(\Omega)$ sense. To this end we make the following hypothesis.

Assumption 3.1 We assume that the conditions $\nu > 0$, $a_0(x) + \alpha \ge 0$ for a.a. $x \in \Omega$, $a_0 + \alpha \not\equiv 0$, and $\alpha < \beta \le \infty$ hold.

Assumption 3.2 The function $L : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is Carathéodory and of class of C^2 with respect to the second variable. Further the following properties hold for almost all $x \in \Omega$:

- $L(\cdot, 0) \in L^1(\Omega)$,
- $\forall M > 0,\ \exists L_M \in L^p(\Omega)$ such that $\left|\dfrac{\partial L}{\partial y}(x, y)\right| \leq L_M(x)\ \forall |y| \leq M$,
- $\forall M > 0,\ \exists C_{L,M} \in \mathbb{R}$ such that $\left|\dfrac{\partial^2 L}{\partial y^2}(x, y)\right| \leq C_{L,M}\ \forall |y| \leq M$,
- $\forall \varepsilon > 0$ and $\forall M > 0\ \exists \rho > 0$ such that

$$\left|\frac{\partial^2 L}{\partial y^2}(x, y_1) - \frac{\partial^2 L}{\partial y^2}(x, y_2)\right| \leq \varepsilon\ \forall |y_1|, |y_2| \leq M \text{ with } |y_1 - y_2| \leq \rho.$$

Remark 3.3 1. From Assumption 3.1, the fact that $a_0(x) + m_u \geq a_0(x) + \alpha\ \forall u \in U_{\mathrm{ad}}$, and (2.1) we infer that $U_{\mathrm{ad}} \subset \mathcal{A}_0 \subset \mathcal{A}$.

2. Using again Assumption 3.1 we deduce the existence of a constant $\Lambda > 0$ such that the inequality (2.1) holds for every $u \in U_{\mathrm{ad}}$ with Λ_u replaced by Λ. Since $a_0(x) + m_u \geq a_0(x) + \alpha\ \forall u \in U_{\mathrm{ad}}$, without loss of generality we can assume that $\Lambda_u \geq \Lambda\ \forall u \in U_{\mathrm{ad}}$. Consequently, (2.4) and (2.5) hold with Λ instead of Λ_u, and $C_{\mu,\infty}$ in (2.6) can be chosen independently of $u \in U_{\mathrm{ad}}$.

As a consequence of Theorem 2.5 and Assumption 3.2 we deduce the differentiability of functional J.

Theorem 3.4 *The functional $J : \mathcal{A} \longrightarrow \mathbb{R}$ is of class C^2 and its derivatives are given by the expressions:*

$$J'(u)v = \int_\Omega (\nu u - y_u \varphi_u) v \,\mathrm{d}x\ \forall u \in \mathcal{A},\ \forall v \in L^2(\Omega), \tag{3.1}$$

$$\begin{aligned} J''(u)(v_1, v_2) = &\int_\Omega \Big[\frac{\partial^2 L}{\partial y^2}(x, y_u) - \varphi_u \frac{\partial^2 a}{\partial y^2}(x, y_u)\Big] z_{u,v_1} z_{u,v_2} \,\mathrm{d}x \\ &- \int_\Omega \Big[v_1 z_{u,v_2} + v_2 z_{u,v_1}\Big]\varphi_u \,\mathrm{d}x + \nu \int_\Omega v_1 v_2 \,\mathrm{d}x,\ \forall u \in \mathcal{A},\ \forall v_1, v_2 \in L^2(\Omega), \end{aligned} \tag{3.2}$$

where $z_{u,v_i} = G'(u)v_i$, $i = 1, 2$ and $\varphi_u \in H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$ is the adjoint state, the unique solution of the equation

$$\begin{cases} A^*\varphi + \dfrac{\partial a}{\partial y}(x, y_u)\varphi + u\varphi = \dfrac{\partial L}{\partial y}(x, y_u) & \text{in } \Omega, \\ \partial_{n_{A^*}}\varphi = 0 & \text{on } \Gamma. \end{cases} \tag{3.3}$$

Proof First let us analyze (3.3). To prove existence, uniqueness and regularity of solution (3.3) we first observe that there exists $\bar{u} \in \mathcal{A}_0$ such that $u \in B_{\varepsilon_{\bar{u}}}(\bar{u})$, where

$\varepsilon_{\bar u}$ is defined in the proof of Theorem 2.5. Then, (3.3) can be written as

$$\begin{cases} A^*\varphi + \Big[\dfrac{\partial a}{\partial y}(x, y_u) + \bar u\Big]\varphi + (u - \bar u)\varphi = \dfrac{\partial L}{\partial y}(x, y_u) \text{ in } \Omega, \\ \partial_{n_{A^*}}\varphi = 0 \text{ on } \Gamma. \end{cases}$$

Setting $M = \|y_u\|_{L^\infty(\Omega)}$, from Assumption 3.2 we obtain $\left|\frac{\partial L}{\partial y}(x, y_u)\right| \le L_M(x)$ with $L_M \in L^p(\Omega)$. Then, arguing as for (2.13) we obtain the coercivity of the linear equation. The existence and uniqueness of a solution in $H^1(\Omega)$ follows from Lax-Milgram theorem. Finally, using again [16] we deduce the $C^{0,\mu}(\bar\Omega)$ regularity.

The fact that J is of class C^2 is an immediate consequence of the chain rule, Theorem 2.5, and Assumption 3.2. Moreover, we have

$$\begin{aligned} J'(u)v &= \int_\Omega \Big[\frac{\partial L}{\partial y}(x, y_u)z_{u,v} + \nu u v\Big]\,\mathrm{d}x, \\ J''(u)(v_1, v_2) &= \int_\Omega \Big[\frac{\partial L}{\partial y}(x, y_u)w + \frac{\partial^2 L}{\partial y^2}(x, y_u)z_{u,v_1}z_{u,v_2} + \nu v_1 v_2\Big]\,\mathrm{d}x, \end{aligned}$$

where $z_{u,v} = G'(u)v$, $z_{u,v_i} = G'(u)v_i$, $i = 1, 2$, and $w = G''(u)(v_1, v_2)$. Combining these expressions with (2.9), (2.10), and (3.3) the formulas (3.1) and (3.2) follow. □

In the above theorem we have proved that the mapping $\Phi : \mathcal{A} \longrightarrow H^1(\Omega) \cap C^{0,\mu}(\bar\Omega)$ given by $\Phi(u) := \varphi_u$ is well defined. In the next theorem its differentiability is established.

Theorem 3.5 *The mapping Φ is of class C^1 and for all $u \in \mathcal{A}$ and $v \in L^2(\Omega)$ the function $\eta_{u,v} = \Phi'(u)v$ is the unique solution of*

$$\begin{cases} A^*\eta + \dfrac{\partial a}{\partial y}(x, y_u)\eta + u\eta = \Big[\dfrac{\partial^2 L}{\partial y^2}(x, y_u) - \varphi_u \dfrac{\partial^2 a}{\partial y^2}(x, y_u)\Big]z_{u,v} - v\varphi_u \ \textit{in } \Omega, \\ \partial_{n_{A^*}}\eta = 0 \ \textit{on } \Gamma, \end{cases} \tag{3.4}$$

where $z_{u,v} = G'(u)v$.

Proof According to Assumption 3.2 and the fact that $y_u, \varphi_u, z_{u,v} \in L^\infty(\Omega)$ we deduce that the right hand side of (3.4) belongs to $L^2(\Omega)$. As for (3.3) the existence, uniqueness, and regularity of $\eta_{u,v}$ follows. To prove the differentiability of Φ we define

$$Y_{A^*} = \{\varphi \in H^1(\Omega) \cap C^{0,\mu}(\bar\Omega) : A^*\varphi \in L^p(\Omega) \text{ and } \partial_{n_{A^*}}\varphi = 0\}$$

and $\mathcal{G} : \mathcal{A} \times Y_{A^*} \longrightarrow L^p(\Omega)$ by

$$\mathcal{G}(u, \varphi) := A^*\varphi + \frac{\partial a}{\partial y}(x, y_u)\varphi + u\varphi - \frac{\partial L}{\partial y}(x, y_u).$$

From Assumptions 2.2 and 3.2 we deduce that $\mathcal{G}$ is a C^1 mapping. We have that

$$\frac{\partial \mathcal{G}}{\partial \varphi}(u,\varphi)\eta = A^*\eta + \frac{\partial a}{\partial y}(x, y_u)\eta + u\eta$$

and $\frac{\partial \mathcal{G}}{\partial \varphi}(u,\varphi) : Y_{A^*} \longrightarrow L^p(\Omega)$ is an isomorphism. Then, applying the implicit function theorem and differentiating the identity $\mathcal{G}(u, \Phi(u)) = 0$ the result follows. □

The following corollary is a straightforward application of formula (3.2) and equation (3.4).

Corollary 3.6 *For every $v_1, v_2 \in L^2(\Omega)$ and all $u \in \mathcal{A}$, the following identities hold*

$$\begin{aligned} J''(u)(v_1, v_2) &= \int_\Omega \Big[\nu v_1 - (\varphi_u z_{u,v_1} + y_u \eta_{u,v_1})\Big] v_2 \, \mathrm{d}x \\ &= \int_\Omega \Big[\nu v_2 - (\varphi_u z_{u,v_2} + y_u \eta_{u,v_2})\Big] v_1 \, \mathrm{d}x. \end{aligned}$$

Theorem 3.7 *Problem* (P) *has at least one solution. Moreover, if $\bar u \in U_{\mathrm{ad}}$ is a local minimizer of* (P) *then there exist $\bar y, \bar\varphi \in H^1(\Omega) \cap C^{0,\mu}(\bar\Omega)$ such that*

$$\begin{cases} A\bar y + a(x, \bar y) + \bar u \bar y = 0 \ \ in\ \Omega, \\ \partial_{n_A} \bar y = g \ \ on\ \Gamma, \end{cases} \tag{3.5}$$

$$\begin{cases} A^*\bar\varphi + \dfrac{\partial a}{\partial y}(x, \bar y)\bar\varphi + \bar u\bar\varphi = \dfrac{\partial L}{\partial y}(x, \bar y) \ \ in\ \Omega, \\ \partial_{n_{A^*}} \bar\varphi = 0 \ \ on\ \Gamma, \end{cases} \tag{3.6}$$

$$\bar u(x) = \mathrm{Proj}_{[\alpha,\beta]}\left(\frac{1}{\nu}\bar y(x)\bar\varphi(x)\right). \tag{3.7}$$

The existence of a solution follows by usual arguments, taking a minimizing sequence, and observing that if $u_k \rightharpoonup \bar u$ in $L^2(\Omega)$ then $y_{u_k} \to \bar y = y_{\bar u}$ strongly in $H^1(\Omega) \cap C(\bar\Omega)$. This statement is an immediate consequence of estimates (2.4), (2.5) and (2.6) and Remark 3.3. The optimality system follows from (3.1), (3.3), and the fact that U_{ad} is convex.

From now on $(\bar u, \bar y, \bar\varphi) \in U_{\mathrm{ad}} \times [H^1(\Omega) \cap C^{0,\mu}(\bar\Omega)]^2$ will denote a triplet that satisfies (3.5), (3.6) and (3.7). Associated with this triplet we define the cone of critical directions

$$C_{\bar u} = \{v \in L^2(\Omega) : v(x) {=} 0 \text{ if } \nu\bar u(x) - \bar y(x)\bar\varphi(x) {\neq} 0 \text{ a.e. in } \Omega \text{ and (3.9) holds}\}, \tag{3.8}$$

$$v(x) \begin{cases} \geq 0 \ \text{ if } \bar u(x) = \alpha, \\ \leq 0 \ \text{ if } \bar u(x) = \beta. \end{cases} \tag{3.9}$$

Regarding the second order optimality conditions we have the following result.

Theorem 3.8 *If $\bar u$ is a local minimizer of* (P)*, then $J''(\bar u)v^2 \geq 0\ \forall v \in C_{\bar u}$ holds. Conversely, if $\bar u \in U_{\rm ad}$ satisfies the first order optimality conditions* (3.5)*,* (3.6) *and* (3.7) *and $J''(\bar u)v^2 > 0\ \forall v \in C_{\bar u} \setminus \{0\}$, then there exist $\varepsilon > 0$ and $\delta > 0$ such that*

$$J(\bar u) + \frac{\delta}{2}\|u - \bar u\|^2_{L^2(\Omega)} \leq J(u) \quad \forall u \in U_{\rm ad} \text{ with } \|u - \bar u\|_{L^2(\Omega)} \leq \varepsilon. \tag{3.10}$$

The proof of this theorem is standard. The reader is referred, for instance, to [7]. For the subsequent analysis the strict complementarity condition will be needed.

Definition 3.9 Let us define

$$\Sigma_{\bar u} = \{x \in \Omega : \bar u(x) \in \{\alpha, \beta\} \text{ and } \nu\bar u(x) - \bar y(x)\bar\varphi(x) = 0\}.$$

We say that the strict complementarity condition is satisfied at $\bar u$ if $|\Sigma_{\bar u}| = 0$, where $|\cdot|$ stands for the Lebesgue measure.

This notion is an extension to the case of infinite constraints of the usual strict complementarity condition in finite dimensional nonlinear programming.

For every $\tau \geq 0$, we define the subspace

$$T^\tau_{\bar u} = \{v \in L^2(\Omega) : v(x) = 0 \text{ if } |\nu\bar u(x) - \bar y(x)\bar\varphi(x)| > \tau \text{ a.e. in } \Omega\}. \tag{3.11}$$

If $\tau = 0$ we simply denote $T_{\bar u} = T^0_{\bar u}$.

Theorem 3.10 *Assume that $\bar u$ satisfies the strict complementarity condition. Then, the following properties hold:*

1- $T_{\bar u} = C_{\bar u}$,

2- If $\bar u$ satisfies the second order optimality condition $J''(\bar u)v^2 > 0\ \forall v \in C_{\bar u} \setminus \{0\}$, then

$$\exists \tau > 0 \text{ and } \kappa > 0 \text{ such that } J''(\bar u)v^2 \geq \kappa\|v\|^2_{L^2(\Omega)}\ \forall v \in T^\tau_{\bar u}. \tag{3.12}$$

Proof 1- It is obvious that $C_{\bar u} \subset T_{\bar u}$. Let us prove the converse inclusion. If $v \in T_{\bar u}$ we have to prove that v satisfies the sign conditions (3.9). If $\bar u(x) = \alpha$, then from (3.7) we deduce $\nu\bar u(x) - \bar y(x)\bar\varphi(x) \geq 0$. Hence, with the strict complementarity condition we get that $\nu\bar u(x) - \bar y(x)\bar\varphi(x) > 0$ for almost all x such that $\bar u(x) = \alpha$. Since $v \in T_{\bar u}$ we conclude $v(x) = 0$ for almost all x such that $\bar u(x) = \alpha$. In a similar way we argue when $\bar u(x) = \beta$.

2- We argue by contradiction. If the statement is false, then $\forall k \geq 1\ \exists v_k \in T^{1/k}_{\bar u}$ such that $J''(\bar u)v_k^2 < \frac{1}{k}\|v_k\|^2_{L^2(\Omega)}$. Dividing v_k by $\|v_k\|_{L^2(\Omega)}$ and denoting the result again by v_k, we obtain

$$v_k \in T^{1/k}_{\bar u},\ \ \|v_k\|_{L^2(\Omega)} = 1,\ \text{ and } J''(\bar u)v_k^2 < \frac{1}{k}. \tag{3.13}$$

Taking a subsequence that we denote again by v_k we have that $v_k \rightharpoonup v$ in $L^2(\Omega)$. The rest of the proof is divided in three steps.

Step 1- $v \in C_{\bar{u}}$. According to statement 1 of the theorem we only need to prove that $v(x) = 0$ if $\nu\bar{u}(x) - \bar{y}(x)\bar{\varphi}(x) \neq 0$. Given $\varepsilon > 0$ we define $\Omega_\varepsilon = \{x \in \Omega : |\nu\bar{u}(x) - \bar{y}(x)\bar{\varphi}(x)| > \varepsilon\}$. By definition of $T_{\bar{u}}^{1/k}$, we have that $v_k(x) = 0$ a.e. in Ω_ε $\forall k > \frac{1}{\varepsilon}$. Therefore, $v(x) = 0$ a.e in Ω_ε holds. Since $\varepsilon > 0$ can be selected arbitrarily small we conclude that $v(x) = 0$ for almost all x in Ω such that $\nu\bar{u}(x) - \bar{y}(x)\bar{\varphi}(x) \neq 0$, hence $v \in T_{\bar{u}} = C_{\bar{u}}$.

Step 2- $J''(\bar{u})v^2 \leq 0$. Since $v_k \rightharpoonup v$ in $L^2(\Omega)$ we get that $z_{\bar{u},v_k} = G'(\bar{u})v_k \rightharpoonup G'(\bar{u})v = z_{\bar{u},v}$ in $H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$. Therefore, the convergence $z_{\bar{u},v_k} \to z_{\bar{u},v}$ is strong in $C(\bar{\Omega})$. Then, we easily pass to the limit in (3.2) to deduce with (3.13) that $J''(\bar{u})v^2 \leq \liminf_{k\to\infty} J''(\bar{u})v_k^2 \leq 0$.

Step 3- Final Contradiction. From Steps 1 and 2, and the fact that $\bar{u}$ satisfies the second order sufficient optimality condition, we deduce that $v = 0$ and consequently, $z_{\bar{u},v_k} \to 0$ in $C(\bar{\Omega})$. Using the fact that $\|v_k\|_{L^2(\Omega)} = 1$, we have that $J''(\bar{u})v_k^2 = \nu + \varepsilon_k$ with $\varepsilon_k \to 0$. Therefore, $\nu = \lim_{k\to\infty} J''(\bar{u})v_k^2 \leq 0$ which contradicts the strict positivity of ν. □

4 Convergence of the semismooth Newton method

Following [23, Chapter 3] we are going to describe the abstract framework where our numerical algorithm fits.

Definition 4.1 Given two Banach spaces U and X, an open subset $\mathcal{A}$ of U, a continuous function $F : \mathcal{A} \longrightarrow X$, and a set-valued mapping $\partial F : \mathcal{A} \longrightarrow \mathcal{L}(U, X)$ such that $\partial F(u) \neq \emptyset$ $\forall u \in \mathcal{A}$, we say that F is ∂F-semismooth at $\bar{u} \in \mathcal{A}$ if

$$\lim_{v\to 0} \sup_{M\in\partial F(\bar{u}+v)} \frac{\|F(\bar{u}+v) - F(\bar{u}) - Mv\|_X}{\|v\|_U} = 0.$$

F is said ∂F-semismooth at $\mathcal{A}$ if it is ∂F-semismooth at every $u \in \mathcal{A}$.

The abstract semismooth Newton method is given in Algorithm 1.

Algorithm 1: Semismooth Newton method.

Initialize. Choose $u_0 \in \mathcal{A}$. Set $j = 0$.
for $j \geq 0$ **do**
  Choose $M_j \in \partial F(u_j)$ and solve $M_j v_j = -F(u_j)$.
  Set $u_{j+1} = u_j + v_j$ and $j = j + 1$.
end

Theorem 4.2 *Suppose that $F : \mathcal{A} \longrightarrow X$ is ∂F-semismooth at $\bar{u} \in \mathcal{A}$ solution of $F(u) = 0$ locally unique. Assume, furthermore, that for all j the operator $M_j \in \partial F(u_j)$ is an isomorphism and there exists $C_F > 0$ such that*

$$\|M_j^{-1}\|_{\mathcal{L}(X,U)} \le C_F \quad \forall j \ge 0. \tag{4.1}$$

Then, there exists $\delta > 0$ such that for all $u_0 \in \mathcal{A}$ with $\|u_0 - \bar{u}\|_U \le \delta$ the sequence $\{u_j\}_{j\ge 0}$ generated by the semismooth Newton method (Algorithm 1) converges superlinearly to $\bar{u}$.

The proof is this theorem can be found in [23, Theorem 3.13]; see also [11]. Let us put our problem into this particular framework. Let $X = U = L^2(\Omega)$ and $\mathcal{A}$ be the open set introduced in Theorem 2.5. We define by $F : \mathcal{A} \longrightarrow L^2(\Omega)$ by

$$F(u) = u - \mathrm{Proj}_{[\alpha,\beta]}\left(\frac{1}{\nu} y_u \varphi_u\right).$$

Due to Theorem 3.7 any local minimizer of (P) is a solution of $F(u) = 0$. In order to define $\partial F(u)$ $\forall u \in \mathcal{A}$ we introduce some additional functions.

$$\begin{aligned}
&S : \mathcal{A} \longrightarrow L^2(\Omega), \quad S(u) = \frac{1}{\nu} G(u)\Phi(u),\\
&\psi : \mathbb{R} \longrightarrow \mathbb{R}, \quad \psi(t) = \mathrm{Proj}_{[\alpha,\beta]}(t),\\
&\Psi : \mathcal{A} \longrightarrow L^2(\Omega), \quad \Psi(u)(x) = \psi(S(u)(x)).
\end{aligned}$$

For every $u \in \mathcal{A}$ we define

$$\begin{aligned}
\partial\Psi(u) = \big\{&N \in \mathcal{L}(L^2(\Omega), L^2(\Omega)) : Nv = hS'(u)v \ \forall v \in L^2(\Omega) \text{ for some}\\
&\text{measurable function } h : \Omega \longrightarrow \mathbb{R} \text{ such that } h(x) \in \partial\psi(S(u)(x))\big\}.
\end{aligned}$$

We observe that ψ is a Lipschitz function and by $\partial\psi(t)$ we denote the subdifferential in Clarke's sense; see [8, Chapter 2]. Note that

$$\partial\psi(t) = \begin{cases} \{1\} & \text{if } t \in (\alpha,\beta),\\ \{0\} & \text{if } t \notin [\alpha,\beta],\\ [0,1] & \text{if } t \in \{\alpha,\beta\}. \end{cases}$$

According to [23, Proposition 2.26], ψ is 1-order $\partial\psi$-semismooth.

Theorem 4.3 *Ψ is $\partial\Psi$-semismooth in $\mathcal{A}$.*

Proof Since Ψ is a superposition operator of ψ and S, we will apply [23, Theorem 3.49] to deduce that $\partial\Psi$-semismooth in $\mathcal{A}$. To this end it is enough to prove that $S : \mathcal{A} \longrightarrow L^2(\Omega)$ is C^1 and that $S : \mathcal{A} \longrightarrow L^6(\Omega)$ is locally Lipschitz. Indeed, noting that

$$S'(u)v = \frac{1}{\nu}[G'(u)v\Phi(u) + G(u)\Phi'(u)v] = \frac{1}{\nu}[z_{u,v}\varphi_u + y_u\eta_{u,v}],$$

using Theorems 2.5 and 3.5 and taking into account that the product of two functions of $H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$ is in the same space, we deduce that the mapping $S : \mathcal{A} \longrightarrow H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})$ is C^1.

The local Lipschitz property is obtained as follows: Since S is C^1, the mapping $DS : \mathcal{A} \longrightarrow \mathcal{L}(L^2(\Omega), H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega}))$ is continuous. Therefore, given $u \in \mathcal{A}$ there exists $\delta_u > 0$ and $L_{S,u} \geq 0$ such that $B_{\delta_u}(u) \subset A$ and

$$\|DS(\hat{u})\|_{\mathcal{L}(L^2(\Omega), H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega}))} \leq L_{S,u} \; \forall \hat{u} \in B_{\delta_u}(u). \tag{4.2}$$

The Lipschitz property in this ball is now a consequence of the generalized mean value theorem. □

Corollary 4.4 *The function $F : \mathcal{A} \longrightarrow L^2(\Omega)$ is ∂F-semismooth in $\mathcal{A}$, where*

$$\partial F(u) = \{M = I - N : N \in \partial \Psi(u)\}.$$

This is a straightforward consequence of Theorem 4.3.

To implement Algorithm 1, we select the operators $M_u : L^2(\Omega) \longrightarrow L^2(\Omega)$ for every $u \in \mathcal{A}$ as follows. First, we define the function $\lambda : \mathbb{R} \longrightarrow \mathbb{R}$ by

$$\lambda(t) = \begin{cases} 1 \text{ if } t \in (\alpha, \beta), \\ 0 \text{ otherwise.} \end{cases}$$

It is obvious that $\lambda(t) \in \partial \psi(t)$ for every $t \in \mathbb{R}$. We define $M_u : L^2(\Omega) \longrightarrow L^2(\Omega)$ by $M_u v = v - h_u \cdot S'(u)v$, where $h_u(x) = \lambda(S(u)(x)) = \lambda\big(\frac{1}{\nu} y_u(x)\varphi_u(x)\big)$. It is immediate that $M_u \in \partial F(u)$. For this selection we have the following result.

Theorem 4.5 *Let $(\bar{u}, \bar{y}, \bar{\varphi}) \in U_{\rm ad} \times [H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})]^2$ satisfy the first order optimality conditions (3.5), (3.6) and (3.7), the strict complementarity condition $|\Sigma_{\bar{u}}| = 0$, and the second order sufficient optimality condition $J''(\bar{u})v^2 > 0$ for every $v \in C_{\bar{u}} \setminus \{0\}$. Then, there exist $\delta > 0$ and $C > 0$ such that for every $u \in B_\delta(\bar{u}) \subset \mathcal{A}$ the linear operator $M_u : L^2(\Omega) \longrightarrow L^2(\Omega)$ is an isomorphism and $\|M_u^{-1}\| \leq C$ holds.*

Proof Given $u \in \mathcal{A}$ we define the active and inactive sets for u as follows

$$\mathbb{A}_u = \left\{ x \in \Omega : \frac{1}{\nu} y_u(x)\varphi_u(x) \notin (\alpha, \beta) \right\},$$
$$\mathbb{I}_u = \left\{ x \in \Omega : \frac{1}{\nu} y_u(x)\varphi_u(x) \in (\alpha, \beta) \right\}.$$

We denote by $\chi_{\mathbb{A}_u}$ and $\chi_{\mathbb{I}_u}$ the characteristic functions of $\mathbb{A}_u$ and $\mathbb{I}_u$, respectively. According to the definition of M_u we have $M_u v = v - \frac{1}{\nu}[z_{u,v}\varphi_u + y_u \eta_{u,v}]\chi_{\mathbb{I}_u}$. It is obvious that M_u is a continuous operator. Let us prove that for every $w \in L^2(\Omega)$ there exists a unique $v \in L^2(\Omega)$ such that $M_u v = w$. This equation can be written in the form

$$\begin{cases} v(x) = w(x) & \text{if } x \in \mathbb{A}_u, \\ v(x) - \frac{1}{\nu}[z_{u,v}(x)\varphi_u(x) + y_u(x)\eta_{u,v}(x)] = w(x) & \text{if } x \in \mathbb{I}_u. \end{cases}$$

Taking into account that v coincides with w in $\mathbb{A}_u$ and, hence, $v = \chi_{\mathbb{I}_u} v + \chi_{\mathbb{A}_u} w$, the second equation can be written

$$\chi_{\mathbb{I}_u} v - \frac{1}{\nu}[z_{u,\chi_{\mathbb{I}_u} v}\varphi_u + y_u\eta_{u,\chi_{\mathbb{I}_u} v}] = w + \frac{1}{\nu}[z_{u,\chi_{\mathbb{A}_u} w}\varphi_u + y_u\eta_{u,\chi_{\mathbb{A}_u} w}]. \tag{4.3}$$

In order to prove the existence and uniqueness of a solution of (4.3) we introduce the quadratic functional $\mathbb{J} : L^2(\mathbb{I}_u) \longrightarrow \mathbb{R}$ defined by

$$\begin{aligned}\mathbb{J}(v) &= \frac{1}{2\nu} J''(u)(\chi_{\mathbb{I}_u} v)^2 - \int_{\mathbb{I}_u} \Big(w + \frac{1}{\nu}[z_{u,\chi_{\mathbb{A}_u} w}\varphi_u + y_u\eta_{u,\chi_{\mathbb{A}_u} w}]\Big)v\,\mathrm{d}x \\ &= \frac{1}{2}\int_{\mathbb{I}_u} \Big[v^2 - \frac{1}{\nu}(\varphi_u z_{u\chi_{\mathbb{I}_u} v} + y_u\eta_{u,\chi_{\mathbb{I}_u} v})v\Big]\,\mathrm{d}x \\ &\quad - \int_{\mathbb{I}_u} \Big(w + \frac{1}{\nu}[z_{u,\chi_{\mathbb{A}_u} w}\varphi_u + y_u\eta_{u,\chi_{\mathbb{A}_u} w}]\Big)v\,\mathrm{d}x,\end{aligned}$$

where we have used the expression of $J''(u)$ given in Corollary 3.6. We observe that $\mathbb{J}'(v) = 0$ if and only if v satisfies (4.3). Therefore, if we prove that $\mathbb{J}$ has a unique stationary point, then the existence and uniqueness of a solution of (4.3) follows. From Theorem 3.10 we get that (3.12) holds for some $\tau > 0$ and $\kappa > 0$. Since J'' is a continuous functional in $\mathcal{A}$, we deduce the existence of $\delta_0 > 0$ such that $|[J''(u) - J''(\bar{u})]v^2| \le \frac{\kappa}{2}\|v\|^2_{L^2(\Omega)}$ for all $v \in L^2(\Omega)$ if $\|u - \bar{u}\|_{L^2(\Omega)} \le \delta_0$. This inequality and (3.12) imply that

$$J''(u)v^2 \ge \frac{\kappa}{2}\|v\|^2_{L^2(\Omega)} \quad \forall v \in T^\tau_{\bar{u}} \text{ and } \forall u \in B_{\delta_0}(\bar{u}). \tag{4.4}$$

Now, we prove that $\chi_{\mathbb{I}_u} v \in T^\tau_{\bar{u}}$ for all $v \in L^2(\mathbb{I}_u)$ if u is sufficiently close to $\bar{u}$. As a consequence we infer that the quadratic form $\mathbb{J}$ is strictly convex and coercive, which proves the existence of a unique stationary point, the unique minimizer. To prove that $\chi_{\mathbb{I}_u} v \in T^\tau_{\bar{u}}$ we select $\delta = \min\{\delta_0, \delta_{\bar{u}}, \varepsilon_{\bar{u}}, \frac{\tau}{\nu L_{S,\bar{u}}}\}$, where $\delta_{\bar{u}}$ and $L_{S,\bar{u}}$ were given in (4.2). Hence, we have that $\|S(u) - S(\bar{u})\|_{C(\bar{\Omega})} \le L_S\delta \le \frac{\tau}{\nu}$ for every $u \in B_\delta(\bar{u})$. If $\nu\bar{u}(x) - \bar{y}(x)\bar{\varphi}(x) > \tau$, then (3.7) implies that $\bar{u}(x) = \alpha$ and, hence, $\frac{1}{\nu}\bar{y}(x)\bar{\varphi}(x) < \alpha - \frac{\tau}{\nu}$. This yields $\frac{1}{\nu}y_u(x)\varphi_u(x) < \alpha$, therefore we have $x \in \mathbb{A}_u$ and $(\chi_{\mathbb{I}_u} v)(x) = 0$. Analogously we proceed if $\nu\bar{u}(x) - \bar{y}(x)\bar{\varphi}(x) < -\tau$.

It remains to deduce the existence of a constant C such that $\|M_u^{-1}\| \le C$ for every $u \in B_\delta(\bar{u})$. From (4.4), Corollary 3.6, and (4.3) we infer

$$\begin{aligned}\frac{\kappa}{2}\|\chi_{\mathbb{I}_u} v\|^2_{L^2(\Omega)} \le J''(u)(\chi_{\mathbb{I}_u} v)^2 &= \int_\Omega (\nu\chi_{\mathbb{I}_u} v - [\varphi_u z_{u,\chi_{\mathbb{I}_u} v} + y_u\eta_{u,\chi_{\mathbb{I}_u} v}])\chi_{\mathbb{I}_u} v\,\mathrm{d}x \\ &= \int_\Omega (\nu w + [\varphi_u z_{u,\chi_{\mathbb{A}_u} w} + y_u\eta_{u,\chi_{\mathbb{A}_u} w}])\chi_{\mathbb{I}_u} v\,\mathrm{d}x.\end{aligned}$$

Since $\chi_{\mathbb{A}_u} w = \chi_{\mathbb{A}_u} v$ we have

$$\nu \|\chi_{\mathbb{A}_u} v\|^2_{L^2(\Omega)} = \nu \int_\Omega w \chi_{\mathbb{A}_u} v \,\mathrm{d}x.$$

From the last two relations, the uniform estimates for y_u and φ_u in the ball $B_{\varepsilon_u}(\bar{u})$ established in the proof of Theorem 4.3, and equations (2.9) and (3.4) we get the existence of a constant C' such that

$$\begin{aligned}\min\left\{\nu, \frac{\kappa}{2}\right\}\|v\|^2_{L^2(\Omega)} &\le \nu \int_\Omega w v \,\mathrm{d}x + \int_\Omega [\varphi_u z_{u,\chi_{\mathbb{A}_u} w} + y_u \eta_{u,\chi_{\mathbb{A}_u} w}]\chi_{\mathbb{I}_u} v \,\mathrm{d}x \\ &\le C' \|w\|_{L^2(\Omega)} \|v\|_{L^2(\Omega)},\end{aligned}$$

which proves that $\|M_u^{-1}\| \le \frac{C}{\min\{\nu, \frac{\kappa}{2}\}}$.

Algorithm 2 implements the semismooth Newton method to solve (P). The following corollary establishes its convergence.

Corollary 4.6 *Let $(\bar{u}, \bar{y}, \bar{\varphi}) \in U_{\mathrm{ad}} \times [H^1(\Omega) \cap C^{0,\mu}(\bar{\Omega})]^2$ satisfy the first order optimality conditions (3.5), (3.6) and (3.7), the strict complementarity condition $|\Sigma_{\bar{u}}| = 0$, and the second order sufficient optimality condition $J''(\bar{u})v^2 > 0$ for every $v \in C_{\bar{u}} \setminus \{0\}$. Then, there exists $\delta > 0$ such that for all $u_0 \in B_\delta(\bar{u})$ the sequence $\{u_j\}$ generated by Algorithm 2 is contained in the ball $B_\delta(\bar{u})$ and converges superlinearly to $\bar{u}$.*

Proof Since any local solution of (P) satisfying the second order sufficient condition is locally the unique stationary point of (P), see [7], this result is a straightforward consequence of Theorem 4.2, Corollary 4.4, and Theorem 4.5. □

Remark 4.7 To solve the quadratic problem (Q_j) that appears in line 8 of Algorithm 2 we can use, e.g., the conjugate gradient method. Notice that we can write $\mathbb{J}_j(v) = \frac{1}{2}(v, A_j v)_{L^2(\mathbb{I}_j)} - (b_j, v)_{L^2(\mathbb{I}_j)}$, where $b_j = \chi_{\mathbb{I}_j}(w_j + \frac{1}{\nu}[z_j \varphi_j + y_j \eta_j])$ and we can compute $A_j v$ using Algorithm 3. Therefore (Q_j) can be solved without need of the explicit computation of the Hessian $J''(u_j)$.

5 A numerical example

We have used Algorithm 2 to solve the problem with the following data: $\Omega = (0, 1)^2 \subset \mathbb{R}^2$, $A = -\Delta$, $g(x) = 0$,

$$f(x, y) = y^3|y| + 2y - 100\sin(2\pi x_1)\sin(\pi x_2),$$

$\nu = 0.05$, $\alpha = -1$, $\beta = 1$, $L(x, y) = 0.5(y - y_d(x))^2$, where

$$y_d(x) = -64x_1(1 - x_1)x_2(1 - x_2).$$

Algorithm 2: Semismooth Newton method to solve (P).

Initialize. Choose $u_0 \in L^2(\Omega)$. Set $j = 0$.

for $j \geq 0$ **do**

Compute $y_j = G(u_j)$ solving the nonlinear equation

$$Ay + a(x, y) + u_j y = 0 \text{ in } \Omega,\ \partial_{\nu_A} y = g \text{ in } \Gamma.$$

Compute $\varphi_j = \Phi(u_j)$ solving the linear equation

$$A^*\varphi + \frac{\partial a}{\partial y}(x, y_j)\varphi + u_j\varphi = \frac{\partial L}{\partial y}(x, y_j) \text{ in } \Omega,\ \partial_{n_{A^*}}\varphi = 0 \text{ on } \Gamma.$$

Set $\mathbb{A}_j = \mathbb{A}_j^\beta \cup \mathbb{A}_j^\alpha$ and $\mathbb{I}_j = \Omega \setminus \mathbb{A}_j$, where

$$\mathbb{A}_j^\beta = \{x \in \Omega : y_j(x)\varphi_j(x) \geq \nu\beta\},$$

$$\mathbb{A}_j^\alpha = \{x \in \Omega : y_j(x)\varphi_j(x) \leq \nu\alpha\}.$$

Set

$$w_j(x) = -F(u_j)(x) = \begin{cases} -u_j(x) + \beta & \text{if } x \in \mathbb{A}_j^\beta \\ -u_j(x) + \frac{1}{\nu}\varphi_j(x)y_j(x) & \text{if } x \in \mathbb{I}_j \\ -u_j(x) + \alpha & \text{if } x \in \mathbb{A}_j^\alpha \end{cases}$$

Compute $z_j = z_{u_j,\chi_{\mathbb{A}_j} w_j}$ and $\eta_j = \eta_{u_j,\chi_{\mathbb{A}_j} w_j}$ solving the linear equations

$$Az_j + \frac{\partial a}{\partial y}(x, y_j)z_j + u_j z_j = -\, y_j\chi_{\mathbb{A}_j} w_j \text{ in } \Omega,\ \partial_{\nu_A} z_j = 0 \text{ on } \Gamma$$

$$A^*\eta_j + \frac{\partial a}{\partial y}(x, y_j)\eta_j + u_j\eta_j = \left(\frac{\partial^2 L}{\partial y^2}(x, y_j) - \varphi_j \frac{\partial^2 f}{\partial y^2}(x, y_j)\right) z_j - \varphi_j\chi_{\mathbb{A}_j} w_j \text{ in } \Omega,\ \partial_{n_{A^*}}\eta_j = 0 \text{ on } \Gamma$$

Solve the quadratic problem

$$(Q_j) \qquad \min_{v \in L^2(\mathbb{I}_j)} \mathbb{J}_j(v) := \frac{1}{2\nu} J''(u_j)(\chi_{\mathbb{I}_j} v)^2 - \int_{\mathbb{I}_j} (w_j + \frac{1}{\nu}[z_j\varphi_j + y_j\eta_j])v\mathrm{d}x$$

Name $v_{\mathbb{I}_j}$ its solution.

Set $v_j = \chi_{\mathbb{A}_j} w_j + \chi_{\mathbb{I}_j} v_{\mathbb{I}_j}$.

Set $u_{j+1} = u_j + v_j$ and $j = j + 1$.

end

We solve a finite element discretization of (P). Continuous piecewise linear functions are used for the state, the adjoint state, and the control. The Tichonov regularization

Algorithm 3: Computation of the product Hessian-vector

1 Solve

$$Az + \frac{\partial a}{\partial y}(x, y_j)z + u_j z = -\, y_j \chi_{\mathbb{I}_j} v \text{ in } \Omega,\ \partial_{\nu_A} z = 0 \text{ on } \Gamma$$

2 Solve

$$A^*\eta + \frac{\partial a}{\partial y}(x, y_j)\eta + u_j\eta = \left(\frac{\partial^2 L}{\partial y^2}(x, y_j) - \varphi_j \frac{\partial^2 f}{\partial y^2}(x, y_j)\right) z \\ -\, \varphi_j \chi_{\mathbb{I}_j} v \text{ in } \Omega,\ \partial_{n_{A^*}} \eta = 0 \text{ on } \Gamma$$

3 Set $A_j v = \chi_{\mathbb{I}_j}(v - \frac{1}{\nu}[z\varphi_j + \eta y_j])$

term is discretized using the lump mass matrix. In this way, the optimization algorithm for the discrete problem is exactly the discrete version of Algorithm 2.

The chosen initial point is $u_0 = 0$. The algorithm stops when

$$\delta_j = \frac{\|v_j\|_{L^2(\Omega)}}{\max\{1, \|u_{j+1}\|_{L^2(\Omega)}\}} < 5 \times 10^{-14}$$

or when $J(u_j)$ and $J(u_{j+1})$ are equal up to machine precision. At each iteration, the solution of the quadratic subproblem (Q_j) is obtained using the conjugate gradient method implemented in Matlab built-in command `pcg` and the nonlinear equation in line 3 of Algorithm 2 is solved using Newton's method. The tolerance 5×10^{-14} is used for both subproblems.

We show the convergence results for different mesh sizes in Tables 1, 2 and 3. Not only the predicted superlinear convergence can be observed in all of them, but also the mesh-independence of the convergence history, which is explained thanks to the convergence result for the algorithm in the infinite-dimensional setting.

♯Newton is the number of Newton iterations to solve the nonlinear PDE in line 3 of Algorithm 2 and ♯CG is the number of iterations of the conjugate gradient method used to solve (Q_j) in Algorithm 2. In terms of computational cost, the hard work is

Table 1 Convergence history of the problem in the example for $h = 2^{-7}$

j	$J(u_j)$	δ_j	♯Newton	♯CG
0	3.9142466314434916e+00	8.8e−01	8	13
1	3.8210815158943565e+00	8.8e−03	5	12
2	3.8210805974920747e+00	3.5e−09	3	13
3	3.821080597492071 2e+00	5.9e−15	2	14
4	3.8210805974920659e+00		1	

Table 2 Convergence history of the problem in the example for $h = 2^{-8}$

j	$J(u_j)$	δ_j	♯Newton	♯CG
0	3.9149225191422081e+00	8.8e−01	8	12
1	3.8217486072447922e+00	9.4e−03	5	12
2	3.8217477897599132e+00	4.4e−07	3	12
3	3.8217477897599528e+00	3.7e−14	2	12
4	3.8217477897599559e+00		1	

Table 3 Convergence history of the problem in the example for $h = 2^{-9}$

j	$J(u_j)$	δ_j	♯Newton	♯CG
0	3.9150915018042891e+00	8.8e−01	8	12
1	3.8219154943064222e+00	9.6e−03	5	12
2	3.8219145854336314e+00	4.1e−07	3	12
3	3.8219145854337437e+00	3.6e−14	2	13
4	3.8219145854337260e+00		1	

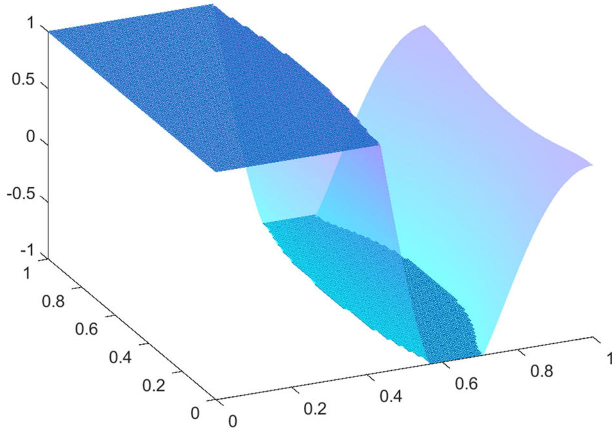

Fig. 1 Optimal control for the example

done to solve the nonlinear PDE: at each of the ♯Newton iterations we have to factorize a (sparse) matrix. The last factorization obtained at this step can be used to solve the rest of the linear PDEs that appear in Algorithm 2 and also the ones involved in the conjugate gradient method, so a good measure of the computational cost is given by the total amount of ♯Newton iterations.

A picture of the optimal control can be seen in Fig. 1. For the finest mesh, we find that $|\{x \in \Omega : \bar{u}(x) = \beta\}| = 0.459$, $|\{x \in \Omega : \bar{u}(x) = \alpha\}| = 0.233$, $|\{x \in \Omega : \alpha < \bar{u}(x) < \beta\}| = 0.308$ and $|\Sigma_{\bar{u}}| = 0$.

Funding Open Access funding provided thanks to the CRUE-CSIC agreement with Springer Nature.

References

1. Alibert, J.J., Raymond, J.P.: Boundary control of semilinear elliptic equations with discontinuous leading coefficients and unbounded controls. Numer. Funct. Anal. Optim. **18**(3–4), 235–250 (1997). https://doi.org/10.1080/01630569708816758
2. Amstutz, S., Laurain, A.: A semismooth Newton method for a class of semilinear optimal control problems with box and volume constraints. Comput. Optim. Appl. **56**(2), 369–403 (2013). https://doi.org/10.1007/s10589-013-9555-6
3. Bruni, C., DiPillo, G., Koch, G.: Bilinear systems: an appealing class of 'nearly linear' systems in theory and applications. IEEE Trans. Autom. Control **19**, 334–348 (1974). https://doi.org/10.1109/TAC.1974.1100617
4. Casas, E.: Boundary control of semilinear elliptic equations with pointwise state constraints. SIAM J. Control Optim. **31**(4), 993–1006 (1993). https://doi.org/10.1137/0331044
5. Casas, E., Fernández, L.A.: A Green's formula for quasilinear elliptic operators. J. Math. Anal. Appl. **142**(1), 62–73 (1989). https://doi.org/10.1016/0022-247X(89)90164-9
6. Casas, E., Mateos, M.: Convergence analysis of the semismooth newton method for sparse control problems governed by semilinear elliptic equations. SIAM J. Control Optim. **62**(6), 3076–3090 (2024). https://doi.org/10.1137/23M1585945
7. Casas, E., Tröltzsch, F.: Second order analysis for optimal control problems: improving results expected from abstract theory. SIAM J. Optim. **22**(1), 261–279 (2012). https://doi.org/10.1137/110840406
8. Clarke, F.H.: Optimization and nonsmooth analysis. John Wiley & Sons, New York (1983)
9. Clason, C., Jin, B.: A semismooth Newton method for nonlinear parameter identification problems with impulsive noise. SIAM J. Imaging Sci. **5**(2), 505–538 (2012). https://doi.org/10.1137/110826187
10. Glowinski, R., Song, Y., Yuan, X., Yue, H.: Bilinear optimal control of an advection-reaction-diffusion system. SIAM Rev. **64**(2), 392–421 (2022). https://doi.org/10.1137/21M1389778
11. Hintermüller, M., Ito, K., Kunisch, K.: The primal-dual active set strategy as a semismooth Newton method. SIAM J. Optim. **13**(3), 865–888 (2003). https://doi.org/10.1137/S1052623401383558
12. Hinze, M., Pinnau, R., Ulbrich, M., Ulbrich, S.: Optimization with PDE constraints. Springer, Dordrecht (2009)
13. Ito, K., Kunisch, K.: Lagrange multiplier approach to variational problems and applications. Soc. Industr. Appl. Math. (SIAM) (2008). https://doi.org/10.1137/1.9780898718614
14. Mannel, F., Rund, A.: A hybrid semismooth quasi-Newton method for nonsmooth optimal control with PDEs. Optim. Eng. **22**(4), 2087–2125 (2021). https://doi.org/10.1007/s11081-020-09523-w
15. Mohler, R.R.: Bilinear control processes: with applications to engineering, ecology, and medicine. Elsevier, Amsterdam (1973)
16. Nittka, R.: Regularity of solutions of linear second order elliptic and parabolic boundary value problems on Lipschitz domains. J. Differ. Equ. **251**(4–5), 860–880 (2011). https://doi.org/10.1016/j.jde.2011.05.019
17. Nocedal, J., Wright, S.J.: Numerical optimization. Springer-Verlag, New York (1999)
18. Pieper, K.: Finite element discretization and efficient numerical solution of elliptic and parabolic sparse control problems. Ph.D. thesis, Technische Universität München (2015). https://nbn-resolving.de/urn/resolver.pl?urn:nbn:de:bvb:91-diss-20150420-1241413-1-4
19. Reyes, J.C., Kunisch, K.: A semi-smooth Newton method for control constrained boundary optimal control of the Navier-Stokes equations. Nonlinear Anal. **62**(7), 1289–1316 (2005). https://doi.org/10.1016/j.na.2005.04.035

20. Stampacchia, G.: Le problème de Dirichlet pour les équations elliptiques du second ordre à coefficients discontinus. Ann. Inst. Fourier **15**(1), 189–257 (1965). https://doi.org/10.5802/aif.204
21. Tröltzsch, F.: Optimal control of partial differential equations: theory, methods and applications. American Mathematical Society, Providence, RI (2010)
22. Ulbrich, M.: Constrained optimal control of Navier-Stokes flow by semismooth Newton methods. Syst. Control Lett. **48**(3), 297–311 (2003). https://doi.org/10.1016/S0167-6911(02)00274-8
23. Ulbrich, M.: Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces, *MOS/SIAM Ser. Optim.*, vol. 11. Philadelphia, PA: Society for Industrial and Applied Mathematics (2011)