SIAM J. CONTROL OPTIM.
Vol. 62, No. 6, pp. 3076–3090

CONVERGENCE ANALYSIS OF THE SEMISMOOTH NEWTON METHOD FOR SPARSE CONTROL PROBLEMS GOVERNED BY SEMILINEAR ELLIPTIC EQUATIONS*

EDUARDO CASAS† AND MARIANO MATEOS‡

Abstract. We show that a second order sufficient condition for local optimality, along with a strict complementarity condition, is enough to get the superlinear convergence of the semismooth Newton method for an optimal control problem governed by a semilinear elliptic equation. The objective functional may include a sparsity promoting term and we allow for box control constraints.

Key words. optimal control, semilinear elliptic equations, semismooth Newton method

MSC codes. 35J61, 49K20, 49M15, 49M05

DOI. 10.1137/23M1585945

1. Introduction.

Let us consider a domain $\Omega \subset \mathbb{R}^n$, $n \leq 3$, with a Lipschitz boundary Γ. We study the following problem:

$$\text{(P)} \qquad \min_{u \in U_{\rm ad}} J(u) := F(u) + \gamma j(u),$$

where

$$F(u) = \int_\Omega L(x, y_u(x))\,\mathrm{d}x + \frac{\kappa}{2}\int_\Omega u(x)^2\,\mathrm{d}x \text{ and } j(u) = \int_\Omega |u(x)|\,\mathrm{d}x.$$

Here $L : \Omega \times \mathbb{R} \to \mathbb{R}$ is a given function, $\kappa > 0$, $\gamma \geq 0$, and

$$U_{\rm ad} = \{u \in L^2(\Omega) : \alpha \leq u(x) \leq \beta \text{ for a.a. } x \in \Omega\},$$

with $-\infty \leq \alpha < \beta \leq \infty$. If $\gamma > 0$, we will further suppose $\alpha < 0 < \beta$.

Above y_u denotes the state associated to the control u related by the following semilinear elliptic state equation:

$$(1.1) \qquad \left\{ \begin{array}{rcll} Ay_u + f(x, y_u) & = & u & \text{in } \Omega, \\ y_u & = & 0 & \text{on } \Gamma. \end{array} \right.$$

Assumptions on the data A, f, L are specified in section 2.

To introduce the main result of the paper and put it in the context of related results in the literature, we briefly describe the semismooth Newton method; precise definitions will be introduced in section 3. Let $\bar{u}$ be a solution of the equation $\Phi(u) = 0$, where Φ is a semismooth function. Given u_k, at every step we select $M_k \in \partial\Phi(u_k)$, the generalized derivative of Φ at u_k, we compute the solution of the linear system

*Received by the editors July 11, 2023; accepted for publication (in revised form) August 14, 2024; published electronically November 19, 2024.
https://doi.org/10.1137/23M1585945
Funding: The authors were partially supported by MCIN/AEI/10.13039/501100011033 under research project PID2020-114837GB-I00.
†Departamento de Matemática Aplicada y Ciencias de la Computación, E.T.S.I. Industriales y de Telecomunicación, Universidad de Cantabria, 39005 Santander, Spain (eduardo.casas@unican.es).
‡Departamento de Matemáticas, Campus de Gijón, Universidad de Oviedo, 33203, Gijón, Spain (mmateos@uniovi.es).

$M_k v_k = -\Phi(u_k)$, and set $u_{k+1} = u_k + v_k$. We have that u_k converges superlinearly to $\bar u$ provided that u_0 is close enough to $\bar u$ and the inverses of the operators M_k exist and are uniformly bounded.

In the case of linear equations and convex objective functionals, the uniform boundedness is obtained assuming the existence of $\nu > 0$ such that $F''(\bar u)v^2 \geq \nu\|v\|^2_{L^2(\Omega)}$ for all $v \in L^2(\Omega)$; see [7, 10, 13]. While this assumption is fully justified in that case, it is too restrictive if the equation is not linear because it is too far from the second order necessary condition $F''(\bar u) \geq 0$ for all $v \in C_{\bar u}$, the cone of critical directions. As far as we know, the only papers dealing with the convergence of the semismooth Newton's method for optimal control problems governed by nonlinear equations are [1], [8], and [10]. In the last two references, the proof of the convergence is done assuming the abovementioned condition, while in the first one a condition implying convexity of the functional is done.

The goal, and the novelty, of our paper is the proof of the superlinear convergence of the semismooth Newton method toward a local solution $\bar u$ of (P) assuming a strict complementarity condition, to be properly established in Definition 2.11, along with a sufficient second order condition for local optimality. The sufficient second order condition is the usual one enjoying a minimal gap with respect to the necessary one. In Theorem 3.4, we prove that these two hypotheses imply the uniform boundedness of the inverses of the selected generalized derivatives. A strict complementary assumption together with a second order sufficient condition are the usual hypotheses to prove the superlinear convergence of numerical algorithms in finite dimensional constrained optimization problems; cf. [11, 12], and see also [9, Chapters 17 and 18] and the references therein. We notice (cf. Remark 3.6) that the strict complementarity condition can be dropped out by assuming the stronger second order sufficient optimality condition (3.6); cf. [14, equation (3.19)].

The plan of the paper is as follows. In section 2 we introduce the assumptions on the control problem and carry out the first and second order analysis. The convergence of the semismooth Newton algorithm is proved in section 3. In the last section, we describe some computational details and present two numerical examples.

2. Assumptions and first and second order analysis of the control problem.

We make the following assumptions on the data of the control problem.

(A1) Throughout the paper, Ω is a bounded open subset of $\mathbb{R}^n$, $1 \leq n \leq 3$. If $n = 2$ or 3 we assume that its boundary Γ is Lipschitz. If $n = 1$, Ω is a bounded interval and Γ is reduced to the two end points of the interval. The operator A is defined in Ω by the expression

$$Ay = -\sum_{i,j=1}^n \partial_{x_j}[a_{ij}\partial_{x_i}y] + a_0 y$$

with $a_0, a_{i,j} \in L^\infty(\Omega)$ for $1 \leq i,j \leq n$, $a_0 \geq 0$, and there exists $\Lambda > 0$ such that

$$\sum_{i,j=1}^n a_{i,j}(x)\xi_i\xi_j \geq \Lambda|\xi|^2 \text{ for a.e. } x \in \Omega \text{ and all } \xi \in \mathbb{R}^n.$$

(A2) We assume that $f : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function of class C^2 with respect to the second variable satisfying the following conditions for almost all $x \in \Omega$:

- $\exists \bar{p} > \dfrac{n}{2}$ such that $f(\cdot,0) \in L^{\bar{p}}(\Omega)$,
- $\dfrac{\partial f}{\partial y}(x,y) \geq 0 \quad \forall y \in \mathbb{R}$,
- $\forall M > 0,\ \exists C_M > 0$ such that $\sum_{j=1}^{2} |\frac{\partial^j f}{\partial y^j}(x,y)| \leq C_{f,M}\ \forall\ |y| \leq M$,
- $\forall \varepsilon > 0$ and $\forall M > 0\ \exists \delta > 0$ such that $|\frac{\partial^2 f}{\partial y^2}(x,y_1) - \frac{\partial^2 f}{\partial y^2}(x,y_2)| \leq \varepsilon$
  $\forall\ |y_1|,|y_2| \leq M$ with $|y_1 - y_2| \leq \delta$.

(A3) For the cost functional we suppose that $L : \Omega \times \mathbb{R} \longrightarrow \mathbb{R}$ is a Carathéodory function of class C^2 with respect to the second variable satisfying the following conditions for almost all $x \in \Omega$:

- $L(\cdot,0) \in L^1(\Omega)$ and $\forall M > 0\ \exists \Psi_{L,M} \in L^{\bar{p}}(\Omega)$ and $C_{L,M} > 0$ such that
$$\left|\frac{\partial L}{\partial y}(x,y)\right| \leq \Psi_{L,M}(x) \quad \text{and} \quad \left|\frac{\partial^2 L}{\partial y^2}(x,y)\right| \leq C_{L,M}\ \forall\ |y| \leq M,$$
- $\forall \varepsilon > 0$ and $\forall M > 0,\ \exists \delta > 0$ such that $|\frac{\partial^2 L}{\partial y^2}(x,y_1) - \frac{\partial^2 L}{\partial y^2}(x,y_2)| \leq \varepsilon$
  $\forall\ |y_1|,|y_2| \leq M$ with $|y_1 - y_2| < \delta$.

Let us consider the Banach space $Y = H_0^1(\Omega) \cap C(\bar{\Omega})$. Under the above assumptions, the following properties are well known; see, for instance, [4, Theorem 1.1.2].

THEOREM 2.1. *For any $u \in L^p(\Omega)$ with $p > n/2$, there exists a unique solution of (1.1) $y_u \in Y$. Moreover, there exists a constant $K > 0$ that depends on A, Ω, p, and $\bar{p}$ such that*

$$\|y_u\|_{H_0^1(\Omega)} + \|y_u\|_{C(\bar{\Omega})} \leq K(\|u\|_{L^p(\Omega)} + \|f(\cdot,0)\|_{L^{\bar{p}}(\Omega)})$$

holds. The mapping $S : L^p(\Omega) \longrightarrow Y$ given by $S(u) = y_u$ is of class C^2. Furthermore, for all $u, v \in L^p(\Omega)$, $z_v = S'(u)v$ is the unique solution to

$$\begin{cases} Az + \dfrac{\partial f}{\partial y}(x,y_u)z = v & \text{in } \Omega, \\ z = 0 & \text{on } \Gamma, \end{cases}$$

and, given $v_1, v_2 \in L^p(\Omega)$, $z_{v_1,v_2} = S''(u)(v_1,v_2)$ is the unique solution to

$$\begin{cases} Az + \dfrac{\partial f}{\partial y}(x,y_u)z = -\dfrac{\partial^2 f}{\partial y^2}(x,y_u)z_{v_1}z_{v_2} & \text{in } \Omega, \\ w = 0 & \text{on } \Gamma, \end{cases}$$

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

For later reference, it will be useful to define the adjoint state in the following form. We consider the mapping $T : L^\infty(\Omega) \longrightarrow Y$ such that $\varphi = T(y)$ is the unique solution to the adjoint state equation:

$$\begin{cases} A^*\varphi + \dfrac{\partial f}{\partial y}(x,y)\varphi = \dfrac{\partial L}{\partial y}(x,y) & \text{in } \Omega, \\ \varphi = 0 & \text{on } \Gamma. \end{cases}$$

Setting $G = T \circ S$, we have that the adjoint state related to u is given by $\varphi_u = G(u)$. From Theorem 2.1 and the chain rule, it is straightforward to deduce the following two results.

THEOREM 2.2. *For every $p > n/2$, the mapping $G : L^p(\Omega) \to Y$ is of class C^1 and for every $u, v \in L^p(\Omega)$, $\eta_v = G'(u)v$ is the unique solution of*

$$\left\{ \begin{array}{ll} A^*\eta_v + \dfrac{\partial f}{\partial y}(x, y_u)\eta_v = \left(\dfrac{\partial^2 L}{\partial y^2}(x, y_u) - \varphi_u \dfrac{\partial^2 f}{\partial y^2}(x, y_u)\right) z_v & \text{in } \Omega, \\ \eta_v = 0 & \text{on } \Gamma. \end{array} \right. \tag{2.1}$$

THEOREM 2.3. *The functional $F : L^2(\Omega) \to \mathbb{R}$ is of class C^2. Further, for every $u, v, v_1, v_2 \in L^2(\Omega)$ the following identities hold:*

$$F'(u)v = \int_\Omega (\varphi_u + \kappa u)v \, \mathrm{d}x, \tag{2.2}$$

$$F''(u)(v_1, v_2) = \int_\Omega \left\{ \left(\frac{\partial^2 L}{\partial y^2}(x, y_u) - \varphi_u \frac{\partial^2 f}{\partial y^2}(x, y_u)\right) z_{v_1} z_{v_2} + \kappa v_1 v_2 \right\} \mathrm{d}x \tag{2.3}$$

$$= \int_\Omega (\eta_{v_1} + \kappa v_1)v_2 \, \mathrm{d}x = \int_\Omega (\eta_{v_2} + \kappa v_2)v_1 \, \mathrm{d}x, \tag{2.4}$$

where $\varphi_u = G(u)$, $z_{v_i} = G'(u)v_i$, and $\eta_{v_i} = G'(u)v_i$ for $i = 1, 2$.

We will need some results about the adjoint states, which we gather in the next lemma.

LEMMA 2.4. *Given $R > 0$ arbitrary, we denote by $\bar{B}_R(0)$ the closed $L^2(\Omega)$-ball centered at 0 with radius R. There exists a constant $K_{G'}(R) > 0$ such that*

$$\|G'(u)v\|_Y \le K_{G'}(R)\|v\|_{L^2(\Omega)} \quad \forall u \in \bar{B}_R(0) \text{ and } \forall v \in L^2(\Omega), \tag{2.5}$$

$$\|G(u_1) - G(u_2)\|_Y \le K_{G'}(R)\|u_1 - u_2\|_{L^2(\Omega)} \quad \forall u_1, u_2 \in \bar{B}_R(0). \tag{2.6}$$

Proof. Let us prove (2.5). From Theorem 2.1, we deduce the existence of a constant $M(R) > 0$ such that $\|y_u\|_{C(\bar{\Omega})} \le M(R)$ for every $u \in \bar{B}_R(0)$. Moreover, from the monotonicity of f we deduce the existence of a constant C_1 such that

$$\|z_v\|_Y \le C_1\|v\|_{L^2(\Omega)} \quad \forall u \in \bar{B}_R(0) \text{ and } \forall v \in L^2(\Omega). \tag{2.7}$$

We also obtain with assumption (A3)

$$\|\varphi_u\|_Y \le C_1 \left\| \frac{\partial L}{\partial y}(\cdot, y_u) \right\|_{L^2(\Omega)} \le C_1\|\Psi_{L,M(R)}\|_{L^2(\Omega)} \le C_R \ \forall u \in \bar{B}_R(0). \tag{2.8}$$

Once again, from (2.1) and using (2.7) and (2.8) along with the assumptions (A2) and (A3) we get

$$\|G'(u)v\|_Y = \|\eta_v\|_Y \le K_{G'}(R)\|v\|_{L^2(\Omega)} \quad \forall u \in \bar{B}_R(0) \text{ and } \forall v \in L^2(\Omega).$$

Thus, (2.5) follows. Estimate (2.6) is readily deduced from (2.5) and the generalized mean value theorem. ∎

Let us also remark that $j(u) = \|u\|_{L^1(\Omega)}$ is convex and Lipschitz. For every $u, v \in L^1(\Omega)$, the directional derivative $j'(u; v)$ is given by

$$j'(u; v) = \int_{\Omega_u^+} v \, \mathrm{d}x - \int_{\Omega_u^-} v \, \mathrm{d}x + \int_{\Omega_u^0} |v| \, \mathrm{d}x, \tag{2.9}$$

where Ω_u^+, Ω_u^-, and Ω_u^0 are the sets of points where u is respectively positive, negative, or zero. We denote $J'(u; v) = F'(u)v + \gamma j'(u; v)$ for every $v \in L^2(\Omega)$. The subdifferential of j at u is given by

$$\partial j(u) = \left\{ \lambda \in L^\infty(\Omega) : \lambda(x) \in \begin{cases} \{+1\} & \text{if } u(x) > 0, \\ \{-1\} & \text{if } u(x) < 0, \\ [-1, 1] & \text{if } u(x) = 0. \end{cases} \right\} \tag{2.10}$$

A local solution of (P) is intended in the $L^2(\Omega)$-sense along this paper. In the following theorem, we summarize necessary and sufficient conditions for local optimality. First, we define the cone of critical directions by

$$C_{\bar u} = \{v \in L^2(\Omega) : \text{satisfying (2.11) and } J'(\bar u)v + \gamma j'(\bar u; v) = 0\},$$

where

$$\begin{cases} v(x) \geq 0 \text{ if } \bar u(x) = \alpha, \\ v(x) \leq 0 \text{ if } \bar u(x) = \beta. \end{cases} \tag{2.11}$$

THEOREM 2.5. *Suppose $\bar u \in U_{ad}$ is a local solution of* (P). *Then, the following conditions hold:*

$$J'(\bar u; u - \bar u) \geq 0 \ \forall u \in U_{ad}, \tag{2.12}$$

$$\exists \bar\lambda \in \partial j(\bar u) \text{ such that } \int_\Omega (\bar\varphi + \kappa \bar u + \gamma \bar\lambda)(u - \bar u) \geq 0 \ \forall u \in U_{ad}, \tag{2.13}$$

$$F''(\bar u)v^2 \geq 0 \ \forall v \in C_{\bar u}, \tag{2.14}$$

where $\bar\varphi = G(\bar u)$. Conversely, suppose that $(\bar u, \bar\lambda) \in U_{ad} \times \partial j(\bar u)$ satisfies (2.13) *and*

$$F''(\bar u)v^2 > 0 \ \forall \ v \in C_{\bar u} \setminus \{0\}. \tag{2.15}$$

Then, there exist $\nu > 0$ and $\delta > 0$ such that

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

The reader is referred to [3, Theorems 3.1, 3.7, and 3.9] for its proof. Notice that the gap between the sufficient condition (2.15) and the necessary condition (2.14) is the minimal one, the same as in finite dimensional optimization.

We quote the following result, whose proof can be found in [3, Corollary 3.2].

COROLLARY 2.6. *Let $(\bar u, \bar\varphi, \bar\lambda) \in U_{ad} \times Y \times \partial j(\bar u)$ satisfy* (2.13) *with $\bar\varphi = G(\bar u)$. Then, the following relation holds:*

$$\bar u(x) = \text{Proj}_{[\alpha,\beta]}\left(-\frac{1}{\kappa}\big(\bar\varphi(x) + \gamma\bar\lambda(x)\big)\right). \tag{2.16}$$

Moreover, if $\gamma > 0$ the following properties are fulfilled:

$$\bar u(x) = 0 \iff |\bar\varphi(x)| \leq \gamma, \tag{2.17}$$

$$\bar\lambda(x) = \text{Proj}_{[-1,+1]}\left(-\frac{1}{\gamma}\varphi(x)\right). \tag{2.18}$$

Remark 2.7. Notice that if $\gamma = 0$, the role of $\bar\lambda$ in (2.13) and (2.16) is irrelevant. Since, nevertheless, the notation is consistent, we leave it there in order to make an exposition as unified as possible of both cases, $\gamma = 0$ and $\gamma > 0$.

Remark 2.8. As an immediate consequence of (2.16) we obtain the following:

(2.19) If $\bar u(x)=\alpha$ then $\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)\geq 0$.

(2.20) If $\bar u(x)=\beta$ then $\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)\leq 0$.

(2.21) If $\alpha<\bar u(x)<\beta$ then $\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)=0$.

(2.22) If $\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)>0$ then $\bar u(x)=\alpha$.

(2.23) If $\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)<0$ then $\bar u(x)=\beta$.

Using (2.2), (2.9), and (2.10) we infer

(2.24)
$$J'(\bar u;v)=\int_{\Omega^+_{\bar u}\cup\Omega^-_{\bar u}}[\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)]v(x)\,\mathrm{d}x+\int_{\Omega^0_{\bar u}}[\bar\varphi(x)v(x)+\gamma|v(x)|]\,\mathrm{d}x.$$

The next lemma establishes an important property of the elements of the critical cone.

Lemma 2.9. *Let $(\bar u,\bar\varphi,\bar\lambda)$ be as in Corollary* 2.6. *Then, the following property holds for $v\in L^2(\Omega)$ and for almost all $x\in\Omega$:*

$$[\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)]v(x)\begin{cases}\geq 0 & \text{if } v \text{ satisfies (2.11)},\\ =0 & \text{if } v\in C_{\bar u}.\end{cases}\tag{2.25}$$

Proof. The inequality of (2.25) is a straightforward consequence of (2.19)–(2.21) and (2.11). To prove the equality of (2.25) we recall that $j'(\bar u;v)\geq\int_\Omega\lambda v\,\mathrm{d}x$ for all $\lambda\in\partial j(\bar u)$. Then, we get

$$0=J'(\bar u;v)\geq\int_\Omega[\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)]v(x)\,\mathrm{d}x\quad\forall v\in C_{\bar u}.$$

Since the integrand is nonnegative for almost all $x\in\Omega$, the above inequality yields $[\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)]v(x)=0$ for almost all $x\in\Omega$. ∎

Lemma 2.10. *Let $(\bar u,\bar\varphi,\bar\lambda)$ be as in Corollary* 2.6. *Then, $C_{\bar u}$ is the set of elements $v\in L^2(\Omega)$ satisfying the following conditions:*

$$v(x)=0 \text{ if } \bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)\neq 0 \text{ or } |\bar\varphi(x)|<\gamma,\tag{2.26}$$

$$\begin{cases} v(x)\geq 0 \text{ if } \bar u(x)=\alpha \text{ or } \bar\varphi(x)=-\gamma,\\ v(x)\leq 0 \text{ if } \bar u(x)=\beta \text{ or } \bar\varphi(x)=+\gamma,\end{cases}\tag{2.27}$$

where the terms involving γ should be removed if $\gamma=0$.

Proof. From (2.17) we infer that $\bar\varphi(x)v(x)+\gamma|v(x)|\geq 0$ for almost all $x\in\Omega^0_{\bar u}$. Using this, the inequality in (2.25), and (2.24) we deduce that $J'(\bar u;v)=0$ if and only if $[\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)]v(x)=0$ a.e. in $\Omega^+_{\bar u}\cup\Omega^-_{\bar u}$ and $\bar\varphi(x)v(x)+\gamma|v(x)|=0$ a.e. in $\Omega^0_{\bar u}$. The last equality holds if and only if ($v(x)=0$ if $|\bar\varphi(x)|<\gamma$), ($v(x)\geq 0$ if $\bar\varphi(x)=-\gamma$), and ($v(x)\leq 0$ if $\bar\varphi(x)=+\gamma$). These equivalences prove the characterization of $C_{\bar u}$ given in the statement of the lemma. ∎

Now, we define the closed vector subspace of $L^2(\Omega)$

$$T_{\bar u}=\{v\in L^2(\Omega):\ v(x)=0 \text{ if } |\bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)|>0 \text{ or } |\bar\varphi(x)|<\gamma\}$$

and the set

$$\Sigma_{\bar u}=\{x\in\Omega:\ (\bar u(x)\in\{\alpha,\beta\} \text{ and } \bar\varphi(x)+\kappa\bar u(x)+\gamma\bar\lambda(x)=0) \text{ or } |\bar\varphi(x)|=\gamma\}.$$

Once again, the terms involving γ should be removed in the case $\gamma=0$.

DEFINITION 2.11. *We say that the strict complementary condition is satisfied at* $\bar u$ *if* $|\Sigma_{\bar u}| = 0$, *where* $|\cdot|$ *stands for the Lebesgue measure.*

LEMMA 2.12. *Let* $(\bar u, \bar\varphi, \bar\lambda)$ *be as in Corollary* 2.6 *and assume that the strict complementary condition holds at* $\bar u$, *then* $C_{\bar u} = T_{\bar u}$.

This lemma is an immediate consequence of Lemma 2.10 and the fact that $|\Sigma_{\bar u}| = 0$.

Given $\tau > 0$, where $\tau < \gamma$ if $\gamma > 0$, we define the extended subspace

$$T^\tau_{\bar u} = \{v \in L^2(\Omega) : v(x) = 0 \text{ if } |\bar\varphi(x) + \kappa\bar u(x) + \gamma\bar\lambda(x)| > \tau \ \text{ or } \ |\bar\varphi(x)| < \gamma - \tau\}.$$

THEOREM 2.13. *Let* $(\bar u, \bar\varphi, \bar\lambda) \in U_{ad} \times Y \times \partial j(\bar u)$ *satisfy* (2.13) *with* $\bar\varphi = G(\bar u)$ *and assume that the strict complementary condition* $|\Sigma_{\bar u}| = 0$ *and the second order sufficient condition* (2.15) *hold at* $\bar u$. *Then, there exist* $\nu > 0$ *and* $\tau > 0$, *with* $\tau < \gamma$ *if* $\gamma > 0$, *such that*

$$F''(\bar u)v^2 \ge \nu \|v\|^2_{L^2(\Omega)} \ \forall\ v \in T^\tau_{\bar u}. \tag{2.28}$$

Proof. We will proceed by contradiction: suppose (2.28) is false. Then, there exists a sequence $\{v_k\}_{k=1}^\infty \subset L^2(\Omega)$ such that $v_k \in T^{1/k}_{\bar u}$ and $F''(\bar u)v_k^2 < \frac{1}{k}\|v_k\|^2_{L^2(\Omega)}$. Of course, we can assume that $\|v_k\|_{L^2(\Omega)} = 1$; otherwise it is enough to divide v_k by its $L^2(\Omega)$-norm to have

$$v_k \in T^{1/k}_{\bar u}, \ \|v_k\|_{L^2(\Omega)} = 1, \text{ and } F''(\bar u)v_k^2 < \frac{1}{k}. \tag{2.29}$$

Then, for a subsequence, denoted in the same way, there exists $v \in L^2(\Omega)$ such that $v_k \rightharpoonup v$ weakly in $L^2(\Omega)$. We observe that $v \in T_{\bar u}$. Indeed, for every $\varepsilon > 0$ we set

$$\Theta^\varepsilon = \{x \in \Omega : v(x) \neq 0 \ \text{ and } \ |\bar\varphi(x) + \kappa\bar u(x) + \gamma\bar\lambda(x)| > \varepsilon \text{ or } \ |\bar\varphi(x)| < \gamma - \varepsilon\}.$$

Because v_k vanishes in Θ^ε for every $k > \frac{1}{\varepsilon}$, its weak limit v vanishes as well in Θ^ε. As $\varepsilon > 0$ is arbitrary, we conclude that $v \in T_{\bar u}$. On the other hand, since the quadratic form $F''(\bar u) : L^2(\Omega) \to \mathbb{R}$ is weakly lower semicontinuous, using (2.29) we have that

$$F''(\bar u)v^2 \le \liminf_{k\to\infty} F''(\bar u)v_k^2 = 0.$$

The strict complementarity condition implies that $C_{\bar u} = T_{\bar u}$. Therefore, as a consequence of (2.15), we deduce that $v = 0$. Moreover, the weak convergence $v_k \rightharpoonup 0$ in $L^2(\Omega)$ implies the strong convergence $z_{v_k} \to 0$ in $C(\bar\Omega)$. Using that $\|v_k\|_{L^2(\Omega)} = 1$ we obtain

$$\lim_{k\to\infty} F''(\bar u)v_k^2 = \lim_{k\to\infty} \int_\Omega \left(\frac{\partial^2 L}{\partial y^2}(x, \bar y) - \bar\varphi \frac{\partial^2 f}{\partial y^2}(x, \bar y)\right) z_{v_k}^2 \,\mathrm{d}x + \kappa = \kappa,$$

which contradicts the fact that $\kappa > 0$. ∎

Remark 2.14. In [14, equation (3.19)], the author makes an assumption similar to (2.28) to prove quadratic convergence for a sequential quadratic programming algorithm. However, (2.28) sounds quite strong as an assumption because it seems to be very far from the second order necessary condition. Theorem 2.13 shows that it is satisfied whenever the no gap second order sufficient condition plus the strict complementarity condition hold.

3. Semismooth Newton method. Next we use (2.16) and (2.18) to define an equation $\Phi(u)=0$ satisfied by any local solution of (P), where Φ is semismooth. We define semismoothness following [15, Definition 3.1]. A slightly different approach using the concept of slant differentiability can be found in [7].

DEFINITION 3.1. *Given two Banach spaces X and Y, an open subset V of X, a continuous function $\Phi : V \to Y$, and a set-valued mapping $\partial\Phi : V \rightrightarrows \mathcal{L}(X,Y)$ such that $\partial\Phi(u) \neq \emptyset$ for every $u \in V$, we say that Φ is $\partial\Phi$-semismooth at $\bar u \in V$ if*

$$\lim_{v\to 0} \sup_{M\in\partial\Phi(\bar u+v)} \frac{\|\Phi(\bar u+v)-\Phi(\bar u)-Mv\|_Y}{\|v\|_X} = 0. \tag{3.1}$$

The multifunction $\partial\Phi$ is called the generalized derivative of Φ.

The semismooth Newton method spans a sequence according to Algorithm 1.

The proof of the following convergence theorem can be found in [15, Theorem 3.13]. See also [7, Theorem 1.1].

THEOREM 3.2. *Suppose that $\Phi : V \to Y$ is $\partial\Phi$-semismooth at $\bar u \in V$ solution of $\Phi(u)=0$ locally unique. Suppose, furthermore, that the following regularity condition is satisfied: for every j, the operator $M_j \in \partial\Phi(u_j)$ is invertible and there exists $C_\Phi > 0$ such that*

$$\|M_j^{-1}\|_{\mathcal{L}(Y,X)} \le C_\Phi \ \ \forall j \ge 0. \tag{3.2}$$

Then, there exists $\delta > 0$ such that for all $u_0 \in V$ with $\|u_0-\bar u\|_X < \delta$ the sequence $\{u_j\}_{j\ge 0}$ spanned by the semismooth Newton method converges superlinearly to $\bar u$.

Taking into account (2.16) and (2.18), we define $\psi:\mathbb{R}\to\mathbb{R}$ as

$$\psi(t) = \mathrm{Proj}_{[\alpha,\beta]}\left\{-\frac{1}{\kappa}\left[t+\mathrm{Proj}_{[-\gamma,+\gamma]}(-t)\right]\right\}$$

and the superposition operator $\Psi_G : L^2(\Omega) \to L^2(\Omega)$ by $\Psi_G(u)(x) = \psi(G(u)(x))$. We recall that $G(u) = \varphi_u$. We consider the mapping $\Phi : L^2(\Omega) \to L^2(\Omega)$ given by $\Phi(u) = u - \Psi_G(u)$. Corollary 2.6 implies that $\bar u$ satisfies the equation $\Phi(\bar u) = 0$. Next, we study the semismoothness properties of Φ. Hereafter, $\partial^{\mathrm{CL}}\psi$ denotes the generalized derivative of ψ in the sense of Clarke [6, Definition 3.10]. We observe that ψ is a Lipschitz function.

LEMMA 3.3. *The function $\Phi : L^2(\Omega) \to L^2(\Omega)$ is $\partial\Phi$-semismooth at every $u \in L^2(\Omega)$ for the set-valued mapping*

$$\partial\Phi(u) = \{M = I - N : \ N \in \partial\Psi_G(u)\},$$

where

$$\partial\Psi_G(u) = \{N \in \mathcal{L}(L^2(\Omega),L^2(\Omega)) : \text{there exists a Lebesgue measurable function } h \text{ such that } h(x) \in \partial^{\mathrm{CL}}\psi(G(u)(x)) \text{ a.e. in } \Omega \text{ and } Nv = h\cdot G'(u)v \ \forall v \in L^2(\Omega)\}.$$

Algorithm 1: Semismooth Newton method.

Initialize. Choose $u_0 \in V$. Set $j = 0$.
repeat
  Choose $M_j \in \partial\Phi(u_j)$ and solve $M_j v_j = -\Phi(u_j)$.
  Set $u_{j+1} = u_j + v_j$ and $j = j+1$.
until *convergence*

Proof. Clearly, Φ is continuous for being the composition of continuous functions.

It is straightforward to check that the generalized derivative of ψ in the sense of Clarke (see [6, Theorem 10.27]) is given by the following expression:

$$\partial^{\mathrm{CL}}\psi(t)=\begin{cases}\{0\} & \text{if } t\in(-\infty,-\gamma-\kappa\beta)\cup(-\gamma,\gamma)\cup(\gamma-\kappa\alpha,+\infty),\\ \left\{-\dfrac{1}{\kappa}\right\} & \text{if } t\in(-\gamma-\kappa\beta,-\gamma)\cup(\gamma,\gamma-\kappa\alpha),\\ \left[-\dfrac{1}{\kappa},0\right] & \text{if } t\in\{-\gamma-\kappa\beta,-\gamma,\gamma,\gamma-\kappa\alpha\},\end{cases}\qquad \text{if } \gamma>0,$$

$$\partial^{\mathrm{CL}}\psi(t)=\begin{cases}\{0\} & \text{if } t\in(-\infty,-\kappa\beta)\cup(-\kappa\alpha,+\infty),\\ \left\{-\dfrac{1}{\kappa}\right\} & \text{if } t\in(-\kappa\beta,-\kappa\alpha),\\ \left[-\dfrac{1}{\kappa},0\right] & \text{if } t\in\{-\kappa\beta,-\kappa\alpha\},\end{cases}\qquad \text{if } \gamma=0.$$

Since ψ is piecewise C^2, it is 1-order semismooth; see [15, Proposition 2.26]. Thanks to the Lipschitz continuity of G (see (2.6)), we deduce straightforwardly from [15, Theorem 3.49] that Ψ_G is $\partial\Psi_G$-semismooth. Hence, defining $\partial\Phi(u)=\{M=I-N: N\in\partial\Psi_G(u)\}$, we readily obtain that Φ is $\partial\Phi$-semismooth. ∎

To perform the step in line 3 of Algorithm 1, we have to choose some element in $\partial\Phi(u)$. In order to do this selection and obtain a family of uniformly invertible operators, we define

$$g(t)=\begin{cases}0 & \text{if } t\in(-\infty,-\gamma-\kappa\beta]\cup[-\gamma,\gamma]\cup[\gamma-\kappa\alpha,+\infty),\\ -\dfrac{1}{\kappa} & \text{if } t\in(-\gamma-\kappa\beta,-\gamma)\cup(\gamma,\gamma-\kappa\alpha),\end{cases}\qquad \text{if } \gamma>0,$$

$$g(t)=\begin{cases}0 & \text{if } t\in(-\infty,-\kappa\beta]\cup[-\kappa\alpha,+\infty),\\ -\dfrac{1}{\kappa} & \text{if } t\in(-\kappa\beta,-\kappa\alpha),\end{cases}\qquad \text{if } \gamma=0.$$

Notice that $g(t)=0$ when $t\in\{-\gamma-\kappa\beta,-\gamma,\gamma,\gamma-\kappa\alpha\}$ if $\gamma>0$ or when $t\in\{-\kappa\beta,-\kappa\alpha\}$ if $\gamma=0$ and hence $g(t)\in\partial^{\mathrm{CL}}\psi(t)$. For a given control $u\in L^2(\Omega)$, we select $M_u\in\partial\Phi(u)$ defined as $M_u v=v-h_u\cdot G'(u)v$, where $h_u(x)=g(\varphi_u(x))$.

THEOREM 3.4. *Let* $(\bar u,\bar\varphi,\bar\lambda)\in U_{ad}\times Y\times\partial j(\bar u)$ *satisfy* (2.13) *with* $\bar\varphi=G(\bar u)$ *and assume that the strict complementary condition* $|\Sigma_{\bar u}|=0$ *and the second order sufficient condition* (2.15) *hold at* $\bar u$. *Then, there exist* $\delta>0$ *and* $C>0$ *such that for all* $u\in B_\delta(\bar u)$ *and all* $w\in L^2(\Omega)$, *the equation* $M_u v=w$ *has a unique solution* $v\in L^2(\Omega)$ *and the inequality* $\|v\|_{L^2(\Omega)}\le C\|w\|_{L^2(\Omega)}$ *holds.*

Proof. We define the active and inactive sets for u. For $\gamma > 0$ we define

$$\begin{aligned}
\mathbb{A}^\beta &= \{x \in \Omega : \ \varphi_u(x) \le -\gamma - \kappa\beta\},\\
\mathbb{J}^+ &= \{x \in \Omega : \ -\gamma - \kappa\beta < \varphi_u(x) < -\gamma\},\\
\mathbb{A}^0 &= \{x \in \Omega : \ |\varphi_u(x)| \le \gamma\},\\
\mathbb{J}^- &= \{x \in \Omega : \ \gamma < \varphi_u(x) < \gamma - \kappa\alpha\},\\
\mathbb{A}^\alpha &= \{x \in \Omega : \ \gamma - \kappa\alpha \le \varphi_u(x)\}.
\end{aligned}$$

Notice that all five sets are disjoint and their union is Ω. We set $\mathbb{A} = \mathbb{A}^\alpha \cup \mathbb{A}^\beta \cup \mathbb{A}^0$ and $\mathbb{J} = \mathbb{J}^- \cup \mathbb{J}^+$. In the case $\gamma = 0$ we define $\mathbb{A} = \mathbb{A}^\alpha \cup \mathbb{A}^\beta$ and

$$\mathbb{J} = \{x \in \Omega : \ -\kappa\beta < \varphi_u(x) < -\kappa\alpha\}.$$

Using the notation $\eta_v = G'(u)v$, we have that

$$M_u v = \begin{cases} v & \text{in } \mathbb{A},\\ v + \dfrac{1}{\kappa}\eta_v & \text{in } \mathbb{J}, \end{cases}$$

and the equation $M_u v = w$ is equivalent to the system

$$\left\{\begin{array}{rcll} v &=& w & \text{in } \mathbb{A},\\ v + \dfrac{1}{\kappa}\eta_v &=& w & \text{in } \mathbb{J}. \end{array}\right. \tag{3.3}$$

We write $v = \chi_{\mathbb{J}} v + \chi_{\mathbb{A}} v$. The first equation determines v in the active set $\mathbb{A}$ and we write the second equation as

$$\chi_{\mathbb{J}} v + \frac{1}{\kappa}\eta_{\chi_{\mathbb{J}} v} = w - \frac{1}{\kappa}\eta_{\chi_{\mathbb{A}} w} \text{ in } \mathbb{J}. \tag{3.4}$$

From (2.4) we get that this equation is the optimality condition of the unconstrained quadratic optimization problem

$$\min_{v \in L^2(\mathbb{J})} H(v) := \frac{1}{2}F''(u)(\chi_{\mathbb{J}} v)^2 - \int_{\mathbb{J}} (\kappa w - \eta_{\chi_{\mathbb{A}} w}) v \, \mathrm{d}x. \tag{3.5}$$

Therefore, if we prove that H has a unique local minimizer in $L^2(\mathbb{J})$, the existence and uniqueness of a solution of (3.4) follows. Using the continuity of the functional $u \to F''(u)$, we deduce the existence of $\delta_0 > 0$ such that if $\|u - \bar u\|_{L^2(\Omega)} < \delta_0$, then $|(F''(u) - F''(\bar u))v^2| < \nu/2 \|v\|^2_{L^2(\Omega)}$. From Theorem 2.13, we deduce the existence of $\tau > 0$, with $\tau < \gamma$ if $\gamma > 0$, such that

$$F''(u)v^2 \ge \frac{\nu}{2}\|v\|^2_{L^2(\Omega)} \ \forall v \in T^\tau_{\bar u} \text{ if } \|u - \bar u\|_{L^2(\Omega)} < \delta_0. \tag{3.6}$$

Therefore, (3.5) has a unique local minimizer, that is also global, if $L^2(\mathbb{J}) \subset T^\tau_{\bar u}$. This embedding follows from the inclusion

$$\mathbb{J} \subset \{x \in \Omega : \ |\bar\varphi(x) + \kappa \bar u(x) + \gamma\bar\lambda(x)| \le \tau \text{ and } |\bar\varphi(x)| \ge \gamma - \tau\}, \tag{3.7}$$

or equivalently

$$\{x \in \Omega : \ |\bar\varphi(x) + \kappa \bar u(x) + \gamma\bar\lambda(x)| > \tau \text{ or } |\bar\varphi(x)| < \gamma - \tau\} \subset \mathbb{A}.$$

Let us check this inclusion. Taking $\delta = \min\{\delta_0, 1, \frac{\tau}{K_{G'}(\bar R)}\}$ with $\bar R = \|\bar u\|_{L^2(\Omega)} + 1$, we deduce from (2.6) that $\|\varphi_u - \bar\varphi\|_{C(\bar\Omega)} < \tau$ if $\|u - \bar u\|_{L^2(\Omega)} < \delta$.

Case 1. Suppose $\bar\varphi(x) + \kappa\bar u(x) + \gamma\bar\lambda(x) > \tau$. From (2.22), we have that $\bar u(x) = \alpha$. If $\gamma > 0$, we also deduce from (2.17) and (2.18) that $\bar\lambda(x) = -1$. We can write that $\bar\varphi(x) > \tau - \kappa\bar u(x) - \gamma\bar\lambda(x) = \tau - \kappa\alpha + \gamma$. Since $\varphi_u(x) - \bar\varphi(x) > -\tau$, we have that

$$\varphi_u(x) = \varphi_u(x) - \bar\varphi(x) + \bar\varphi(x) > -\tau + \tau - \kappa\alpha + \gamma = -\kappa\alpha + \gamma$$

and, hence, $x \in \mathbb{A}^\alpha \subset \mathbb{A}$.

Case 2. Suppose $\bar\varphi(x) + \kappa\bar u(x) + \gamma\bar\lambda(x) < -\tau$. From (2.23), we have that $\bar u(x) = \beta$. If $\gamma > 0$, we also deduce from (2.17) and (2.18) that $\bar\lambda(x) = 1$. We can write that $\bar\varphi(x) < -\tau - \kappa\bar u(x) - \gamma\lambda(x) = -\tau - \kappa\beta - \gamma$. Since $\varphi_u(x) - \bar\varphi(x) < \tau$, we have that

$$\varphi_u(x) = \varphi_u(x) - \bar\varphi(x) + \bar\varphi(x) < \tau - \tau - \kappa\beta - \gamma = -\kappa\beta - \gamma$$

and, consequently, $x \in \mathbb{A}^\beta \subset \mathbb{A}$.

For $\gamma = 0$, Cases 1 and 2 imply (3.7).

Case 3. Suppose $\gamma > 0$ and $|\bar\varphi(x)| < \gamma - \tau$. Then $|\varphi_u(x)| \le |\varphi_u(x) - \bar\varphi(x)| + |\bar\varphi(x)| < \tau + \gamma - \tau = \gamma$, which yields $x \in \mathbb{A}^0 \subset \mathbb{A}$.

Therefore (3.6) and (3.7) hold and consequently the system (3.3) has a unique solution v and $\chi_{\mathbb{J}} v \in T^\tau_{\bar u}$. It remains to get an estimate for v in terms of w with a constant independent of $u \in B_\delta(\bar u)$. Using (3.6), (2.4), and (3.4) we get

$$\begin{aligned}\frac{\nu}{2}\|\chi_{\mathbb{J}} v\|^2_{L^2(\Omega)} \le F''(u)(\chi_{\mathbb{J}} v)^2 &= \int_\Omega (\eta_{\chi_{\mathbb{J}} v} + \kappa\chi_{\mathbb{J}} v)\chi_{\mathbb{J}} v \,\mathrm{d}x\\ &= \kappa\int_\Omega \left(w - \frac{1}{\kappa}\eta_{\chi_{\mathbb{A}} w}\right)\chi_{\mathbb{J}} v\,\mathrm{d}x.\end{aligned}$$

From the first equation in (3.3), we have that

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

Multiplying this equality by κ and adding it to the previous inequality, we obtain with the Cauchy–Schwarz inequality and estimate (2.5)

$$\begin{aligned}\min\left\{\kappa, \frac{\nu}{2}\right\}\|v\|^2_{L^2(\Omega)} &\le \kappa\int_\Omega wv\,\mathrm{d}x - \int_\Omega \eta_{\chi_{\mathbb{A}} w}\chi_{\mathbb{J}} v\,\mathrm{d}x\\ &\le \left(\kappa\|w\|_{L^2(\Omega)} + \|\eta_{\chi_{\mathbb{A}} w}\|_{L^2(\Omega)}\right)\|v\|_{L^2(\Omega)}\\ &\le \left(\kappa + K_{G'}(\bar R)\right)\|w\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}.\end{aligned}$$

This yields $\|v\|_{L^2(\Omega)} \le C\|w\|_{L^2(\Omega)}$ with $C = \frac{\kappa + K_{G'}(\bar R)}{\min\{\kappa, \frac{\nu}{2}\}}$. ∎

The following result is an immediate consequence of Theorem 3.2, Lemma 3.3, and Theorem 3.4.

COROLLARY 3.5. *Let* $(\bar u, \bar\varphi, \bar\lambda) \in U_{ad} \times Y \times \partial j(\bar u)$ *satisfy* (2.13) *with* $\bar\varphi = G(\bar u)$ *and assume that the strict complementary condition* $|\Sigma_{\bar u}| = 0$ *and the second order sufficient condition* (2.15) *hold at* $\bar u$. *Then, there exists* $\delta > 0$ *such that for all* $u_0 \in B_\delta(\bar u)$, *the sequence spanned by Algorithm* 2 *converges superlinearly to* $\bar u$.

The semismooth Newton's method for problem (P) is detailed in Algorithm 2.

Remark 3.6. Since $C_{\bar u} \subset T^\tau_{\bar u}$, then (3.6) implies the sufficient second order optimality condition (2.15). The proof of Theorem 3.4 uses (3.6), but the strict complementarity condition $|\Sigma_{\bar u}| = 0$ is not used. Consequently, the statement of Theorem 3.4,

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

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

repeat

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

$$Ay_j + f(x, y_j) = u_j \text{ in } \Omega,\ y_j = 0 \text{ in } \Gamma$$

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

$$A^*\varphi_j + \frac{\partial f}{\partial y}(x, y_j)\varphi_j = \frac{\partial L}{\partial y}(x, y_j) \text{ in } \Omega,\ \varphi_j = 0 \text{ in } \Gamma$$

Compute $\mathbb{A}_j^\beta$, $\mathbb{A}_j^0$, $\mathbb{A}_j^\alpha$, $\mathbb{A}_j$, and $\mathbb{J}_j^+$, $\mathbb{J}_j^-$, $\mathbb{J}_j$ using φ_j.

Compute

$$w_j(x) = -\Phi(u_j)(x) = \begin{cases} -u_j(x) + \beta & \text{if } x \in \mathbb{A}_j^\beta \\ -u_j(x) - \frac{1}{\kappa}(\varphi_k(x) + \gamma) & \text{if } x \in \mathbb{J}_j^+ \\ -u_j(x) & \text{if } x \in \mathbb{A}_j^0 \\ -u_j(x) - \frac{1}{\kappa}(\varphi_k(x) - \gamma) & \text{if } x \in \mathbb{J}_j^- \\ -u_j(x) + \alpha & \text{if } x \in \mathbb{A}_j^\alpha \end{cases}$$

Compute $\eta_j = \eta_{\chi_{\mathbb{A}_j} w_j}$ solving the linear equations

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

Solve the quadratic problem

$$(Q_j) \qquad \min_{v \in L^2(\mathbb{J}_j)} H_j(v) := \frac{1}{2} F''(u_j)(\chi_{\mathbb{J}_j} v)^2 - \int_{\mathbb{J}_j} (\kappa w_j - \eta_j) v \,\mathrm{d}x$$

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

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

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

until *convergence*

and also that of Corollary 3.5, can be rewritten replacing (2.15) and the strict complementarity condition by (3.6). Obviously, this is a weaker assumption, but it is a less natural assumption: if the strict complementarity condition is not satisfied, the gap between (3.6) and the second order necessary condition is too large.

Algorithm 3: Computation of the product Hessian vector.

Solve $Az + \frac{\partial f}{\partial y}(x, y_j)z = \chi_{\mathbb{J}_j} v$ in Ω, $z = 0$ in Γ

Solve $A^*\eta + \frac{\partial f}{\partial y}(x, y_j)\eta = \Big(\frac{\partial^2 L}{\partial y^2}(x, y_j) - \varphi_j \frac{\partial^2 f}{\partial y^2}(x, y_j)\Big) z$ in Ω, $\eta = 0$ in Γ

Set $A_j v = \chi_{\mathbb{J}_j}(\eta + \kappa v)$

4. Some computational details and numerical examples. Let us comment on how to solve the quadratic problem (Q_j) that appears in line 8 of Algorithm 2. Notice that we can write $H_j(v) = \frac{1}{2}(v, A_j v)_{L^2(\mathbb{J}_j)} - (b_j, v)_{L^2(\mathbb{J}_j)}$, where $b_j = \chi_{\mathbb{J}_j}(\kappa w_j - \eta_j)$ and we can compute $A_j v$ using Algorithm 3. Therefore (Q_j) can be solved using, e.g., the conjugate gradient method without need of the explicit computation of the Hessian $F''(u_j)$.

From the computational point of view, at each step of Algorithm 2 we have to solve one nonlinear partial differential equation and several linear partial differential equations: three before solving the quadratic problem and two at each step of the conjugate gradient method that we use to solve the quadratic problem. When discretized, all these linear equations share either the same coefficient matrix (or its transpose in the case of a nonsymmetric problem; see, e.g., [5]). Therefore, an advantage can be taken from a single factorization. If the nonlinear equation at iteration $j+1$ is solved using Newton's method, the matrix of the linear problem to be solved in the first iteration in this subproblem is the same as the matrix used for the linear equations at iteration j.

We present one example posed in a two-dimensional domain and another one in a three-dimensional domain. To solve the problem we use the finite element approximation studied in [2]: the state, the adjoint state, and the control are discretized using continuous piecewise linear elements and the Tikhonov and sparsity terms are discretized using the composite trapezoid formula. We stop the algorithm 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 with the MATLAB built-in command `pcg` and the nonlinear equation in line 3 is solved using Newton's method. The tolerance 5×10^{-14} is used for both subproblems.

Since we do not have the exact solutions of the problems presented, we cannot check the assumptions of Corollary 3.5 beforehand. Nevertheless, in both problems we find numerically a solution of the optimality system (2.13), the fast convergence of the conjugate gradient method is a good indication that the second order sufficient condition (2.15) is satisfied, and we check numerically the strict complementarity condition $|\Sigma_{\bar u}| = 0$.

Example 1. We consider the data of Example 1 in [2], where convergence of the finite element approximation of (P) is studied and error estimates in terms of the discretization parameter are obtained: $\Omega = B_1(0,0) \subset \mathbb{R}^2$, $Ay = -\Delta y$, $f(x,y) = y^3$, $L(x,y) = \frac{1}{2}(y - y_d(x))^2$ with $y_d(x) = 3\sin(2\pi x_1)\sin(\pi x_2)\mathrm{e}^{x_1}$, $\kappa = 0.002$, $\gamma = 0.03$, $\alpha = -12$, and $\beta = 12$. It is straightforward to show that assumptions (A1)–(A3) are satisfied.

As in [2], we solve the problem in a mesh of size $h = 2^{-7}$ (1.3×10^5 elements, 66049 nodes). In order to get an initial point u_0 close enough to $\bar u$, we take as u_0 the solution of the discretized problem with mesh size $h = 2^{-6}$.

TABLE 1

Convergence history of the problem in Example 1. *♯Newton is the number of Newton iterations to solve the nonlinear PDE in line* 3 *and ♯CG is the number of iterations of the conjugate gradient method used to solve* (Q_j) *in Algorithm* 2.

j	$J(u_j)$	δ_j	♯Newton	♯CG
0	11.141742584195615	1.5×10^{-2}	4	13
1	11.141687025807151	4.5×10^{-5}	3	12
2	11.141686904484867	1.0×10^{-7}	3	13
3	11.141686904484866	2.4×10^{-14}	2	14
4	11.141686904484862		1	

TABLE 2

Convergence history of the problem in Example 2. *♯Newton is the number of Newton iterations to solve the nonlinear PDE in line* 3 *and ♯CG is the number of iterations of the conjugate gradient method used to solve* (Q_j) *in Algorithm* 2.

j	$J(u_j)$	δ_j	♯Newton	♯CG
0	5.1160436513941248	2.6×10^{0}	4	4
1	4.8088004565179974	1.0×10^{-2}	4	4
2	4.8087950298698070	2.7×10^{-7}	3	4
3	4.8087950298698035	6.6×10^{-16}	2	5
4	4.8087950298698035		1	

We have summarized the convergence history in Table 1. The superlinear order of convergence can be appreciated in the way the order of magnitude of the error between iterations δ_j varies in the first steps: -2, -5, -7, -14. We find numerically that $|\mathbb{J}|=0.678$, $|\mathbb{A}^\beta|=0.310$, $|\mathbb{A}^\alpha|=0.310$, $|\mathbb{A}^0|=1.844$, and $|\Sigma_{\bar u}|=0$.

Example 2. Consider $\Omega=(0,1)^3\subset\mathbb{R}^3$, $Ay=-\Delta y$, $f(x,y)=|y|^3y$, $L(x,y)=\frac{1}{2}(y-y_d(x))^2$ with $y_d=\prod_{i=1}^3 8x_i(1-x_i)$, $\kappa=0.1$, $\gamma=0.05$, $\alpha=-1$, $\beta=1$. Assumptions (A1)–(A3) are clearly satisfied in this setting.

We use a mesh of size $h=2^{-5}$ (1.97×10^5 elements, 35937 nodes) and start with $u_0=y_d$.

We have summarized the convergence history in Table 2. The superlinear order of convergence can be appreciated in the way the order of magnitude of the error between iterations δ_j varies in the first steps: 0, -2, -7, -16. We find numerically that $|\mathbb{J}|=0.323$, $|\mathbb{A}^\beta|=0.157$, $|\mathbb{A}^\alpha|=0$, $|\mathbb{A}^0|=0.520$, and $|\Sigma_{\bar u}|=0$.

REFERENCES

[1] S. Amstutz and A. Laurain, *A semismooth Newton method for a class of semilinear optimal control problems with box and volume constraints*, Comput. Optim. Appl., 56 (2013), pp. 369–403, https://doi.org/10.1007/s10589-013-9555-6.

[2] E. Casas, R. Herzog, and G. Wachsmuth, *Approximation of sparse controls in semilinear equations by piecewise linear functions*, Numer. Math., 122 (2012), pp. 645–669, https://doi.org/10.1007/s00211-012-0475-7.

[3] E. Casas, R. Herzog, and G. Wachsmuth, *Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional*, SIAM J. Optim., 22 (2012), pp. 795–820, https://doi.org/10.1137/110834366.

[4] E. Casas, K. Kunisch, and F. Tröltzsch, *Optimal control of PDEs and FE-approximation*, in Numerical Control Part A, Handb. Numer. Anal. 23, E. Trélat and E. Zuazua, eds., North-Holland, Amsterdam, 2022, pp. 115–163.

[5] E. Casas, M. Mateos, and A. Rösch, *Analysis of control problems of nonmontone semilinear elliptic equations*, ESAIM Control Optim. Calc. Var., 26 (2020), 80, https://doi.org/10.1051/cocv/2020032.

[6] F. Clarke, *Functional Analysis, Calculus of Variations and Optimal Control*, Grad. Texts in Math. 264, Springer, London, 2013, https://doi.org/10.1007/978-1-4471-4820-3.

[7] M. Hintermüller, K. Ito, and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2002), pp. 865–888, https://doi.org/10.1137/S1052623401383558, 2003.

[8] F. Mannel and A. Rund, *A hybrid semismooth quasi-Newton method for nonsmooth optimal control with PDEs*, Optim. Eng., 22 (2021), pp. 2087–2125, https://doi.org/10.1007/s11081-020-09523-w.

[9] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer Ser. Oper. Res., Springer-Verlag, New York, 1999, https://doi.org/10.1007/b98874.

[10] K. Pieper, *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.

[11] S. M. Robinson, *A quadratically-convergent algorithm for general nonlinear programming problems*, Math. Program., 3 (1972), pp. 145–156, https://doi.org/10.1007/BF01584986.

[12] S. M. Robinson, *Perturbed Kuhn-Tucker points and rates of convergence for a class of nonlinear-programming algorithms*, Math. Program., 7 (1974), pp. 1–16, https://doi.org/10.1007/BF01585500.

[13] G. Stadler, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), pp. 159–181, https://doi.org/10.1007/s10589-007-9150-9.

[14] F. Tröltzsch, *On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations*, SIAM J. Control Optim., 38 (1999), pp. 294–312, https://doi.org/10.1137/S0363012998341423.

[15] M. Ulbrich, *Semismooth Newton Methods for Variational Inequalities and Constrained Optimization Problems in Function Spaces*, MOS-SIAM Ser. Optim. 11, SIAM, Philadelphia, 2011, https://doi.org/10.1137/1.9781611970692.