Preprint, IFB-Report No. 11 (04/2008), Institute of Mathematics and Scientific Computing, University of Graz.

MATHEMATICAL PROGRAMS WITH COMPLEMENTARITY CONSTRAINTS IN FUNCTION SPACE: C- AND STRONG STATIONARITY AND A PATH-FOLLOWING ALGORITHM

M. HINTERMÜLLER AND I. KOPACKA

Abstract. An optimal control problem governed by an elliptic variational inequality is studied. The feasible set of the problem is relaxed and a path-following type method is used to regularize the constraint on the state variable. First order optimality conditions for the relaxed-regularized subproblems are derived and convergence of stationary points with respect to the relaxation and regularization parameters is shown. In particular, C- and strong stationarity as well as variants thereof are studied. The subproblems are solved by using semismooth Newton methods. The overall algorithmic concept is provided and its performance is discussed by means of examples, including problems with bilateral constraints and a nonsymmetric operator.

1. Introduction

Mathematical programs with equilibrium constraints (MPECs) received a considerable amount of attention in finite dimensional space in the recent past; see, e.g., the monographs [32, 37] and the many references therein. Concerning problems posed in function space, however, the topic is significantly less researched. In the latter context, MPECs typically arise in connection with optimal control problems for variational inequalities. An account of this problem class together with a state-of-the art overview of the work in the 80's can be found in [3]. Since then there has been a number of research efforts; see, e.g., the work in [4, 5, 6, 7, 9, 17, 23, 27, 29, 31, 34]. We also refer to the recent two-volume monograph [35, 36]. But still, the overall research level is far less complete when compared to finite dimensions and, as far as stationarity principles are concerned, significantly less complete and systematic. Typically only a weak form of stationarity is derived. Further, the literature on numerical solution procedures for function space based problems is extremely scarce.

One of our goals in the present paper is to contribute to systematizing and completing the notions of stationarity in the function space context. In fact, in finite dimensions it is well-known that, depending on what MPEC-based constraint qualification or possibly second order condition is satisfied, respectively, different forms of stationarity arise; see, e.g., [41]. In function space it turns out that, for instance concepts related to C- and strong stationarity are available as well. Moreover, depending on certain conditions, we also introduce the new concepts of (ε-)*almost*

Date: January 13, 2009.

2000 *Mathematics Subject Classification.* 49K20, 49M37, 65K05, 90C33.

Key words and phrases. Constrained optimal control, mathematical programs with complementarity constraints, MPEC, MPCC, Moreau-Yosida regularization, semismooth Newton method, C- and strong stationarity.

The authors gratefully acknowledge support by the Austrian Federal Ministry of Science and Research (bm.w_ f) and the FWF under START-program Y305 "Interfaces and Free Boundaries".

C- and (ε-)*almost strong stationarity*, which are specific to the function space context considered here. We also mention that in [28] a non-pointwise counterpart of M-stationarity was derived for a problem in one spatial dimensional. Secondly we aim at a constructive proof technique which can be cast into a solution algorithm. This results in a numerical method which admits a function space based convergence analysis. As a consequence one expects that the discrete counterpart of the method exhibits some numerical stability under refinements of the discretization of the infinite dimensional problem.

In order to address some of the analytical as well as numerical difficulties attached to MPECs in function space we consider the following optimal control problem in which the state y of the system is defined as the solution of an elliptic variational inequality (VI). We study

$$\begin{aligned} &\min J(y,u) \text{ over } y \in K, u \in \mathcal{U}, \\ &\text{subject to (s.t.)} \quad \langle \mathcal{A}y - g(u), v - y\rangle \geq 0 \quad \forall v \in K, \end{aligned}$$

where $\mathcal{A}$ denotes a second order linear elliptic partial differential operator and K is a closed convex cone in some suitable Banach space $\mathcal{Y}$. The duality pairing between $\mathcal{Y}$ and its dual $\mathcal{Y}'$ is given by $\langle\cdot,\cdot\rangle$. Further, u denotes the control variable and the source term $g(\cdot)$ determines the control action. Assuming, e.g., $K := \{v \in \mathcal{Y} : v \geq 0 \text{ a.e. in } \Omega\}$ and introducing a slack variable ξ, the VI can be equivalently written as the linear complementarity problem

$$\mathcal{A}y - \xi = g(u), \quad y \geq 0 \text{ a.e. in } \Omega, \quad \langle \xi, v - y\rangle \geq 0 \quad \forall v \in K. \tag{1.1}$$

The optimization problem under investigation then becomes a mathematical program with complementarity constraints (MPCC) in function space which is a particular instance of an MPEC. Then, regardless of the dimension of the underlying space, the last two relations in (1.1) are equivalent to

$$y \geq 0, \quad \xi \geq 0, \quad \langle y, \xi\rangle = 0. \tag{1.2}$$

As this system is part of the constraint set of the overall minimization problem, typically all classical constraint qualifications (such as, for instance, the linear independence CQ or the Mangasarian-Fromovitz CQ in finite dimensions) are violated. Hence, deriving optimality conditions from standard mathematical programming theory in Banach space is impossible. Furthermore, in our general function space context, the state constraint $y \geq 0$ is critical as it gives rise to a Lagrange multiplier with low regularity [12]. This fact requires a careful numerical treatment in order to obtain stability of the solution algorithm under mesh refinements; see, e.g., [26].

In the course of this paper we will address a particular instance of the MPCC defined above, where no constraints act on the control u and the function g is linear. In order to overcome the first difficulty mentioned above we relax the constraint $\langle y, \xi\rangle = 0$ and, hence, enlarge the feasible set. This technique has been used in [42] for a finite dimensional problem and in [5] in function space. In [42] both pointwise and integral-type relaxation are used and similar convergence results for the two approaches are proven. We use a relaxation of the form $\langle y, \xi\rangle \leq \alpha$, $\alpha > 0$, as the resulting problem is of lower dimension. In fact, in this case the Lagrange multiplier is only a scalar, as opposed to a function in the case of pointwise relaxation. We observe that the resulting relaxed problem still contains a pointwise state constraint. In order to overcome the low multiplier regularity associated with such constraints, we use a Moreau-Yosida-based regularization [25, 26] and solve the resulting subproblems by a semismooth Newton method [24].

The rest of the paper is organized as follows. In section 2 we define the problem and discuss the relaxation and the regularization approach. In section 3 we investigate the convergence behavior of global solutions with respect to the relaxation and regularization parameters. In section 4 we derive first order optimality systems and investigate the convergence behavior of stationary points as this reflects the typical situation in numerical practice. In section 5 we define and analyze the semismooth Newton method and discuss the overall solution algorithm. In section 6 we illustrate the results via numerical examples.

2. Problem formulation

Let Ω be an open, bounded subset of $\mathbb{R}^n$, $n \leq 3$, with a smooth boundary $\partial\Omega$. Throughout this paper we denote by $(\cdot,\cdot)$ and $\|\cdot\|$ the scalar product and norm in $L^2(\Omega)$ and by $\langle\cdot,\cdot\rangle$ the duality pairing between $H^{-1}(\Omega)$ and $H^1_0(\Omega)$. We consider the bilinear form $a(\cdot,\cdot)$ defined on $H^1_0(\Omega)\times H^1_0(\Omega)$ by

$$a(v,w) = \sum_{i,j=1}^{n} \int_\Omega a_{ij}\frac{\partial v}{\partial x_j}\frac{\partial w}{\partial x_i}dx + \sum_{i=1}^{n}\int_\Omega b_i \frac{\partial v}{\partial x_i} w\,dx + \int_\Omega cvw\,dx \quad \forall v,w \in H^1_0(\Omega),$$

where a_{ij}, b_i and c belong to $L^\infty(\Omega)$. Moreover we suppose that $a_{ij} \in C^{0,1}(\bar\Omega)$ and $c \geq 0$. We further assume that $a(\cdot,\cdot)$ is bounded, i.e.,

$$\text{(H1)} \qquad \exists\, C_b > 0 \text{ such that } |a(v,w)| \leq C_b \|v\|_{H^1_0}\|w\|_{H^1_0} \quad \forall v,w \in H^1_0(\Omega),$$

and coercive, i.e.,

$$\text{(H2)} \qquad \exists\, C_c > 0 \text{ such that } a(v,v) \geq C_c\|v\|^2_{H^1_0} \quad \forall v \in H^1_0(\Omega).$$

Due to (H1) and (H2) the bilinear form $a(\cdot,\cdot)$ defines a norm. We further define the associated operator $\mathcal{A} : H^1_0(\Omega) \to H^{-1}(\Omega)$ by

$$\langle \mathcal{A}v, w\rangle = a(v,w) \quad \forall v,w \in H^1_0(\Omega)$$

and the cone K by

$$K = \{v \in H^1_0(\Omega) : v \geq 0 \text{ a.e. in } \Omega\}.$$

The state variable $y \in H^1_0(\Omega)$ is determined by the control $u \in L^2(\Omega)$ via the solution of the variational inequality

$$\text{(2.1)} \qquad y \in K,\ a(y, v-y) \geq (f+u, v-y) \quad \forall v \in K,$$

where $f \in L^2(\Omega)$ is a fixed data term. This variational problem has a unique solution; see, for instance, [16, 30]. Introducing the slack variable ξ, we can reformulate the variational inequality equivalently as the complementarity system

$$\mathcal{A}y = \xi + u + f,\ y \geq 0 \text{ a.e. in } \Omega,\ \xi \geq 0,\ \langle y, \xi\rangle = 0.$$

A priori, ξ is an element of the dual space $H^{-1}(\Omega)$ and $\xi \geq 0$ has to be interpreted as $\langle \xi, v\rangle \geq 0$ for all $v \in K$. If the domain Ω is sufficiently smooth and if $u+f \in L^2(\Omega)$, then according to [16, 30] the solution y of the variational inequality is an element of $H^2(\Omega) \cap H_0^1(\Omega) =: \mathcal{X}$. Therefore $\xi \in L^2(\Omega)$ and $\xi \geq 0$ a.e. in Ω. We now define the optimal control problem $(\mathcal{P})$

$$\min J(y,u) := \frac{1}{2}\|y - y_d\|^2 + \frac{\nu}{2}\|u\|^2 \text{ over } y \in H_0^1(\Omega), u, \xi \in L^2(\Omega),$$

$$\text{subject to (s.t.) } \mathcal{A}y = u + \xi + f \quad \text{in } H^{-1}(\Omega), \tag{2.2a}$$

$$y \geq 0 \text{ a.e. in } \Omega, \tag{2.2b}$$

$$\xi \geq 0 \text{ a.e. in } \Omega, \tag{2.2c}$$

$$(y, \xi) = 0. \tag{2.2d}$$

Here $y_d \in L^2(\Omega)$ denotes the desired state and $\nu > 0$ is the cost of the control. We call (2.2a) the state equation and (2.2b)-(2.2d) the complementarity system with respect to y and ξ. Problem (2.2) defines a mathematical program with complementarity constraints (MPCC) in function space. Existence of a solution of $(\mathcal{P})$ was proven in [34].

Remark 2.1. Note that (2.1) defines the obstacle problem for the trivial obstacle $\psi = 0$. The MPCC (2.2) can easily be modified to suit a sufficiently smooth obstacle ψ with $\psi|_{\partial\Omega} \leq 0$ by considering the transformation $\tilde{y} = y - \psi$ and modifying f and y_d accordingly.

In order to set up an algorithmic approach which admits an analysis in function space we consider a series of relaxed and regularized problems approximating the original MPCC. To ensure the existence of Lagrange multipliers we inflate the feasible domain by replacing the constraint (2.2d) with the inequality $(y, \xi) \leq \alpha$, where $\alpha > 0$ is called the *relaxation parameter* (see, e.g., [5, 42]). Bergounioux shows by means of a counterexample, that by simply relaxing equation (2.2d), the boundedness of ξ in $L^2(\Omega)$ is no longer given. Therefore, to guarantee the existence of a solution, an additional constraint of the type $\|\xi\| \leq R$ with some sufficiently large fixed constant R is introduced. Using a pointwise relaxation, instead of the integral-type approach, would, in Bergounioux's example, solve the problem of existence of a solution, without having to "force" the quantity ξ to be bounded. Nevertheless the problematic nature remains and other examples can be found for which the relaxed problem has no optimal solution. Instead of invoking the explicit constraint $\|\xi\| \leq R$, we rather add a term containing the L^2-norm of ξ to the cost functional, where the corresponding weight parameter $\kappa > 0$ tends to zero as α tends to zero. This term not only ensures the existence of a solution for positive κ, but also, as we will show in section 5, the term is beneficial to the function space convergence analysis of our solution method. The resulting problem $(\mathcal{P}_\alpha)$ has the following structure:

$$\min J(y,u) + \frac{\kappa}{2}\|\xi\|^2 \text{ over } y \in \mathcal{X}, u, \xi \in L^2(\Omega),$$

$$\text{s.t. } \mathcal{A}y = u + \xi + f \quad \text{in } L^2(\Omega), \tag{2.3a}$$

$$y \geq 0 \text{ a.e. in } \Omega, \tag{2.3b}$$

$$\xi \geq 0 \text{ a.e. in } \Omega, \tag{2.3c}$$

$$(y, \xi) \leq \alpha. \tag{2.3d}$$

Note that in (2.3) the space for the state variable $y \in \mathcal{X}$ was chosen. Due to the nature of the constraint (2.3b) standard constraint qualifications (e.g., [44]) do not hold for $y \in H_0^1(\Omega)$. This mirrors the fact that the relaxed problem $(\mathcal{P}_\alpha)$ belongs to the problem class of state-constrained optimal control problems, a class which typically features low multiplier regularity; see, e.g., [12]. Since $y \in H^2(\Omega)$, which continuously embeds into $C(\bar{\Omega})$ for $n \leq 3$, the multiplier corresponding to the pointwise constraint $y \geq 0$ is an element of the space of regular Borel-measures. To regularize the problem, we use a Moreau-Yosida based regularization (e.g. [25]) with a *regularization parameter* $\gamma > 0$. This allows us to achieve higher regularity for the multiplier associated to the state constraint (2.2b). In fact, we will show that it is an element of the dual space $H^{-1}(\Omega)$. The relaxed-regularized problem $(\mathcal{P}_{\alpha,\gamma})$ is defined by

$$\min \tilde{J}_\gamma(y,u,\xi) := J(y,u) + \frac{\kappa}{2}\|\xi\|^2 + \frac{1}{2\gamma}\|\max(0,\bar{\lambda} - \gamma y)\|^2$$

$$\text{over } y \in H_0^1(\Omega), u, \xi \in L^2(\Omega),$$

$$\text{s.t. } \mathcal{A}y = u + \xi + f \quad \text{in } H^{-1}(\Omega), \tag{2.4a}$$

$$\xi \geq 0 \ \text{ a.e. in } \Omega, \tag{2.4b}$$

$$(y,\xi) \leq \alpha, \tag{2.4c}$$

where $(\gamma,\alpha,\kappa) > 0$ and $\bar{\lambda} \in L^p(\Omega)$ with $p > 2$, $\bar{\lambda} \geq 0$ a.e. in Ω. Above the max-operation is understood in the pointwise almost everywhere sense.

3. Global solutions

In this section we prove existence of global solutions of the relaxed-regularized problem and discuss convergence of these solutions with respect to the relaxation and regularization parameters. Below we frequently operate on subsequences, which we will, for the sake of readability, not denote specifically.

Theorem 3.1. *For each triple $(\gamma,\alpha,\kappa) > 0$ the relaxed-regularized problem* (2.4) *admits at least one globally optimal solution.*

Proof. First note that the feasible set

$$\mathcal{D} := \{(y,u,\xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega) : (y,u,\xi) \text{ satisfies (2.4a)-(2.4c)}\}$$

of $(\mathcal{P}_{\alpha,\gamma})$ is nonempty, as the point $(\tilde{y},\tilde{u},\tilde{\xi}) := (0,-f,0) \in \mathcal{D}$. Now let $\{(y_n,u_n,\xi_n)\} \subset \mathcal{D}$ be a minimizing sequence, such that

$$\lim_{n\to\infty} \tilde{J}_\gamma(y_n,u_n,\xi_n) = \inf\{\tilde{J}_\gamma(y,u,\xi) : (y,u,\xi) \in \mathcal{D}\} := d \geq 0.$$

Since $L^2(\Omega)$ is a reflexive, separable Banach space, every bounded sequence in $L^2(\Omega)$ has a weakly convergent subsequence. As $\{\tilde{J}_\gamma(y_n,u_n,\xi_n)\}$ is bounded, $\{u_n\}$ and $\{\xi_n\}$ are bounded in the $L^2(\Omega)$-norm. Therefore there exist $(\bar{u},\bar{\xi}) \in L^2(\Omega) \times L^2(\Omega)$ and a subsequence (also denoted by $\{(y_n,u_n,\xi_n)\}$) such that

$$u_n \rightharpoonup \bar{u} \ \text{ and } \ \xi_n \rightharpoonup \bar{\xi} \quad \text{in } L^2(\Omega).$$

Using equation (2.4a) and (H2) we infer

$$C_c\|y_n\|_{H_0^1}^2 \leq \langle \mathcal{A}y_n, y_n\rangle \leq (\|u_n\|_{H^{-1}} + \|\xi_n\|_{H^{-1}} + \|f\|_{H^{-1}})\,\|y_n\|_{H_0^1}.$$

Since $\{u_n\}$, $\{\xi_n\}$ and f are bounded in $H^{-1}(\Omega)$ due to the continuous embedding of $L^2(\Omega)$ in $H^{-1}(\Omega)$, $\{\|y_n\|_{H_0^1}\}$ is bounded and, by the reflexivity and separability of $H_0^1(\Omega)$, there exists $\bar{y} \in H_0^1(\Omega)$ and a further subsequence such that $y_n \rightharpoonup \bar{y}$ in

$H_0^1(\Omega)$. We next show that the limit point $(\bar{y}, \bar{u}, \bar{\xi})$ is feasible. From the compact embedding of $H_0^1(\Omega)$ in $L^2(\Omega)$ we infer

$$\alpha \geq (y_n, \xi_n) \to (\bar{y}, \bar{\xi}),$$

consequently $(\bar{y}, \bar{u}, \bar{\xi})$ satisfies (2.4c). As each element of the minimizing sequence satisfies the state equation (2.4a), taking the limit $n \to \infty$ yields

$$\mathcal{A}\bar{y} = \bar{u} + \bar{\xi} + f \quad \text{in } H^{-1}(\Omega).$$

Further, $\bar{\xi}$ satisfies (2.4b) due to the fact that the set

$$\mathcal{D}_\xi := \{\xi \in L^2(\Omega) : \xi \geq 0 \text{ a.e.}\}$$

is closed and convex and hence weakly closed (see, e.g., [10], p.38) and $\xi_n \in \mathcal{D}_\xi$ for all $n \in \mathbb{N}$. The weak convergence of $\{(y_n, u_n, \xi_n)\}$, the feasibility of $(\bar{y}, \bar{u}, \bar{\xi})$ and the lower semi-continuity of $\tilde{J}_\gamma$ give

$$d = \liminf_{n\to\infty} \tilde{J}_\gamma(y_n, u_n, \xi_n) \geq \tilde{J}_\gamma(\bar{y}, \bar{u}, \bar{\xi}) \geq d.$$

Therefore $(\bar{y}, \bar{u}, \bar{\xi})$ is an optimal solution of $(\mathcal{P}_{\alpha,\gamma})$. □

Next we are interested in the convergence behavior of optimal solutions with respect to the regularization and relaxation parameters. For each $\gamma > 0$, let α_γ and $\kappa_\gamma > 0$ satisfy $\alpha_\gamma \to 0$, $\kappa_\gamma \to 0$ as $\gamma \to \infty$. We now show that the global solutions of the relaxed-regularized problems (2.4) converge to a global solution of the original problem (2.2).

Theorem 3.2. *For every $\gamma > 0$ let $(y_\gamma, u_\gamma, \xi_\gamma) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ be a solution of $(\mathcal{P}_{\alpha_\gamma,\gamma})$. Then there exists $(\tilde{y}, \tilde{u}, \tilde{\xi}) \in \mathcal{X} \times L^2(\Omega) \times L^2(\Omega)$ such that $y_\gamma \to \tilde{y}$ in $H_0^1(\Omega)$, $u_\gamma \to \tilde{u}$ in $L^2(\Omega)$ and $\xi_\gamma \to \tilde{\xi}$ in $H^{-1}(\Omega)$ as $\gamma \to \infty$, where $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is a solution of $(\mathcal{P})$. Furthermore $\frac{1}{2\gamma}\|\max(0, \bar{\lambda} - \gamma y_\gamma)\|^2 \to 0$ and $\frac{\kappa_\gamma}{2}\|\xi_\gamma\|^2 \to 0$ as $\gamma \to \infty$.*

Proof. Again we argue by using the feasible point $(0, -f, 0)$ for all γ. For each $\gamma > 1$ we estimate

$$\tilde{J}_\gamma(y_\gamma, u_\gamma, \xi_\gamma) \leq \tilde{J}_\gamma(0, -f, 0) \leq \frac{1}{2}\|y_d\|^2 + \frac{\nu}{2}\|f\|^2 + \frac{1}{2}\|\max(0, \bar{\lambda})\|^2.$$

¿From this it follows that the sequences $\{y_\gamma\}$, $\{u_\gamma\}$ and $\left\{\frac{1}{\sqrt{2\gamma}}\max(0, \bar{\lambda} - \gamma y_\gamma)\right\}$ are uniformly bounded in $L^2(\Omega)$ as $\gamma \to \infty$. Furthermore, due (H2), (2.4a) and (2.4c), we obtain

$$C_c\|y_\gamma\|^2_{H_0^1} \leq \langle \mathcal{A}y_\gamma, y_\gamma\rangle = \langle u_\gamma + f + \xi_\gamma, y_\gamma\rangle \leq \|u_\gamma + f\|_{H^{-1}}\|y_\gamma\|_{H_0^1} + \alpha_\gamma.$$

Thus there exist $\tilde{y} \in H_0^1(\Omega)$, $\tilde{u} \in L^2(\Omega)$ and a subsequence (again denoted by $\{(y_\gamma, u_\gamma)\}$), such that y_γ converges to $\tilde{y}$ weakly in $H_0^1(\Omega)$ and u_γ converges to $\tilde{u}$ weakly in $L^2(\Omega)$ and strongly in $H^{-1}(\Omega)$ due to the compact embedding of $L^2(\Omega)$ in $H^{-1}(\Omega)$. As $\left\{\frac{1}{\sqrt{2\gamma}}\max(0, \bar{\lambda} - \gamma y_\gamma)\right\}$ is bounded in $L^2(\Omega)$ we obtain

$$\|\max(0, \frac{\bar{\lambda}}{\gamma} - y_\gamma)\| \overset{\gamma\to\infty}{\longrightarrow} 0.$$

Since y_γ converges strongly in $L^2(\Omega)$, without loss of generality we may assume that y_γ converges to $\tilde{y}$ a.e. in Ω. Taking the limit $\gamma \to \infty$ and applying Fatou's lemma (see, e.g., [2]), we conclude that $\|\max(0,-\tilde{y})\| = 0$ and consequently $\tilde{y} \geq 0$ a.e. in Ω.

The triple $(y_\gamma, u_\gamma, \xi_\gamma)$ satisfies the state equation (2.4a). Due to (H1) and the boundedness of $\{(y_\gamma, u_\gamma)\}$ in $H_0^1(\Omega) \times L^2(\Omega)$ the sequence $\{\xi_\gamma\}$ is bounded in $H^{-1}(\Omega)$. Hence there exists $\tilde{\xi} \in H^{-1}(\Omega)$ such that (on a further subsequence denoted the same) $\xi_\gamma \rightharpoonup \tilde{\xi}$ in $H^{-1}(\Omega)$ and

$$\mathcal{A}\tilde{y} = \tilde{u} + \tilde{\xi} + f \quad \text{in } H^1(\Omega).$$

Note that $\xi_\gamma \geq 0$ a.e. in Ω. We therefore obtain

$$\langle \tilde{\xi}, v\rangle = \lim_{\gamma\to\infty} \langle \xi_\gamma, v\rangle \geq 0 \text{ for all } v \in H_0^1(\Omega),\ v \geq 0 \text{ a.e. in } \Omega. \tag{3.1}$$

Next we estimate

$$\begin{aligned} 0 \leq C_c\|y_\gamma - \tilde{y}\|^2_{H_0^1(\Omega)} &\leq \langle \mathcal{A}(y_\gamma - \tilde{y}), y_\gamma - \tilde{y}\rangle = \langle u_\gamma - \tilde{u} + \xi_\gamma - \tilde{\xi}, y_\gamma - \tilde{y}\rangle \\ &\leq \langle u_\gamma - \tilde{u}, y_\gamma - \tilde{y}\rangle + \alpha_\gamma - \langle \xi_\gamma, \tilde{y}\rangle - \langle \tilde{\xi}, y_\gamma\rangle + \langle \tilde{\xi}, \tilde{y}\rangle. \end{aligned} \tag{3.2}$$

Due to the strong convergence of u_γ in $H^{-1}(\Omega)$ and the weak convergence of (y_γ, ξ_γ) in $H_0^1(\Omega) \times H^{-1}(\Omega)$, the expression on the right hand side of (3.2) converges to $-\langle \tilde{\xi}, \tilde{y}\rangle$, which is nonpositive due to (3.1). Therefore y_γ converges strongly in $H_0^1(\Omega)$. Furthermore we find

$$0 \leq \langle \tilde{\xi}, \tilde{y}\rangle = \lim_{\gamma\to\infty} \langle \xi_\gamma, y_\gamma\rangle \leq \lim_{\gamma\to\infty} \alpha_\gamma = 0$$

and

$$\begin{aligned} 0 \leq \|\xi_\gamma - \tilde{\xi}\|_{H^{-1}} &\leq \|\mathcal{A}(y_\gamma - \tilde{y})\|_{H^{-1}} + \|u_\gamma - \tilde{u}\|_{H^{-1}} \\ &\leq \|\mathcal{A}\|_{\mathcal{L}(H_0^1, H^{-1})}\|y_\gamma - \tilde{y}\|_{H_0^1} + \|u_\gamma - \tilde{u}\|_{H^{-1}} \to 0. \end{aligned}$$

Hence ξ_γ converges strongly in $H^{-1}(\Omega)$. Consequently $(\tilde{y}, \tilde{u}, \tilde{\xi})$ solves the variational inequality (2.1) which implies $\tilde{y} \in \mathcal{X}$ (see section 2) and further $\tilde{\xi} \in L^2(\Omega)$.

Now let $(y^*, u^*, \xi^*) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ be an optimal solution of $(\mathcal{P})$. Note that (y^*, u^*, ξ^*) is feasible for the relaxed-regularized problem $(\mathcal{P}_{\alpha,\gamma})$ and $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is feasible for the original problem $(\mathcal{P})$. We therefore conclude

$$\begin{aligned} J(y^*, u^*) &\leq J(\tilde{y}, \tilde{u}), \\ \tilde{J}_\gamma(y_\gamma, u_\gamma, \xi_\gamma) &\leq \tilde{J}_\gamma(y^*, u^*, \xi^*) \quad \forall \gamma > 0. \end{aligned}$$

Using the lower semi-continuity of J, the definition of $\tilde{J}_\gamma$ and the non-negativity of y^* it follows that

$$\begin{aligned} J(\tilde{y}, \tilde{u}) &\leq \liminf_{\gamma\to\infty} J(y_\gamma, u_\gamma) \leq \liminf_{\gamma\to\infty} \tilde{J}_\gamma(y_\gamma, u_\gamma, \xi_\gamma) \\ &\leq \limsup_{\gamma\to\infty} \tilde{J}_\gamma(y_\gamma, u_\gamma, \xi_\gamma) \leq \limsup_{\gamma\to\infty} \tilde{J}_\gamma(y^*, u^*, \xi^*) = J(y^*, u^*) \\ &\leq J(\tilde{y}, \tilde{u}). \end{aligned}$$

Therefore $\lim_{\gamma\to\infty} \tilde{J}_\gamma(y_\gamma, u_\gamma, \xi_\gamma) = J(\tilde{y}, \tilde{u}) = J(y^*, u^*)$ and $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is optimal for $(\mathcal{P})$. From the convergence of the objective function values, due to the strong convergence of y_γ in $L^2(\Omega)$ we deduce that

$$\frac{1}{2\gamma}\|\max(0, \bar{\lambda} - \gamma y_\gamma)\|^2 \to 0 \quad \wedge \quad \|u_\gamma\|^2 \to \|\tilde{u}\|^2 \quad \wedge \quad \frac{\kappa_\gamma}{2}\|\xi_\gamma\|^2 \to 0.$$

As weak convergence together with norm-convergence in Hilbert spaces imply strong convergence (see, e.g., [2], p. 250), this yields the strong convergence $u_\gamma \to \tilde{u}$ in $L^2(\Omega)$. □

4. Stationary points

In the previous section, our analysis required global solutions of the relaxed-regularized problems. However, finding globally optimal solutions (in particular by means of numerical algorithms) is difficult in practice. Often, one rather has to rely on stationary points, i.e., points satisfying first order optimality conditions, or on local solutions. Concerning stationarity, for finite dimensional MPECs there exists a hierarchy of concepts; see, e.g., [41, 42, 43]. In our present context, the notions of C- and strong stationarity are of particular interest. In fact, based on stationarity for the relaxed-regularized problems (the corresponding conditions can be derived from classical results of mathematical programming theory in Banach spaces) we investigate the behavior of accumulation points of sequences of such stationary points. Depending on very mild assumptions we show that accumulation points are *ε-almost C-stationary* for the original MPCC $(\mathcal{P})$. We also provide conditions for the stronger stationarity concepts. See Definitions 4.2 and 4.3 for detailed descriptions of these new concepts.

4.1. Optimality Systems.

We commence this section by defining suitable stationarity concepts for the original problem $(\mathcal{P})$. Our denotations parallel concepts in finite dimensions; see [41]. For the sake of brevity we set $\Omega^+ := \{\mathbf{x} \in \Omega : y(\mathbf{x}) > 0\}$.

Definition 4.1. (i) *The point $(y, u, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ is called a* C-stationary *point for problem (2.2), if there exist $p \in H_0^1(\Omega)$ and $\lambda \in H^{-1}(\Omega)$ such that the following system of equations is satisfied.*

$$y - \lambda + \mathcal{A}^* p = y_d, \tag{4.1a}$$

$$\nu u - p = 0, \tag{4.1b}$$

$$\mathcal{A} y - u - \xi = f, \tag{4.1c}$$

$$\xi \geq 0 \text{ a.e.}, \; y \geq 0 \text{ a.e.}, \; (y, \xi) = 0, \tag{4.1d}$$

$$\langle \lambda, p \rangle \leq 0, \tag{4.1e}$$

$$p = 0 \text{ a.e. in } \{\xi > 0\}. \tag{4.1f}$$

Furthermore we impose

$$\langle \lambda, \phi \rangle = 0 \; \forall \phi \in H_0^1(\Omega), \; \phi = 0 \text{ a.e. in } \Omega \setminus \Omega^+. \tag{4.2}$$

(ii) *The point (y, u, ξ) is called* strongly stationary, *if (4.1) is satisfied and additionally p and λ have the following sign properties:*

$$p \leq 0 \text{ a.e. in } B, \tag{4.3a}$$

$$\langle \lambda, \phi \rangle \geq 0 \; \forall \phi \in H_0^1(\Omega), \; \phi \geq 0 \text{ a.e. in } B, \phi = 0 \text{ a.e. in } \Omega \setminus (\Omega^+ \cup B), \tag{4.3b}$$

where $B = \{y = 0\} \cap \{\xi = 0\}$ denotes the biactive set.

In [34], Mignot and Puel show that every global solution of $(\mathcal{P})$ satisfies a first order system which is equivalent to the one characterizing strong stationarity. We mention here that the arguments of [34] remain true in the case of local solutions. Their proof technique, however, requires the knowledge of a global (local) solution beforehand and is therefore difficult to realize in solution algorithms. Our subsequent proof technique, on the other hand, does not need knowledge of a global or local solution in advance. Moreover it allows to design a corresponding solution algorithm as it only relies on stationary points (which need not be global or local solutions). Without further assumptions, however, only a weaker form of stationarity can be guaranteed.

Note that the finite dimensional analogue of (4.2) is "$\lambda = 0$ in Ω^+". In function space, however, $\lambda \in H^{-1}(\Omega)$ does not admit a pointwise interpretation. The finite dimensional condition therefore has no unique infinite dimensional counterpart. We introduce weaker forms of the stationarity concepts of Definition 4.1 reflecting this ambiguity.

Definition 4.2. (i) *The point* $(y, u, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ *is called* ε-almost C-stationary, *if there exist* $p \in H_0^1(\Omega)$ *and* $\lambda \in H^{-1}(\Omega)$ *such that* (4.1) *is satisfied and further*

(a) $\langle \lambda, y \rangle = 0$,

(b) *for every* $\varepsilon > 0$ *there exists* $E_\varepsilon \subset \Omega^+$ *with* $meas(\Omega^+ \setminus E_\varepsilon) \leq \varepsilon$ *such that*

$$\langle \lambda, \phi \rangle = 0 \; \forall \phi \in H_0^1(\Omega),\ \phi = 0 \ \text{a.e. in } \Omega \setminus E_\varepsilon. \tag{4.4}$$

(ii) *The point* (y, u, ξ) *is called* ε-almost strongly stationary *if it satisfies* (4.1) *together with*

$$p \leq 0 \ \text{a.e. in } B, \tag{4.5a}$$

$$\langle \lambda, y \rangle = 0, \tag{4.5b}$$

$$\langle \lambda, \phi \rangle \geq 0 \; \forall \phi \in H_0^1(\Omega),\ \phi \geq 0 \ \text{a.e. in } B, \phi = 0 \ \text{a.e. in } \Omega \setminus (E_\varepsilon \cup B), \tag{4.5c}$$

where the sets E_ε *are defined as in* (ib).

We furthermore define a notion lying between ε-stationarity and the concepts defined in Definition 4.1.

Definition 4.3. (i) *The point* (y, u, ξ) *is called* almost C-stationary *if equations* (4.1)*, together with*

$$\begin{aligned} &\langle \lambda, y \rangle = 0, \\ &\langle \lambda, \phi \rangle = 0 \; \forall \phi \in H_0^1(\Omega),\ \phi = 0 \ \text{a.e. in } \Omega \setminus \Omega^+, \phi|_{\Omega^+} \in H_0^1(\Omega^+) \end{aligned} \tag{4.6}$$

are satisfied.

(ii) *The point* (y, u, ξ) *is called* almost strongly-stationary *if* (4.1) *holds true and furthermore*

$$p \leq 0 \ \text{a.e. in } B, \tag{4.7a}$$

$$\langle \lambda, y \rangle = 0, \tag{4.7b}$$

$$\begin{aligned} &\langle \lambda, \phi \rangle \geq 0 \; \forall \phi \in H_0^1(\Omega),\ \phi \geq 0 \ \text{a.e. in } B, \phi = 0 \ \text{a.e. in } \Omega \setminus (\Omega^+ \cup B), \\ &\qquad \max(0, -\phi)|_{\Omega^+} \in H_0^1(\Omega^+). \end{aligned} \tag{4.7c}$$

Remark 4.4. If Ω^+ is a Lipschitz domain, the concepts of Definitions 4.1 and 4.3 coincide; see, e.g., [22]. Furthermore note that while in finite dimensions all three concepts are equivalent, in function space there exists a hierarchy as illustrated below:

$$\begin{array}{ccccc} \text{strong-stat.} & \Rightarrow & \text{alm. strong-stat.} & \Rightarrow & \varepsilon\text{-alm. strong-stat.} \\ \Downarrow & & \Downarrow & & \Downarrow \\ \text{C-stat.} & \Rightarrow & \text{alm. C-stat.} & \Rightarrow & \varepsilon\text{-alm. C-stat.} \end{array}$$

As, due to [34], each global or local solution of $(\mathcal{P})$ is strongly stationary, these optimal points therefore automatically satisfy all weaker notions of stationarity defined in this section.

Next we turn towards the relaxed-regularized problem. Using results due to Zowe and Kurcyusz [44] we are able to formulate necessary optimality conditions for $(\mathcal{P}_{\alpha,\gamma})$.

Corollary 4.5. *Let $(y, u, \xi) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ be an optimal solution of $(\mathcal{P}_{\alpha,\gamma})$. Then there exist Lagrange multipliers $(p, r, \mu) \in H_0^1(\Omega) \times \mathbb{R} \times L^2(\Omega)$ that satisfy the following system of equations:*

$$y - \max(0, \bar{\lambda} - \gamma y) + \mathcal{A}^* p + r\xi = y_d, \tag{4.8a}$$

$$\nu u - p = 0, \tag{4.8b}$$

$$\kappa\xi - p + ry - \mu = 0, \tag{4.8c}$$

$$\xi \geq 0 \text{ a.e.}, \; \mu \geq 0 \text{ a.e.}, \; (\xi, \mu) = 0, \tag{4.8d}$$

$$(y, \xi) \leq \alpha, \; r \geq 0, \; r\left((y, \xi) - \alpha\right) = 0, \tag{4.8e}$$

$$\mathcal{A}y - u - \xi = f. \tag{4.8f}$$

The proof of this corollary can be found in the Appendix. Note that $p \in H_0^1(\Omega)$ and therefore $u \in H_0^1(\Omega)$ due to (4.8b).

4.2. **ε-almost C-stationarity.** For each $\gamma > 0$ we define $\alpha_\gamma > 0$ and $\kappa_\gamma > 0$ such that

$$(\alpha_\gamma, \kappa_\gamma) \overset{\gamma \to \infty}{\longrightarrow} (0, 0) \quad \wedge \quad \max\left\{(\alpha_\gamma \sqrt{\gamma})^{-1}, \kappa_\gamma \sqrt{\gamma}\right\} \leq C \tag{4.9}$$

with $C > 0$ independent of γ. We further assume that the stationary points of the relaxed-regularized problems stay inside some uniformly bounded set. Based on these assumptions we show that limit-points of such a sequence of stationary points for the relaxed-regularized problems are ε-almost C-stationary for the original problem $(\mathcal{P})$. Further, we provide conditions for the limit-points to comply with the stronger stationarity concepts.

Theorem 4.6. *For each $\gamma > 0$ let $\alpha_\gamma > 0$, $\kappa_\gamma > 0$ be given such that* (4.9) *holds true. Further let $(y_\gamma, u_\gamma, \xi_\gamma) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ and $(p_\gamma, r_\gamma, \mu_\gamma) \in H_0^1(\Omega) \times \mathbb{R} \times L^2(\Omega)$ satisfy the optimality system* (4.8) *of the relaxed-regularized problem $(\mathcal{P}_{\alpha,\gamma})$.*

If $\{(u_\gamma, \xi_\gamma)\}$ is uniformly bounded in $L^2(\Omega) \times L^2(\Omega)$, then there exists a point $(\tilde{y}, \tilde{u}, \tilde{\xi}) \in \mathcal{X} \times H_0^1(\Omega) \times L^2(\Omega)$, which is ε-almost C-stationary for $(\mathcal{P})$, and corresponding multipliers $(\tilde{p}, \tilde{\lambda})$ in $H_0^1(\Omega) \times H^{-1}(\Omega)$, such that on a suitable subsequence (denoted the same) $y_\gamma \to \tilde{y}$ in $H_0^1(\Omega)$, $u_\gamma \rightharpoonup \tilde{u}$ in $H_0^1(\Omega)$, $\xi_\gamma \rightharpoonup \tilde{\xi}$ in $L^2(\Omega)$, $p_\gamma \rightharpoonup \tilde{p}$ in $H_0^1(\Omega)$ and $\left(\max(0, \bar{\lambda} - \gamma y_\gamma) - r_\gamma \xi_\gamma\right) \rightharpoonup \tilde{\lambda}$ in $H^{-1}(\Omega)$.

Remark 4.7. Note that $(\tilde{y}, \tilde{u}) \in \mathcal{X} \times H_0^1(\Omega)$. This a posteriori regularity gain is due to (2.1) and (4.1b).

Further note that in Theorem 4.6 the boundedness of ξ_γ in $L^2(\Omega)$ is required, whereas in Theorem 3.2 we obtain convergence (only) in $H^{-1}(\Omega)$ for global solutions. One possible way to circumvent this is the addition of the term $\|\xi_\gamma - \tilde{\xi}\|^2$ to the cost functional of the relaxed-regularized problem, where $\tilde{\xi} \in L^2(\Omega)$ is the solution of the original problem; see, e.g., [5, 7, 34]. This latter technique, however, requires knowledge of the solution $\tilde{\xi}$, which appears to be impractical with respect to designing numerical solution algorithms.

For the proof of Theorem 4.6 we will need some auxiliary results. For this purpose we subsequently assume that the prerequisites of Theorem 4.6 hold true. For each $\gamma > 0$ we define the sets

$$N_\gamma := \{\mathbf{x} \in \Omega : y_\gamma(\mathbf{x}) < 0\},\ P_\gamma := \Omega \setminus N_\gamma,\ \Lambda_\gamma := \{\mathbf{x} \in \Omega : \bar{\lambda}(\mathbf{x}) - \gamma y_\gamma(\mathbf{x}) > 0\} \tag{4.10}$$

and introduce the notation $\lambda_\gamma := \max(0, \bar{\lambda} - \gamma y_\gamma)$.

Lemma 4.8. *Let $\{y_\gamma\}$ be a bounded sequence in $L^2(\Omega)$ such that $\{|(\lambda_\gamma, y_\gamma)|\}$ is bounded. Then the following assertions hold:*

(i) *There exists $C > 0$ independent of γ such that $\gamma \int_{N_\gamma} y_\gamma^2 \leq C$ for all $\gamma > 0$.*

(ii) $\limsup_{\gamma\to\infty} (\lambda_\gamma, y_\gamma) \leq 0$.

Proof. The parameter $\bar{\lambda}$ is pointwise non-negative. Hence, we have $0 \leq \lambda_\gamma \leq \bar{\lambda}$ in P_γ and $\lambda_\gamma = \bar{\lambda} - \gamma y_\gamma$ in N_γ. By assumption there exists $\tilde{C} > 0$ such that $|(\lambda_\gamma, y_\gamma)| \leq \tilde{C}$ for all $\gamma > 0$. Therefore

$$-\tilde{C} \leq (\lambda_\gamma, y_\gamma) \leq \int_{N_\gamma} \bar{\lambda} y_\gamma - \gamma \int_{N_\gamma} y_\gamma^2 + \int_{P_\gamma} \bar{\lambda} y_\gamma = \int_{\Omega} \bar{\lambda} y_\gamma - \gamma \int_{N_\gamma} y_\gamma^2.$$

Consequently,

$$\gamma \int_{N_\gamma} y_\gamma^2 \leq \tilde{C} + (\bar{\lambda}, y_\gamma) \leq \tilde{C} + \|\bar{\lambda}\| \|y_\gamma\|. \tag{4.11}$$

As $\{y_\gamma\}$ is bounded in $L^2(\Omega)$, (i) is established. We further estimate

$$(y_\gamma, \lambda_\gamma) = \int_{\Lambda_\gamma} \bar{\lambda} y_\gamma - \gamma \int_{\Lambda_\gamma} y_\gamma^2 \leq \int_{\{\bar{\lambda} > \gamma y_\gamma \geq 0\}} \bar{\lambda} y_\gamma \leq \frac{1}{\gamma} \int_{\{\bar{\lambda} > \gamma y_\gamma \geq 0\}} \bar{\lambda}^2 \to 0,$$

which completes the proof. □

Now we are ready to prove the main theorem of this section.

Proof of Theorem 4.6. Using the same arguments as in the proofs of Theorems 3.1 and 3.2 we obtain the boundedness of $\{y_\gamma\}$ in $H_0^1(\Omega)$ and the existence of $(\tilde{y}, \tilde{u}, \tilde{\xi}) \in H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega)$ and a subsequence $\{(y_\gamma, u_\gamma, \xi_\gamma)\}$ such that $y_\gamma \to \tilde{y}$ in $H_0^1(\Omega)$ and $(u_\gamma, \xi_\gamma) \rightharpoonup (\tilde{u}, \tilde{\xi})$ in $L^2(\Omega) \times L^2(\Omega)$. Further we find that $(\tilde{y}, \tilde{u}, \tilde{\xi})$ satisfies (4.1c), i.e.,

$$\mathcal{A}\tilde{y} = \tilde{u} + \tilde{\xi} + f \quad \text{in } H^{-1}(\Omega), \tag{4.12}$$

and that

$$\tilde{\xi} \geq 0 \text{ a.e. in } \Omega \quad \wedge \quad (\tilde{y}, \tilde{\xi}) \leq 0 \tag{4.13}$$

holds. Multiplication of (4.8c) by ξ_γ yields:

$$\kappa_\gamma\|\xi_\gamma\|^2 - (p_\gamma, \xi_\gamma) + r_\gamma(y_\gamma, \xi_\gamma) - (\mu_\gamma, \xi_\gamma) = 0.$$

Using the complementarity conditions (4.8d) and (4.8e) and the optimality condition (4.8b) we estimate

$$r_\gamma\alpha_\gamma = (p_\gamma, \xi_\gamma) - \kappa_\gamma\|\xi_\gamma\|^2 \le \nu\|u_\gamma\|\|\xi_\gamma\| - \kappa_\gamma\|\xi_\gamma\|^2. \tag{4.14}$$

Therefore, $\{r_\gamma\alpha_\gamma\}$ is bounded.

Next we show that $\left\{\frac{1}{\sqrt{\gamma}}\lambda_\gamma\right\}$ is bounded in $L^2(\Omega)$. In the case of global solutions the existence of a feasible point is sufficient to guarantee this result. Unfortunately we cannot use the optimality of the cost functional under our present assumptions. Rather we multiply (4.8a) by y_γ and use (4.8f) to find

$$\begin{aligned}(\lambda_\gamma, y_\gamma) &= \|y_\gamma\|^2 + (\mathcal{A}^* p_\gamma, y_\gamma) + r_\gamma(\xi_\gamma, y_\gamma) - (y_d, y_\gamma)\\ &= \|y_\gamma\|^2 + (p_\gamma, u_\gamma + \xi_\gamma + f) + r_\gamma(\xi_\gamma, y_\gamma) - (y_d, y_\gamma).\end{aligned}$$

Using equation (4.8e) we infer

$$|(\lambda_\gamma, y_\gamma)| \le \|y_\gamma\|^2 + \|p_\gamma\|\left(\|u_\gamma\| + \|\xi_\gamma\| + \|f\|\right) + r_\gamma\alpha_\gamma + \|y_d\|\|y_\gamma\|. \tag{4.15}$$

As $\{u_\gamma\}$ and hence $\{p_\gamma\}$ due to (4.8b) are bounded in $L^2(\Omega)$, (4.15) yields the boundedness of $\{|(\lambda_\gamma, y_\gamma)|\}$. We further estimate

$$\|\frac{1}{\sqrt{\gamma}}\lambda_\gamma\|^2 \le \frac{1}{\gamma}\int_{P_\gamma}\bar\lambda^2 + \frac{1}{\gamma}\int_{N_\gamma}(\bar\lambda - \gamma y_\gamma)^2 = \frac{1}{\gamma}\int_{\Omega}\bar\lambda^2 - 2\int_{N_\gamma}\bar\lambda y_\gamma + \gamma\int_{N_\gamma} y_\gamma^2.$$

Lemma 4.8 (i) then yields the boundedness of $\left\{\frac{1}{\sqrt{\gamma}}\lambda_\gamma\right\}$ in $L^2(\Omega)$. Similar to the proof of Theorem 3.2 we obtain

$$\tilde y \ge 0. \tag{4.16}$$

The non-negativity of $\tilde y$ together with (4.13) yields

$$(\tilde y, \tilde\xi) = 0. \tag{4.17}$$

Hence, $(\tilde y, \tilde u, \tilde \xi)$ satisfies (4.1d). This, together with (4.12), implies $\tilde y \in \mathcal{X} = H^2(\Omega)\cap H_0^1(\Omega)$; see section 2. Next we show the boundedness of $\{p_\gamma\}$ in $H_0^1(\Omega)$. Using (H2), equations (4.8a), (4.8c), as well as the non-negativity of λ_γ and μ_γ, we find

$$\begin{aligned}C_c\|p_\gamma\|_{H_0^1}^2 &\le \langle \mathcal{A}^* p_\gamma, p_\gamma\rangle = (\lambda_\gamma - r_\gamma\xi_\gamma, p_\gamma) + (y_d, p_\gamma) - (y_\gamma, p_\gamma)\\ &= (\lambda_\gamma - r_\gamma\xi_\gamma, \kappa_\gamma\xi_\gamma + r_\gamma y_\gamma - \mu_\gamma) + (y_d, p_\gamma) - (y_\gamma, p_\gamma)\\ &\le \kappa_\gamma(\lambda_\gamma, \xi_\gamma) + r_\gamma(\lambda_\gamma, y_\gamma) + \left(\|y_d\| + \|y_\gamma\|\right)\|p_\gamma\|\\ &= \kappa_\gamma(\bar\lambda, \xi_\gamma)_{\Lambda_\gamma} - \kappa_\gamma\gamma(y_\gamma, \xi_\gamma)_{\{\bar\lambda > \gamma y_\gamma \ge 0\}} - \kappa_\gamma\gamma(y_\gamma, \xi_\gamma)_{N_\gamma}\\ &\quad + r_\gamma(\bar\lambda, y_\gamma)_{\{\bar\lambda > \gamma y_\gamma \ge 0\}} + r_\gamma(\bar\lambda, y_\gamma)_{N_\gamma} - r_\gamma\gamma(y_\gamma, y_\gamma)_{\Lambda_\gamma}\\ &\quad + \left(\|y_d\| + \|y_\gamma\|\right)\|p_\gamma\|\\ &\le \kappa_\gamma\|\bar\lambda\|\|\xi_\gamma\| - \kappa_\gamma\sqrt{\gamma}(\sqrt{\gamma}y_\gamma, \xi_\gamma)_{N_\gamma} + \frac{r_\gamma}{\gamma}(\bar\lambda, \bar\lambda)_{\{\bar\lambda > \gamma y_\gamma > 0\}}\\ &\quad + \left(\|y_d\| + \|y_\gamma\|\right)\|p_\gamma\|_{H_0^1}.\end{aligned}$$

Recall that $\{\|\sqrt{\gamma}y_\gamma\|_{L^2(N_\gamma)}\}$ is bounded due to Lemma 4.8, $\{\kappa_\gamma\sqrt{\gamma}\}$ and $\{\frac{r_\gamma}{\gamma}\} = \{\frac{r_\gamma\alpha_\gamma}{\alpha_\gamma\gamma}\}$ are bounded due to (4.9) and (4.14). Hence $\{p_\gamma\}$ is bounded in $H_0^1(\Omega)$ and by (4.8a) $\{\lambda_\gamma - r_\gamma\xi_\gamma\}$ is bounded in $H^{-1}(\Omega)$. Therefore there exist $\tilde p \in H_0^1(\Omega)$ and

$\tilde\lambda \in H^{-1}(\Omega)$ and a subsequence (again denoted by the index γ) such that $p_\gamma \rightharpoonup \tilde p$ in $H_0^1(\Omega)$ and $(\lambda_\gamma - r_\gamma \xi_\gamma) \rightharpoonup \tilde\lambda$ in $H^{-1}(\Omega)$. Equation (4.8b) immediately yields weak convergence of u_γ to $\tilde u$ in $H_0^1(\Omega)$ and, together with (4.8a), we find

$$\tilde y - \tilde\lambda + \mathcal{A}^* \tilde p = y_d \quad \text{in } H^{-1}(\Omega), \tag{4.18}$$

$$\nu \tilde u - \tilde p = 0, \tag{4.19}$$

i.e., (4.1a) and (4.1b) are satisfied.

We next show that $(\tilde p, \tilde\xi) = 0$. From (4.14) it follows that

$$(\tilde p, \tilde\xi) = \lim_{\gamma\to\infty} r_\gamma \alpha_\gamma \geq 0. \tag{4.20}$$

If $\{r_\gamma\}$ is bounded, then the assertion is evident. Let us now assume that $r_\gamma \to \infty$. Using the adjoint equation (4.8a) we deduce that

$$(p_\gamma, \xi_\gamma) = \frac{1}{r_\gamma}\big((p_\gamma, y_d) - (p_\gamma, y_\gamma) - \langle p_\gamma, \mathcal{A}^* p_\gamma\rangle\big) + \frac{1}{r_\gamma}(p_\gamma, \lambda_\gamma). \tag{4.21}$$

The first term of the sum on the right hand side of (4.21) tends to zero, as $\{p_\gamma\}$ and $\{y_\gamma\}$ are bounded in $H_0^1(\Omega)$ and $r_\gamma \to \infty$. Using (4.8c) we find that

$$\frac{1}{r_\gamma}(p_\gamma, \lambda_\gamma) = (y_\gamma, \lambda_\gamma) + \frac{\kappa_\gamma}{r_\gamma}(\xi_\gamma, \lambda_\gamma) - \frac{1}{r_\gamma}(\mu_\gamma, \lambda_\gamma) \leq (y_\gamma, \lambda_\gamma) + \frac{\kappa_\gamma}{r_\gamma}(\xi_\gamma, \lambda_\gamma). \tag{4.22}$$

The first term on the right hand side of (4.22) is bounded from above by Lemma 4.8 (ii). The second term can be estimated as follows:

$$\begin{aligned} 0 \leq \frac{\kappa_\gamma}{r_\gamma}(\xi_\gamma, \lambda_\gamma) &= \frac{\kappa_\gamma}{r_\gamma}\left(\int_{\Lambda_\gamma} \bar\lambda \xi_\gamma - \gamma \int_{\Lambda_\gamma} y_\gamma \xi_\gamma\right) \\ &\leq \frac{1}{r_\gamma}\left(\kappa_\gamma(\bar\lambda, \xi_\gamma) - \kappa_\gamma \gamma \int_{N_\gamma} y_\gamma \xi_\gamma\right) \\ &\leq \frac{1}{r_\gamma}\left(\kappa_\gamma(\bar\lambda, \xi_\gamma) + \|\sqrt{\gamma} y_\gamma\|_{L^2(N_\gamma)} \kappa_\gamma \sqrt{\gamma}\|\xi_\gamma\|\right). \end{aligned} \tag{4.23}$$

Due to the boundedness of $\{\|\sqrt{\gamma} y_\gamma\|_{L^2(N_\gamma)}\}$ (c.f. Lemma 4.8) and $\{\kappa_\gamma \sqrt{\gamma}\}$ (c.f. (4.9)), the expression tends to zero. Inserting (4.22) and (4.23) into (4.21), we find that $(\tilde p, \tilde\xi) = \lim_{\gamma\to\infty}(p_\gamma, \xi_\gamma) \leq 0$. Therefore, by (4.20) it follows that

$$(\tilde p, \tilde\xi) = 0 \quad \wedge \quad r_\gamma \alpha_\gamma \to 0 \text{ as } \gamma \to \infty. \tag{4.24}$$

Note that due to (4.9), (4.24) and Lemma 4.8 we find that

$$-r_\gamma\big(\xi_\gamma, \max(0, -y_\gamma)\big) = r_\gamma \alpha_\gamma (\alpha_\gamma \sqrt{\gamma})^{-1} (\xi_\gamma, \sqrt{\gamma} y_\gamma)_{L^2(N_\gamma)} \to 0. \tag{4.25}$$

Consequently this, together with (4.8e) and (4.24), implies

$$\lim_{\gamma\to\infty} r_\gamma(\xi_\gamma, y_\gamma^+) = \lim_{\gamma\to\infty} r_\gamma(\xi_\gamma, y_\gamma^-) = 0, \tag{4.26}$$

where $y_\gamma^+ := \max(0, y_\gamma)$ and $y_\gamma^- := \max(0, -y_\gamma)$. Now let $\omega \subset \Omega$ be an arbitrary subset. Then (4.8c) yields

$$(p_\gamma, \xi_\gamma)_\omega = \kappa_\gamma \|\xi_\gamma\|^2_{L^2(\omega)} + r_\gamma(y_\gamma, \xi_\gamma)_\omega - (\mu_\gamma, \xi_\gamma)_\omega,$$

where $(\cdot,\cdot)_\omega$ denotes the inner product in $L^2(\omega)$. Due to the boundedness of $\{\xi_\gamma\}$ in $L^2(\omega)$, (4.26) and (4.8d) we find that the right hand side of the above equation tends to zero, as $\gamma \to \infty$. Hence

$$(\tilde{p}, \tilde{\xi})_\omega = 0 \;\forall\, \omega \subset \Omega. \tag{4.27}$$

If we choose $\omega = \{\tilde{p} > 0\}$ and $\omega = \{\tilde{p} < 0\}$, respectively, in (4.27), we find that $(\tilde{p}^+, \tilde{\xi}) = (\tilde{p}^-, \tilde{\xi}) = 0$ and therefore

$$\tilde{p} = 0 \text{ a.e. in } \{\tilde{\xi} > 0\},$$

i.e., (4.1f) holds.

We next prove that $\langle \tilde{\lambda}, \tilde{y}\rangle = 0$. Using (4.24) and (4.8e), it follows from Lemma 4.8 (ii) that

$$\langle \tilde{\lambda}, \tilde{y}\rangle = \lim_{\gamma\to\infty} \big((\lambda_\gamma, y_\gamma) - r_\gamma(\xi_\gamma, y_\gamma)\big) = \lim_{\gamma\to\infty} (\lambda_\gamma, y_\gamma) \le 0. \tag{4.28}$$

On the other hand, as $y_\gamma \to \tilde{y}$ in $H_0^1(\Omega)$ also $y_\gamma^+ \to \tilde{y}^+$ in $H_0^1(\Omega)$. For the proof of this property see Appendix B. Due to (4.16) and (4.26) we then find that

$$\langle \tilde{\lambda}, \tilde{y}\rangle = \langle \tilde{\lambda}, \tilde{y}^+\rangle = \lim_{\gamma\to\infty} \big((\lambda_\gamma, y_\gamma^+) - r_\gamma(\xi_\gamma, y_\gamma^+)\big) = \lim_{\gamma\to\infty} (\lambda_\gamma, y_\gamma^+) \ge 0.$$

Hence, from (4.28) it follows that

$$\langle \tilde{\lambda}, \tilde{y}\rangle = 0. \tag{4.29}$$

As a direct consequence we obtain

$$\gamma(y_\gamma, y_\gamma)_{L^2(\Lambda_\gamma)} \to 0, \tag{4.30}$$

as, due to the definition of Λ_γ (see (4.10)) and (4.29), we can estimate

$$\begin{aligned} 0 \le \lim_{\gamma\to 0} \gamma(y_\gamma, y_\gamma)_{L^2(\Lambda_\gamma)} &= \lim_{\gamma\to\infty} \big((\bar{\lambda}, y_\gamma)_{L^2(\Lambda_\gamma)} - (\lambda_\gamma, y_\gamma)\big) \\ &\le \lim_{\gamma\to\infty} \big(\gamma^{-1}(\bar{\lambda}, \bar{\lambda})_{L^2(\Lambda_\gamma)} - (\lambda_\gamma, y_\gamma)\big) = 0. \end{aligned}$$

Next we show that $(\lambda_\gamma - r_\gamma \xi_\gamma) \to 0$ point-wise a.e. in $\Omega^+ = \{\tilde{y} > 0\}$. We begin by examining $\lambda_\gamma = \gamma \max(0, \frac{\bar{\lambda}}{\gamma} - y_\gamma)$. We know that (on a subsequence) $y_\gamma \to \tilde{y}$ point-wise a.e. in Ω. Hence for almost every $\mathbf{x} \in \Omega^+$ the quantity $(\frac{\bar{\lambda}}{\gamma} - y_\gamma)(\mathbf{x}) < 0$ for γ sufficiently large. Therefore

$$\lambda_\gamma \to 0 \;\text{ point-wise a.e. in } \Omega^+. \tag{4.31}$$

Due to (4.26) we find that

$$\lim_{\gamma\to\infty} \|r_\gamma \xi_\gamma y_\gamma\|_{L^1(\Omega^+)} = 0.$$

Therefore there exists a further subsequence (without loss of generality denoted the same) such that $r_\gamma \xi_\gamma y_\gamma \to 0$ point-wise a.e. in Ω^+. As y_γ converges point-wise on that subset to a strictly positive value, we can deduce that

$$r_\gamma \xi_\gamma \to 0 \;\text{ point-wise a.e. in } \Omega^+. \tag{4.32}$$

Combining (4.31) with (4.32) we find that

$$(\lambda_\gamma - r_\gamma \xi_\gamma) \to 0 \;\text{ point-wise a.e. in } \Omega^+. \tag{4.33}$$

Due to Egorov's Theorem (see, e.g., [2]) the quantity $(\lambda_\gamma - r_\gamma \xi_\gamma)|_{\Omega^+}$ then converges uniformly with respect to the underlying measure to zero, i.e., for every $\varepsilon > 0$ there exists a subset $E_\varepsilon \subset \Omega^+$ with $\text{meas}\,(\Omega^+ \setminus E_\varepsilon) \le \varepsilon$, such that $(\lambda_\gamma - r_\gamma \xi_\gamma) \to 0$ uniformly in E_ε. For every $\varepsilon > 0$ this yields

$$\langle \tilde{\lambda}, \varphi \rangle = \lim_{\gamma\to\infty} \langle \lambda_\gamma - r_\gamma \xi_\gamma, \varphi \rangle = \lim_{\gamma\to\infty} \int_\Omega (\lambda_\gamma - r_\gamma \xi_\gamma)\, \varphi = \lim_{\gamma\to\infty} \int_{E_\varepsilon} (\lambda_\gamma - r_\gamma \xi_\gamma)\, \varphi = 0$$

for all $\varphi \in H_0^1(\Omega)$ with $\varphi = 0$ a.e. in $\Omega \setminus E_\varepsilon$. This implies (ib) in Definition 4.2

In order to prove ε-almost C-stationarity of the limit point $(\tilde{y}, \tilde{u}, \tilde{\xi})$ it remains to show that $\langle \tilde{\lambda}, \tilde{p} \rangle \le 0$. Using equation (4.8c) we see that

$$(p_\gamma, \xi_\gamma) = \kappa_\gamma \|\xi_\gamma\|^2 + r_\gamma (y_\gamma, \xi_\gamma) - (\mu_\gamma, \xi_\gamma) = \kappa_\gamma \|\xi_\gamma\|^2 + r_\gamma \alpha_\gamma \ge 0.$$

Utilizing (4.8c) again, we find

$$\begin{aligned}
(\lambda_\gamma - r_\gamma \xi_\gamma, p_\gamma) &\le (\lambda_\gamma, p_\gamma) = (\bar{\lambda}, p_\gamma)_{\Lambda_\gamma} - \gamma (y_\gamma, p_\gamma)_{\Lambda_\gamma} \\
&= (\bar{\lambda}, p_\gamma)_{\Lambda_\gamma} - \gamma \left(\kappa_\gamma (y_\gamma, \xi_\gamma)_{\Lambda_\gamma} + r_\gamma (y_\gamma, y_\gamma)_{\Lambda_\gamma} - (y_\gamma, \mu_\gamma)_{\Lambda_\gamma} \right) \\
&\le (\bar{\lambda}, p_\gamma)_{\Lambda_\gamma} - \gamma \kappa_\gamma (y_\gamma, \xi_\gamma)_{\Lambda_\gamma} + (\bar{\lambda}, \mu_\gamma)_{\Lambda_\gamma} \\
&= (\bar{\lambda}, p_\gamma)_{\Lambda_\gamma} - \gamma \kappa_\gamma (y_\gamma, \xi_\gamma)_{\Lambda_\gamma} + \kappa_\gamma (\bar{\lambda}, \xi_\gamma)_{\Lambda_\gamma} - (\bar{\lambda}, p_\gamma)_{\Lambda_\gamma} + r_\gamma (\bar{\lambda}, y_\gamma)_{\Lambda_\gamma} \\
&\le -\gamma \kappa_\gamma (y_\gamma, \xi_\gamma)_{\Lambda_\gamma} + \kappa_\gamma (\bar{\lambda}, \xi_\gamma)_{\Lambda_\gamma} + \frac{r_\gamma}{\gamma} (\bar{\lambda}, \bar{\lambda})_{\Lambda_\gamma}.
\end{aligned}$$

The first term on the right hand side above tends to zero by (4.9) and (4.30), the second term vanishes because $\{\xi_\gamma\}$ is bounded and the last term tends to zero due to (4.9) and (4.24). Therefore

$$\limsup_{\gamma\to\infty} \langle \lambda_\gamma - r_\gamma \xi_\gamma, p_\gamma \rangle \le 0.$$

For the $H_0^1(\Omega)$-weakly convergent sequence $p_\gamma \rightharpoonup \tilde{p}$ we obtain

$$\langle \mathcal{A}^* \tilde{p}, \tilde{p} \rangle \le \liminf_{\gamma\to\infty} \langle \mathcal{A}^* p_\gamma, p_\gamma \rangle,$$

as the bilinear form $a(\cdot,\cdot)$ defines a norm on $H_0^1(\Omega)$. Using the strong convergence of y_γ in $H_0^1(\Omega)$ and the adjoint equations of both the relaxed-regularized problem and the original MPCC, (4.8a) and (4.18), we find

$$\begin{aligned}
\langle \tilde{\lambda}, \tilde{p} \rangle = \langle \tilde{y} - y_d, \tilde{p} \rangle + \langle \mathcal{A}^* \tilde{p}, \tilde{p} \rangle &\le \liminf_{\gamma\to\infty} \big(\langle y_\gamma - y_d, p_\gamma \rangle + \langle \mathcal{A}^* p_\gamma, p_\gamma \rangle \big) \\
= \liminf_{\gamma\to\infty} \langle \lambda_\gamma - r_\gamma \xi_\gamma, p_\gamma \rangle \le 0,
\end{aligned}$$

which completes the proof. □

Note that our ε-almost C-stationarity result relies on rather mild assumptions concerning the convergence behavior of the sequences $\{\alpha_\gamma\}$ and $\{\kappa_\gamma\}$. For the stronger concepts, however, we have to impose further conditions. We conclude this section by discussing some of these conditions.

4.3. **From ε-almost- to almost- and C-stationarity.**

Lemma 4.9. *Let the assumptions of Theorem 4.6 hold true. If*

$$\langle \lambda_\gamma, \tilde{y} \rangle \to 0 \;\; \textit{or equivalently} \;\; r_\gamma (\xi_\gamma, \tilde{y}) \to 0 \;\; \textit{as } \gamma \to \infty \tag{4.34}$$

then $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is almost C-stationary for $(\mathcal{P})$. If furthermore Ω^+ is a Lipschitz domain, then $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is C-stationary.

Proof. The equivalence in (4.34) follows from (4.29). Now let $\omega \subset \Omega^+$ be a compact subset. As $\tilde{y} \in H^2(\Omega) \cap H_0^1(\Omega)$ and $H^2(\Omega)$ embeds into $C(\Omega)$ for $n \leq 3$, we find that there exists $\epsilon > 0$, such that

$$\tilde{y} \geq \epsilon \text{ in } \omega.$$

The non-negativity of λ_γ, together with (4.34) then yields

$$0 \leq \|\lambda_\gamma\|_{L^1(\omega)} \leq \frac{1}{\epsilon}\int_\omega \lambda_\gamma \tilde{y} \leq \frac{1}{\epsilon}\int_\Omega \lambda_\gamma \tilde{y} \to 0.$$

Analogously we find that $r_\gamma \|\xi_\gamma\|_{L^1(\omega)} \to 0$. As ω was chosen arbitrarily, we conclude that for every $\phi \in C_\circ^\infty(\Omega^+)$, with $\hat{\phi}$ being the trivial extension of ϕ to Ω,

$$\begin{aligned} |\langle \tilde{\lambda}, \hat{\phi}\rangle| &= |\lim_{\gamma\to\infty}(\lambda_\gamma - r_\gamma \xi_\gamma, \hat{\phi})| \\ &\leq \max_{\mathbf{x}\in\Omega^+}|\phi(\mathbf{x})| \lim_{\gamma\to\infty}\left(\|\lambda_\gamma\|_{L^1(\mathrm{supp}\phi)} + r_\gamma\|\xi_\gamma\|_{L^1(\mathrm{supp}\phi)}\right) = 0. \end{aligned}$$

The density of $C_\circ^\infty(\Omega^+)$ in $H_0^1(\Omega^+)$ then implies (4.6). If Ω^+ is a Lipschitz-domain then $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is C-stationary; see Remark 4.4. □

Remark 4.10. Note that (4.34) is satisfied if $\{r_\gamma\}$ is bounded. The boundedness of $\{r_\gamma\}$ could numerically be verified for a variety of test problems, including degenerate problems, and ones that violated strict complementarity.

An alternative sufficient condition for (4.34) can be formulated using the convergence speed of the state variable y_γ. In particular, if

$$\|y_\gamma - \tilde{y}\| = \mathcal{O}(\gamma^{-1/2}), \tag{4.35}$$

we find that, due to (4.9) and (4.24),

$$|r_\gamma(\xi_\gamma, \tilde{y})| \leq r_\gamma(\xi_\gamma, y_\gamma) + r_\gamma(\xi_\gamma, |\tilde{y} - y_\gamma|) \leq r_\gamma(\xi_\gamma, y_\gamma) + Cr_\gamma\gamma^{-1/2}\|\xi_\gamma\| \to 0,$$

where $C > 0$ is independent of γ. Condition (4.35) was also verified for our test problems; see figure 6.1.

Next we focus on strong stationarity. The conditions we consider essentially deal with the behavior of the quantities $r_\gamma y_\gamma$ and $r_\gamma \xi_\gamma$ on the biactive set

$$B = \{\tilde{y} = 0\} \cap \{\tilde{\xi} = 0\}.$$

4.4. **From C- to strong stationarity.**

Lemma 4.11. *Let the assumptions of Theorem 4.6 be satisfied. Furthermore let*

$$r_\gamma(y_\gamma, v)_B \to 0 \ \forall v \in L^2(\Omega), \tag{4.36a}$$

$$r_\gamma(\xi_\gamma, \phi)_{B\cup\Omega^+} \to 0 \ \forall \phi \in H_0^1(\Omega). \tag{4.36b}$$

Then $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is ε-almost strongly stationary. Furthermore if $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is almost C-stationary or C-stationary, then (4.36) *implies almost strong stationarity or strong stationarity, respectively.*

Proof. Let $v \in L^2(\Omega)$, $v \geq 0$ in B. Then due to the non-negativity of μ_γ we find that

$$(\tilde{p}, v)_B = \lim_{\gamma\to\infty}(\kappa_\gamma\xi_\gamma + r_\gamma y_\gamma - \mu_\gamma, v)_B \leq \lim_{\gamma\to\infty}(\kappa_\gamma\xi_\gamma + r_\gamma y_\gamma, v)_B = 0.$$

As v was chosen arbitrarily this proves the sign condition for $\tilde{p}$.

For the condition on $\tilde{\lambda}$ we give the proof for the case of an ε-almost C-stationary point. The proofs for almost C-stationarity and C-stationarity are similar. Let $\varepsilon > 0$ be given. As $(\tilde{y}, \tilde{u}, \tilde{\xi})$ is ε-almost C-stationary there exists a set E_ε as in Definition 4.2. Now let $\phi \in H_0^1(\Omega)$ be given as in (4.5c). Note that $\phi^- \in H_0^1(\Omega)$ vanishes a.e. outside of E_ε. Hence

$$\langle \tilde{\lambda}, \phi \rangle = \lim_{\gamma \to \infty} \big((\lambda_\gamma, \phi^+) - r_\gamma (\xi_\gamma, \phi^+) \big) - \langle \tilde{\lambda}, \phi^- \rangle = \lim_{\gamma \to \infty} (\lambda_\gamma, \phi^+) \geq 0.$$

□

Note that the assumptions of Lemma 4.11 are again satisfied in the case of a bounded sequence $\{r_\gamma\}$.

Remark 4.12. Condition (4.36b) seems rather restrictive. But in fact we have already established that for every $\varepsilon > 0$ there exists $E_\varepsilon \subset \Omega^+$ with $\operatorname{meas}(\Omega^+ \setminus E_\varepsilon) \leq \varepsilon$ such that $r_\gamma \xi_\gamma$ converges to zero uniformly on E_ε. Therefore condition (4.36b) can be weakened to

$$r_\gamma (\xi_\gamma, \phi)_B \to 0 \ \forall \phi \in H_0^1(\Omega) \tag{4.37}$$

in the case of ε-almost C-stationarity, and to

$$r_\gamma (\xi_\gamma, \phi)_{B \cup (\Omega^+ \setminus E_\varepsilon)} \to 0 \ \forall \phi \in H_0^1(\Omega) \tag{4.38}$$

for some $\varepsilon > 0$ in the case of almost C-stationarity and C-stationarity.

We will now give an alternative condition for the satisfaction of the sign property of $\tilde{p}$ in the biactive set (replacing (4.36a)). For this purpose we define the sets

$$Y_\gamma := \{\mathbf{x} \in \Omega : y_\gamma(\mathbf{x}) \leq \alpha_\gamma\}, \ \tilde{Y} := \{\mathbf{x} \in \Omega : \tilde{y}(\mathbf{x}) = 0\}.$$

Using (4.8c), (4.8d) and (4.24) we find that

$$\begin{aligned} 0 \leq (p_\gamma, \chi_{Y_\gamma} p_\gamma^+) &= \kappa_\gamma (\xi_\gamma, \chi_{Y_\gamma} p_\gamma^+) + r_\gamma (y_\gamma, \chi_{Y_\gamma} p_\gamma^+) - (\mu_\gamma, \chi_{Y_\gamma} p_\gamma^+) \\ &\leq \kappa_\gamma (\xi_\gamma, \chi_{Y_\gamma} p_\gamma^+) + r_\gamma \alpha_\gamma \| \chi_{Y_\gamma} p_\gamma^+ \|_{L^1(\Omega)} \to 0, \end{aligned}$$

i.e.,

$$\lim_{\gamma \to \infty} \int_{Y_\gamma} \max(0, p_\gamma)^2 = 0, \tag{4.39}$$

where χ_{Y_γ} denotes the characteristic function of the set Y_γ. As $\tilde{p} = 0$ a.e. in $\{\tilde{\xi} > 0\}$, we find that due to (4.39) the sign condition for $\tilde{p}$ in the biactive set in (4.3) is in fact equivalent to

$$\int_{\tilde{Y}} \max(0, \tilde{p})^2 = \lim_{\gamma \to \infty} \int_{Y_\gamma} \max(0, p_\gamma)^2. \tag{4.40}$$

Note that due to the convergence of p_γ to $\tilde{p}$ in $H_0^1(\Omega)$, the compact embedding of $H_0^1(\Omega)$ into $L^q(\Omega)$ for $1 - \frac{n}{2} > -\frac{n}{q}$ (see, e.g., see [1, 2]) and the Lipschitz continuity of the $\max(0, \cdot)$-operator, we find that

$$\max(0, p_\gamma) \to \max(0, \tilde{p}) \text{ in } L^q(\Omega) \text{ for all } 1 - \frac{n}{2} > -\frac{n}{q}. \tag{4.41}$$

The satisfaction of equation (4.40) therefore depends on the behavior of the sets Y_γ. We give a sufficient condition for (4.40) using a notion of set convergence; see [22].

Definition 4.13. *Let* $\{E_k\}_{k\geq 0}$ *and* E *be measurable subsets of* $\mathbb{R}^n$*. The sequence* $\{E_k\}_{k\geq 0}$ *is said to converge to* E in the sense of characteristic functions*, if*

$$\chi_{E_k} \to \chi_E \quad \textit{in } L^s_{loc}(\mathbb{R}^n)\ \forall s \in [1,\infty) \textit{ as } k \to \infty.$$

Lemma 4.14. *If* $Y_\gamma \to \tilde{Y}$ *in the sense of characteristic functions, then* (4.40) *holds.*

Proof. Let $q > 2$ satisfy $1 - \frac{n}{2} > -\frac{n}{q}$. Further define t and $s \in \mathbb{R}$ such that

$$\frac{1}{q} + \frac{1}{t} = 1,\ \frac{1}{q} + \frac{1}{s} = \frac{1}{t},$$

i.e.,

$$t = \frac{q}{q-1},\ s = \frac{q}{q-2}.$$

Then

$$\begin{aligned}\|\chi_{Y_\gamma} p_\gamma^+ - \chi_{\tilde{Y}} \tilde{p}^+\|_{L^t} &\leq \|\left(\chi_{Y_\gamma} - \chi_{\tilde{Y}}\right) p_\gamma^+\|_{L^t} + \|\chi_{\tilde{Y}}\left(p_\gamma^+ - \tilde{p}^+\right)\|_{L^t}\\ &\leq \|\chi_{Y_\gamma} - \chi_{\tilde{Y}}\|_{L^s}\|p_\gamma^+\|_{L^q} + \|\chi_{\tilde{Y}}\|_{L^s}\|p_\gamma^+ - \tilde{p}^+\|_{L^q},\end{aligned}$$

where $\|\chi_{Y_\gamma} - \chi_{\tilde{Y}}\|_{L^s} \to 0$ due to the convergence of the sets in the sense of characteristic functions and $\|p_\gamma^+ - \tilde{p}^+\|_{L^q} \to 0$ due to (4.41). Therefore

$$\chi_{Y_\gamma} p_\gamma^+ \to \chi_{\tilde{Y}} \tilde{p}^+ \text{ in } L^t(\Omega).$$

Using (4.41) we can further estimate

$$\begin{aligned}|(\chi_{Y_\gamma} p_\gamma^+, p_\gamma^+) - (\chi_{\tilde{Y}} \tilde{p}^+, \tilde{p}^+)| =& |(\chi_{Y_\gamma} p_\gamma^+, p_\gamma^+ - \tilde{p}^+) + (\chi_{Y_\gamma} p_\gamma^+ - \chi_{\tilde{Y}} \tilde{p}^+, \tilde{p}^+)|\\ \leq& \|\chi_{Y_\gamma} p_\gamma^+\|_{L^t}\|p_\gamma^+ - \tilde{p}^+\|_{L^q} +\\ & \|\chi_{Y_\gamma} p_\gamma^+ - \chi_{\tilde{Y}} \tilde{p}^+\|_{L^t}\|\tilde{p}^+\|_{L^q} \to 0.\end{aligned}$$

□

5. The Algorithm

Theorem 4.6 proves the convergence of stationary points of the regularized problem $(\mathcal{P}_{\alpha,\gamma})$ to an ε-almost C-stationary point of the MPEC $(\mathcal{P})$. The nature of the proof technique allows the construction of a solution algorithm that exhibits the same function space convergence properties as stated in Theorem 4.6.

5.1. The outer loop.

In this section we specify the outer loop for the solution of $(\mathcal{P})$. The regularization parameter γ is initialized by $\gamma_0 > 0$ and increased by a factor $\beta_\gamma > 1$ after each outer iteration. The quantities α and κ are updated such that (4.9) is satisfied. In particular we choose

$$\kappa = \gamma^{-1/2} \quad \wedge \quad \alpha = \alpha_0 \left(\frac{\gamma_0}{\gamma}\right)^{\frac{1}{2}}. \tag{5.1}$$

The relaxation parameter α is initialized by $\alpha_0 := (y_0, \xi_0)$, where y_0 and ξ_0 are initial values determined as described below in section 5.3. The subsequent outer iterations are initialized by the solutions of their respective preceding outer iteration. The outer loop is described in Algorithm 1.

Remark 5.1. From Theorem 4.6 it follows that every accumulation point of the sequence $\{(y^k, u^k, \xi^k, r_k)\}$ determined in Algorithm 1 is ε-almost C-stationary for $(\mathcal{P})$.

Algorithm 1 (Outer loop)

Data: y_d, $f \in L^2(\Omega)$, $\bar{\lambda} \in L^q(\Omega)$ with $q > 2$, $\bar{\lambda} \geq 0$, $c_r > 0$.

1: Choose $(y^0, u^0, \xi^0, r_0) \in H_0^1(\Omega) \times H_0^1(\Omega) \times L^2(\Omega) \times \mathbb{R}_+$, $(\gamma_0, \alpha_0, \kappa_0) > 0$, $\beta_\gamma > 1$ and set $k := 0$.
2: **repeat**
3: Compute a stationary point $(y^{k+1}, u^{k+1}, \xi^{k+1}, r_{k+1})$ of $(\mathcal{P}_{\alpha_k,\gamma_k})$ with $\kappa = \kappa_k$, using an iterative scheme with initial values (y^k, u^k, ξ^k, r_k).
4: Set $\gamma_{k+1} := \beta_\gamma \gamma_k$ and update α and κ according to (5.1).
5: **until** some stopping rule is satisfied.

5.2. **Solving the subproblem.** Step 3 of Algorithm 1 requires the computation of a stationary point of the relaxed-regularized subproblem. In this section we propose a semismooth Newton method for the solution of the optimality system (4.8).

Note that the complementarity conditions (4.8d) and (4.8e) can be reformulated using the $\max(0,\cdot)$-operator. For arbitrary positive constants c_μ and c_r, (4.8d) and (4.8e) are equivalent to

$$\mu - \max(0, \mu - c_\mu \xi) = 0,$$
$$r - \max\left(0, r + c_r\left((y, \xi) - \alpha\right)\right) = 0,$$

respectively. Using equations (4.8b) and (4.8c) we eliminate the multipliers p and μ and set $c_\mu := \kappa$. This leads to the system

$$F(y, u, \xi, r) = 0 \tag{5.2}$$

with $F : H_0^1(\Omega) \times H_0^1(\Omega) \times L^2 \times \mathbb{R} \to H^{-1}(\Omega) \times L^2(\Omega) \times \mathbb{R} \times H^{-1}(\Omega)$ and

$$F(y, u, \xi, r) = \begin{pmatrix} y - \max(0, \bar{\lambda} - \gamma y) + \nu \mathcal{A}^* u + r\xi - y_d \\ \kappa\xi - \nu u + ry - \max(0, ry - \nu u) \\ r - \max(0, r + c_r((y, \xi) - \alpha)) \\ \mathcal{A}y - u - \xi - f \end{pmatrix}. \tag{5.3}$$

Note that due to the max-operations involved in (5.3), F is not necessarily Fréchet-differentiable. However, it turns out that it admits a weaker derivative. For the sake of recalling the definition of a suitable derivative we proceed in general terms and let X and Z be Banach spaces, $D \subset X$ an open subset of X and $F : D \to Z$.

Definition 5.2. [13, 24] *The mapping $F : D \subset X \to Z$ is called Newton-differentiable in the open subset $U \subset D$, if there exists a family of mappings $G : U \to \mathcal{L}(X, Z)$ such that*

$$\lim_{h \to 0} \frac{1}{\|h\|} \|F(x + h) - F(x) - G(x + h)h\| = 0$$

for every $x \in U$.

We refer to G as the *Newton derivative* or *generalized derivative* for F in U. Note that G is not required to be unique to be a generalized derivative for F in U. We also point out that Definition 5.2 resembles the concept of *semismoothness* known in finite dimensional space [33, 38]. In [24] it was shown that

$$G_\delta(y)(\mathbf{x}) = \begin{cases} 1 & \text{if} \quad y(\mathbf{x}) > 0, \\ 0 & \text{if} \quad y(\mathbf{x}) < 0, \\ \delta & \text{if} \quad y(\mathbf{x}) = 0, \end{cases} \tag{5.4}$$

for $y \in X$ and $\delta \in \mathbb{R}$ is a Newton-derivative of $\max(0,\cdot) : L^{q_1}(\Omega) \to L^{q_2}(\Omega)$, if $1 \le q_2 < q_1 \le \infty$.

Now assume that we are interested in finding $x^* \in X$ such that

$$F(x) = 0. \tag{5.5}$$

Then one may apply a generalized version of Newton's method for computing x^*; see (5.6) below. The following result can be found in [13]; see also [24].

Theorem 5.3. *Suppose that x^* is a solution of* (5.5) *and that F is Newton-differentiable in an open neighborhood U containing x^* with a Newton-derivative $G(x)$. If $G(x)$ is nonsingular for all $x \in U$ and $\{\|G(x)^{-1}\| : x \in U\}$ is bounded, then the semismooth Newton iteration*

$$x^{k+1} = x^k - G(x^k)^{-1}F(x^k) \tag{5.6}$$

converges superlinearly to x^*, provided that $\|x^0 - x^*\|$ is sufficiently small.

Now we are ready to define the semismooth Newton algorithm for (5.2). Let $(y, u, \xi, r) \in H_0^1(\Omega) \times H_0^1(\Omega) \times L^2(\Omega) \times \mathbb{R}$ denote the current iterate. We define the following sets:

$$\begin{aligned} A_y &:= \{\mathbf{x} \in \Omega : \bar\lambda(\mathbf{x}) - \gamma y(\mathbf{x}) > 0\}, \\ A_\mu &:= \{\mathbf{x} \in \Omega : r y(\mathbf{x}) - \nu u(\mathbf{x}) > 0\}, \\ I_\mu &:= \Omega \setminus A_\mu, \\ x_r &:= \begin{cases} 1 & \text{if} \quad r + c_r\left((y,\xi)_{L^2} - \alpha\right) > 0, \\ 0 & \text{else.} \end{cases} \end{aligned}$$

Utilizing G_0 in (5.4) as the Newton-derivative of $\max(\cdot, 0)$ and setting $\delta x^k := x^{k+1} - x^k$, it is straightforward to show that the Newton iteration (5.6) is equivalent to the system

$$\begin{aligned} &(I + \gamma\chi_{A_{y^k}})\delta y^k + \nu\mathcal{A}^*\delta u^k + r_k\delta\xi^k + \delta r_k\xi^k = \\ &\quad = -y^k + \chi_{A_{y^k}}(\bar\lambda - \gamma y^k) - \nu\mathcal{A}^* u^k - r_k\xi^k + y_d, \end{aligned} \tag{5.7a}$$

$$\begin{aligned} &\kappa\delta\xi^k - \nu\delta u^k + r_k\delta y^k + \delta r_k y^k - \chi_{A_{\mu^k}}(r_k\delta y^k + \delta r_k y^k - \nu\delta u^k) = \\ &\quad = -\kappa\xi^k + \nu u^k - r_k y^k + \chi_{A_{\mu^k}}(r_k y^k - \nu u^k), \end{aligned} \tag{5.7b}$$

$$\begin{aligned} &(1 - x_{r_k})\delta r_k - x_{r_k}c_r\left((\delta y^k, \xi^k) + (y^k, \delta\xi^k)\right) = \\ &\quad = -r_k + x_{r_k}\left(r_k + c_r\left((y^k,\xi^k)_{L^2} - \alpha\right)\right), \end{aligned} \tag{5.7c}$$

$$\mathcal{A}\delta y^k - \delta u^k - \delta\xi^k = -\mathcal{A}y^k + u^k + \xi^k + f, \tag{5.7d}$$

where $\chi_{A_{y^k}}$ and $\chi_{A_{\mu^k}}$ denote the characteristic functions of the sets A_{y^k} and A_{μ^k} respectively. As stated in Theorem 5.3, the semismooth Newton method is a locally convergent method only. One possible way to globalize the method is by using backtracking along a so called *Newton path* $\mathfrak{p}$. In the case of semismooth functions, a descent property along such a suitably chosen path can be guaranteed; see, e.g., [14, 15, 39]. To be specific, fix $x \in X$. Then the corresponding path $\mathfrak{p}_x : [0,1] \to X$ is defined by

$$G(\mathfrak{p}_x(\tau))(\mathfrak{p}_x(\tau) - x) = -\tau F(x), \quad 0 \le \tau \le 1, \tag{5.8}$$

where G is the Newton-derivative of F. The resulting method is summarized in Algorithm 2.

Algorithm 2 (Path Newton Method)

Data: y_d, $f \in L^2(\Omega)$, $\bar{\lambda} \in L^q(\Omega)$ with $q > 2$, $\bar{\lambda} \geq 0$, $(\gamma, \alpha, \kappa) > 0$, $c_r > 0$.

1: Choose $x^0 = (y^0, u^0, \xi^0, r_0)$, $\sigma \in (0,1)$, $\beta \in (0,1)$, $tol > 0$ and set $k := 0$.

2: **Stopping criterion:** If $\|F(x^k)\| < tol$ then stop.

3: **Path search:** Find the smallest integer $i_k \geq 0$ such that

$$\|F(\mathfrak{p}_{x^k}(\beta^{i_k}))\| \leq (1 - \sigma\beta^{i_k})\|F(x^k)\| \tag{5.9}$$

and set $\tau_k = \beta_{i_k}$.

4: **Data update:** Set $x^{k+1} = \mathfrak{p}_{x^k}(\tau_k)$, $k = k+1$ and go to step 2.

Remark 5.4. Due to the nonlinearity in (5.8) the computation of a point along the path is very costly. For our test runs we implemented a much cheaper smooth Armijo-type line search, utilizing a path of the form

$$\mathfrak{p}_{x^k}(\tau) := x^k + \tau d^k, \tag{5.10}$$

where $d^k = (\delta y^k, \delta u^k, \delta \xi^k, \delta r^k)$ is determined by (5.7). Although there is no guaranteed descent along such a path, the globalization of this type worked sufficiently well for most problems.

Alternatively, hybrid ideas were studied, where the search direction d^k in (5.10) is replaced by the solution $\tilde{d}^k$ of the problem

$$G(\tilde{x}^k)\tilde{d}^k = -F(x^k),$$

where $\tilde{x}^k$ is a point on the line segment $[x^k, x^k + d^k]$, if the line search along (5.10) fails. Different variations of the above idea, including intermediate steps $x^{k+\frac{1}{2}}$ towards the "first" non-differentiability along the generalized Newton-direction d^k, were considered as well.

We tested all variants above and point out that in our numerical tests the Armijo-type line search worked very well. Whenever the line search, however, encountered problems because of non-differentiabilities, all path/line-search methods ran into difficulties.

Next we state Newton differentiability of the first order system of the subproblem.

Proposition 5.5. *The function $F := (F_1, F_2, F_3, F_4)$ defined in* (5.3)*, is Newton-differentiable along $\{x^k\}$.*

Proof. We note that due to optimality condition (4.8b) the optimal control u gains a posteriori regularity and is in $H_0^1(\Omega)$. Furthermore, if we initialize the algorithm by $u_0 \in H_0^1(\Omega)$, then, due to (5.7a), each update δu solves an elliptic equation with right side in $L^2(\Omega)$. Hence $u^k \in H_0^1(\Omega)$ for all $k \in \mathbb{N}$. In view of Theorem 5.3 we have $U := H_0^1(\Omega) \times H_0^1(\Omega) \times L^2(\Omega) \times \mathbb{R}$. We further note that every C^1-function is Newton-differentiable, hence F_4 is Newton-differentiable. Moreover the sum of Newton-differentiable functions, as well as the $\max(0,\cdot)$-operator in finite dimensions are Newton-differentiable. Furthermore the superposition of a Newton-differentiable mapping after a C^1-mapping is Newton-differentiable again. For a proof we refer to Proposition B.2 in Appendix B. Therefore F_3 is Newton-differentiable. The fact that for spatial dimensions $n \leq 3$ the space $H_0^1(\Omega)$ continuously embeds into $L^q(\Omega)$ for $q \leq 6$ yields Newton-differentiability of the max-term in F_1 and F_2 with image space $L^2(\Omega)$. As $L^2(\Omega)$ continuously embeds into $H^{-1}(\Omega)$ the mapping F_1 is Newton-differentiable. Hence F is Newton-differentiable. □

Note that the $L^2(\Omega)$-term $c_\mu \xi$ in the argument of the $\max(0,\cdot)$-operator in F_2 was eliminated by setting c_μ to κ. Adding the weighted $L^2(\Omega)$-norm of ξ to the cost functional of the relaxed problem therefore gives the necessary smoothing property such that F is Newton-differentiable. In view of Theorem 5.3 we have the following convergence result.

Theorem 5.6. *If $\|G(y,u,\xi,r)^{-1}\|$ is bounded in a neighborhood of a solution (y^*,u^*,ξ^*,r^*) of (5.2) then the semismooth Newton iteration*

$$(y^{k+1},u^{k+1},\xi^{k+1},r^{k+1}) = (y^k,u^k,\xi^k,r^k) + (\delta y^k,\delta u^k,\delta\xi^k,\delta r^k),$$

where $(\delta y^k,\delta u^k,\delta\xi^k,\delta r^k)$ is determined by (5.7) converges superlinearly to (y^*,u^*,ξ^*,r^*), provided that (y^0,u^0,ξ^0,r^0) is sufficiently close to (y^*,u^*,ξ^*,r^*).

The proof follows immediately from Theorem 5.3 and Proposition 5.5. We mention here that for our numerical examples reported on in the next section, the assumption of the boundedness of the inverse was always satisfied on a discrete level for various meshes. However we point out that for instance in special cases where $y^k \equiv \xi^k \equiv 0$ and $r_k - c_r\alpha > 0$ invertibility is problematic. In these cases additional stabilization is required.

5.3. **Initialization.** Due to the local convergence properties of the semismooth Newton method, its initialization becomes an issue. In our tests the following strategies worked well.

5.3.1. *The outer loop.* For the initialization of the outer loop we neglect the constraint $(y,\xi)\le\alpha$ and solve the following constrained optimal control problem using a primal-dual active set strategy (see, e.g., [24]) to obtain initial values (y^0,u^0,ξ^0):
(5.11)

$$\begin{aligned}\min \tilde{J}_{\gamma_0}(y,u,\xi) &= \frac{1}{2}\|y-y_d\|^2 + \frac{\nu}{2}\|u\|^2 + \frac{\kappa_0}{2}\|\xi\|^2 + \frac{1}{2\gamma_0}\|\max(0,\bar\lambda-\gamma_0 y)\|^2\\ &\text{over } y\in H_0^1(\Omega);\, u,\xi\in L^2(\Omega),\\ &\text{s.t. } \mathcal{A}y = u+\xi+f,\\ &\qquad \xi\ge 0 \text{ a.e. in } \Omega.\end{aligned}$$

Here $\gamma_0 > 0$ and $\kappa_0 := \gamma_0^{-\frac{1}{2}}$. We point out that the active-set-strategy employed for solving (5.11) admits a function space analysis and converges locally at a superlinear rate; see, e.g., [24]. The multiplier r is initialized by $r_0 := 0$.

5.3.2. *The inner loop.* As specified in Algorithm 1, each inner loop is initialized by the solution of its preceding outer iteration. The quality of the initialization depends on the update strategy for γ. If the regularization parameter is updated conservatively the initial values are of high quality and the semismooth Newton method requires only a small number of iterations until successful termination. Such a choice, however, results in a large number of outer iterations. By using a more aggressive update strategy for γ the number of outer iterations is kept low, typically at the cost of additional inner iterations.

6. Numerics

We consider the two-dimensional domain $\Omega = (0,1)^2$ and discretize using a uniform grid with mesh size h in each dimension. For the discretization of the Laplace-operator we use a standard five-point finite difference stencil. The test runs are based on a nested iteration technique using a grid hierarchy with mesh sizes $\{h_i\}_{i=0}^5$, where $h_i = 2^{-(i+3)}$.

In each outer iteration the relaxed-regularized problem $(\mathcal{P}_{\alpha,\gamma})$ is solved using a semismooth Newton method (see Algorithm 2) with a stopping tolerance *tol* depending on the mesh size of the current grid (e.g. $tol = 5h^2 10^{-4}$). We note that a convergence result analogous to Theorem 5.6 also holds true for a discretized version of the semismooth Newton method. If

$$\gamma \geq c_{grid}\, h^{-4}, \tag{6.1}$$

where $c_{grid} > 0$ is a constant factor, then the grid is refined by halving the mesh size. This criterion is motivated by approximation results aiming at a balance of the regularization and the discretization errors, respectively. As far as the regularization is concerned, we assume an approximation order of

$$\|y_\gamma - y^*\| \leq C \frac{1}{\sqrt{\gamma}}$$

with respect to γ. This assumption is supported by corresponding estimates for variational inequalities; see, e.g., [19]. On the other hand, we expect a discretization error of the form

$$\|y_h - y^*\| \leq C h^2,$$

where y_h is the solution of the discrete problem (see, e.g, [21]). Assuming similar behavior for our relaxed-regularized problem and the finite difference discretization, this leads to the estimate

$$\|y_{\gamma,h} - y^*\| \leq C \left(\frac{1}{\sqrt{\gamma}} + h^2 \right), \tag{6.2}$$

where $y_{\gamma,h}$ is the solution of the discrete penalized problem. For a fixed mesh size h, the discretization error dominates the approximation error of the regularization if $\gamma > h^{-4}$. Increasing the regularization parameter γ further does not improve the overall approximation error. These considerations motivate (6.1). Let us emphasize here, that the reasoning above is heuristic. A detailed error analysis is beyond the scope of this paper.

A numerical justification for (6.1), respectively (6.2), is provided in figure 6.1, which shows the L^2-errors $\|y_\gamma - y^*\|$ and $\|u_\gamma - u^*\|$ for Example 6.2 for different values of γ on different meshes in a log/log-scale. The graphic illustrates the approximation order $\mathcal{O}(\gamma^{-1/2})$ for both the state and the control. Furthermore we find that for the state variable y the discretization errors are roughly divided by 4 each time the mesh size is halved. To this end observe the convergence in the region where the curves level off. The control shows a similar behavior with a slightly smaller exponent for the order of discretization.

The $H_0^1(\Omega)$-functions y and u are prolongated using a standard nine-point bilinear interpolation with homogeneous Dirichlet boundary conditions. The L^2-function ξ is prolongated using a seven point interpolation scheme without boundary conditions. For the definition of these prolongation operators, see, e.g., [20]. With these initial guesses the relaxed-regularized problem is solved on the next finer grid. This procedure is repeated until the finest mesh size $h_5 = 2^{-8}$ is reached. The algorithm terminates if γ satisfies (6.1) for the finest mesh size.

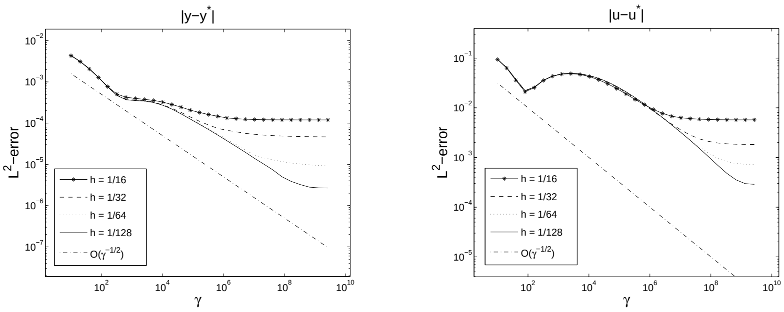

FIGURE 6.1. Approximation errors for the state y (left) and control u (right) for Example 6.2 on different meshes.

6.1. **Examples.** We present numerical test problems to illustrate our theoretical results and the numerical behavior of the new algorithm. For all examples, the regularization parameter is initialized by $\gamma_0 = 10$ and increased by a constant factor $\beta_\gamma = 2$. Further, the parameters take the values $c_r = 10$, $\sigma = 10^{-4}$, $c_{grid} = 1$, $\beta = 0.5$, $tol = 5 \cdot 10^{-4} h^2$.

Example 6.1. **Lack of strict complementarity**. We construct a test problem for which the active set at the solution contains a subset where strict complementarity does not hold, i.e., where the biactive set $B := \{y^* = 0\} \cap \{\xi^* = 0\}$ has a positive measure. This situation is challenging, as the active constraint gradients at the solution are linearly dependent. Here we consider the elliptic operator $\mathcal{A} = -\Delta$

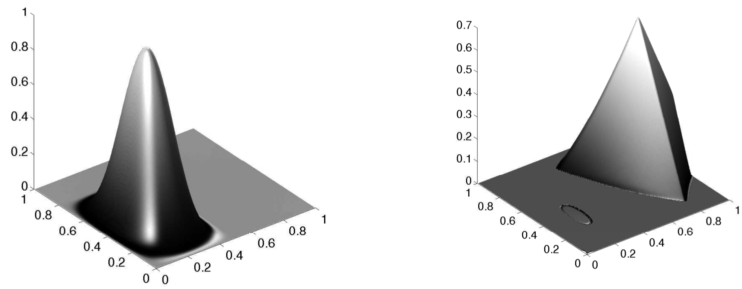

FIGURE 6.2. Optimal State y^* (left) and multiplier ξ^* (right) for Example 6.1.

and define

$$y^*(\mathbf{x}_1, \mathbf{x}_2) = \begin{cases} z_1(\mathbf{x}_1) z_2(\mathbf{x}_2) & \text{in } (0, 0.5) \times (0, 0.8), \\ 0 & \text{else,} \end{cases}$$
$$u^*(\mathbf{x}_1, \mathbf{x}_2) = y^*(\mathbf{x}_1, \mathbf{x}_2),$$
$$\xi^*(\mathbf{x}_1, \mathbf{x}_2) = 2 \max(0, -|\mathbf{x}_1 - 0.8| - |(\mathbf{x}_2 - 0.2)\mathbf{x}_1 - 0.3| + 0.35)$$

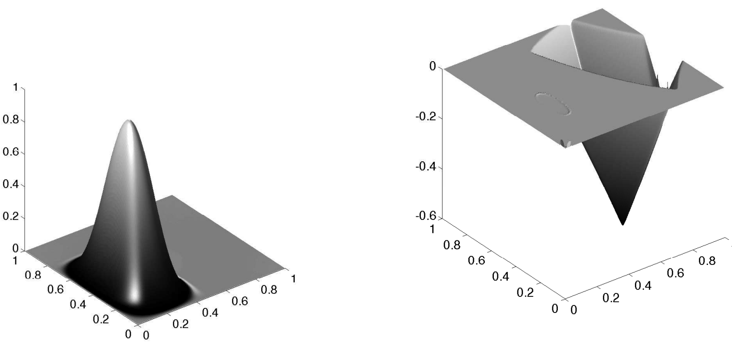

FIGURE 6.3. Optimal control u^* (left) and multiplier λ^* (right) for Example 6.1.

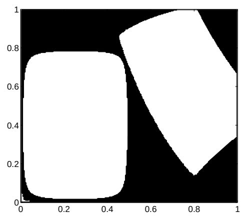

FIGURE 6.4. Biactive set (black) for Example 6.1.

with

$$z_1(\mathbf{x}_1) = -4096\mathbf{x}_1^6 + 6144\mathbf{x}_1^5 - 3072\mathbf{x}_1^4 + 512\mathbf{x}_1^3,$$
$$z_2(\mathbf{x}_2) = -244.140625\mathbf{x}_2^6 + 585.9375\mathbf{x}_2^5 - 468.75\mathbf{x}_2^4 + 125\mathbf{x}_2^3.$$

We further set

$$f = -\Delta y^* - u^* - \xi^*,$$
$$y_d = y^* + \xi^* - \nu\Delta u^*,$$

with $\nu = 1$. The optimal solutions are displayed in figures 6.2-6.4.

Example 6.2. **Degenerate solution**. For this example the optimal state y^* exhibits a very flat transition into the active set. This makes the active set detection challenging. Purely primal active set techniques usually perform poorly in such situations. Again we consider the operator $\mathcal{A} = -\Delta$, this time we set $\nu = 0.01$. The example is defined by the data

$$f(\mathbf{x}_1, \mathbf{x}_2) = y_d(\mathbf{x}_1, \mathbf{x}_2) = -|\mathbf{x}_1\mathbf{x}_2 - 0.5| + 0.25.$$

The optimal solution is shown in figures 6.5-6.6.

Example 6.3. **Elasto-plastic-torsion problem**. In this example we consider an infinitely long cylindrical bar with cross section Ω. We assume the bar to be isotropic and elastic. Starting from a zero-stress initial state, an increasing torsion moment is applied to the bar. The torsion is characterized by $c \geq 0$, which

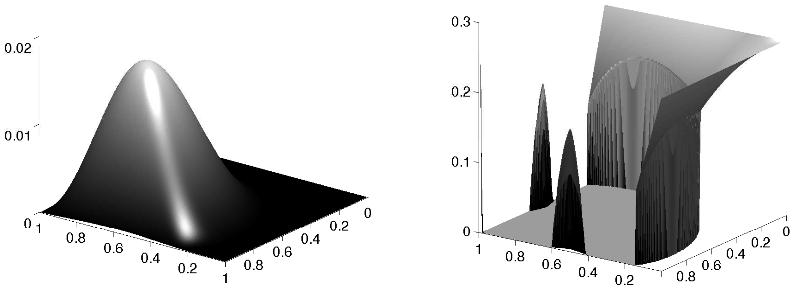

FIGURE 6.5. Optimal State y^* (left) and multiplier ξ^* (right) for Example 6.2.

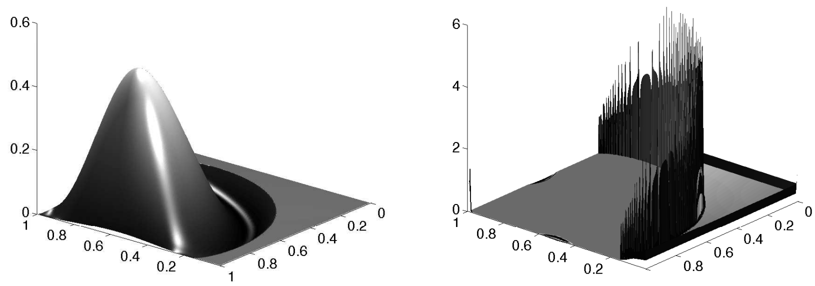

FIGURE 6.6. Optimal control u^* (left) and multiplier λ^* (right) for Example 6.2.

is defined as the torsion angle per unit length. The determination of the stress field y is equivalent to the solution of the following variational inequality (see, e.g., [18, 40]): Find $y \in K$ such that

$$\int_\Omega \nabla y \cdot \nabla(z-y) \geq c \int_\Omega (z-y) \; \forall z \in K, \tag{6.3}$$

where the cone K is defined by

$$K = \{z \in H_0^1(\Omega) : |\nabla z| \leq 1 \text{ a.e. in } \Omega\}.$$

In [11], Brezis and Sibony show, that if $\Omega \subset \mathbb{R}^2$ is bounded and has a smooth boundary Γ, then the variational inequality problem (6.3) is equivalent to finding $y \in \hat{K}$ such that

$$\int_\Omega \nabla y \cdot \nabla(z-y)\, dx \leq c \int_\Omega (z-y)\, dx \; \forall z \in \hat{K}, \tag{6.4}$$

with $\hat{K} = \{z \in H_0^1(\Omega) : |z(x)| \leq d(x,\Gamma) \text{ a.e. in } \Omega\}$. Using slack variables (non-negative Lagrange multipliers), the variational inequality (6.4) can be equivalently

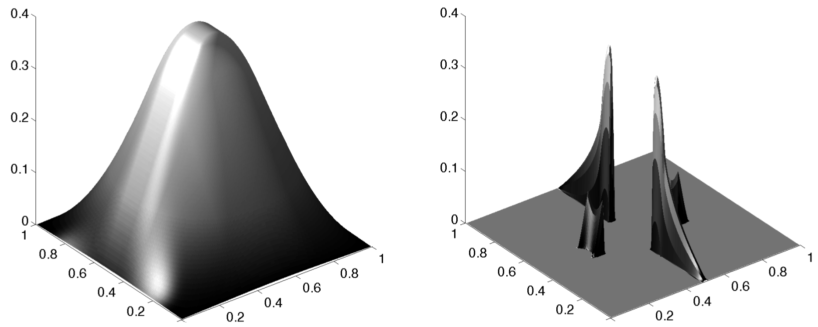

FIGURE 6.7. Optimal State y^* (left) and multiplier ξ_u^* (right) for Example 6.3.

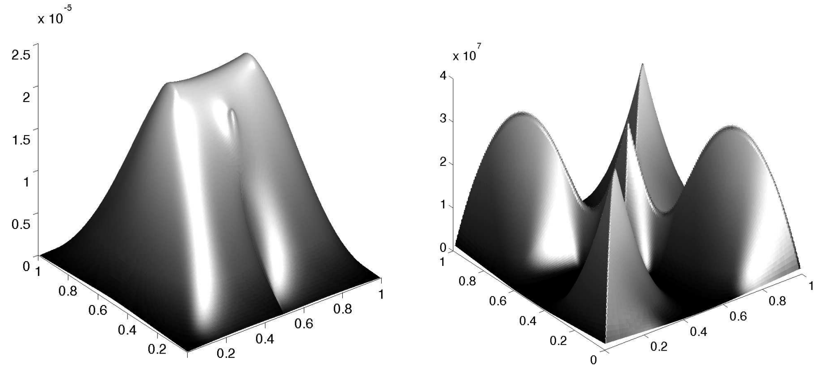

FIGURE 6.8. Optimal control u^* (left) and multiplier λ_u^* (right) for Example 6.3.

written as

$$
\begin{aligned}
-\Delta y - c\mathbf{1} &= \xi_u - \xi_l, \\
y + d \geq 0,\ \xi_l \geq 0,\ (y+d, \xi_l) &= 0, \\
d - y \geq 0,\ \xi_u \geq 0,\ (d-y, \xi_u) &= 0,
\end{aligned} \tag{6.5}
$$

with $d := d(\cdot, \Gamma) \in C(\Omega)$ and $\mathbf{1}$ being the constant function with value 1. Note that the unilateral constraints on the state variable are extended to bilateral ones. Therefore an extra multiplier ξ_u is introduced. Here ξ_l represents the Lagrange multiplier corresponding to the lower bound $-d \leq y$, whereas ξ_u corresponds to the upper bound $y \leq d$. This variational inequality can be treated as the state system in our MPCC using $c\mathbf{1}$ as the control u. In this case the function space for the control would be limited to the space of constant functions. Here we generalize this setting by regarding an L^2-control u and introducing a fixed data term $f \in L^2(\Omega)$.

Consequently we consider the optimal control problem

$$(\mathcal{P})\qquad \begin{array}{rl} \min & J(y,u) := \frac{1}{2}\|y-y_d\|^2_{L^2} + \frac{\nu}{2}\|u\|^2_{L^2} \\ \text{over} & y \in H^1_0(\Omega);\ u,\ \xi_l,\ \xi_u \in L^2(\Omega), \\ \text{s.t.} & -\Delta y = u - \xi_l + \xi_u + f, \\ & y + d \geq 0,\ \xi_l \geq 0,\ (y+d,\xi_l) = 0, \\ & d - y \geq 0,\ \xi_u \geq 0,\ (d-y,\xi_u) = 0. \end{array}$$

Note that similar to the additional multiplier in the lower level problem, the upper level problem also gives rise to an additional multiplier λ_u, where λ_l, $\lambda_u \in H^{-1}(\Omega)$ correspond to the respective constraints on y. We further consider the relaxed-regularized problem

$$(\mathcal{P}_{\alpha,\gamma})\qquad \begin{array}{rl} \min & J_\gamma(y,u,\xi_l,\xi_u) := J(y,u) + \frac{\kappa}{2}\left(\|\xi_l\|^2_{L^2} + \|\xi_u\|^2_{L^2}\right) \\ & +\frac{1}{2\gamma}\left(\|\max(0,\bar{\lambda}-\gamma(y+d))\|^2_{L^2} + \|\max(0,\bar{\lambda}-\gamma(d-y))\|^2_{L^2}\right) \\ \text{over} & y \in H^1_0(\Omega);\ u,\ \xi_l,\ \xi_u \in L^2(\Omega), \\ \text{s.t.} & -\Delta y = u - \xi_l + \xi_u + f, \\ & \xi_l \geq 0,\ \xi_u \geq 0, \\ & (y+d,\xi_l) \leq \alpha,\ (d-y,\xi_u) \leq \alpha. \end{array}$$

On $\Omega = (0,1)\times(0,1)$ we use the following data:

$$y_d = z_1 z_2,\ f = -\Delta y_d - z_3,$$

where

$$z_1(\mathbf{x}_1,\mathbf{x}_2) = \begin{cases} -2.5\mathbf{x}_1^2 + 2.5\mathbf{x}_1 - 0.225 & \text{in } (0.3,0.7)\times(0,1), \\ 0.5 - |\mathbf{x}_1 - 0.5| & \text{else}, \end{cases}$$

$$z_2(\mathbf{x}_1,\mathbf{x}_2) = \begin{cases} -12.5\mathbf{x}_2^2 + 11.25\mathbf{x}_2 - 1.53125 & \text{in } (0,1)\times(0.35,0.45), \\ 1 & \text{in } (0,1)\times[0.45,0.55], \\ -12.5\mathbf{x}_2^2 + 13.75\mathbf{x}_2 - 2.78125 & \text{in } (0,1)\times(0.55,0.65), \\ 1.25 - 2.5|\mathbf{x}_2 - 0.5| & \text{else}, \end{cases}$$

$$z_3(\mathbf{x}_1,\mathbf{x}_2) = \begin{cases} \mathbf{x}_1(1-\mathbf{x}_1) & \text{in } ((0,0.3)\cup(0.7,1))\times[0.45,0.55], \\ 0 & \text{else}, \end{cases}$$

and $\nu = 1$. The solution is inactive with respect to the lower bound, therefore $\lambda_l^* = 0$. Furthermore we find that $\xi_l^* = 0$. The optimal values for y^*, ξ_u^*, u^* and λ_u^* are displayed in figures 6.7-6.8.

Example 6.4. **Non-symmetric operator**. In this example we consider the non-symmetric operator

$$\mathcal{A} := -\Delta + b^T\nabla \tag{6.6}$$

with $b \in \mathbb{R}^n$. The resulting MPCC possesses no bilevel structure, i.e. the variational inequality can no longer be interpreted as the optimality conditions of a lower-level minimization problem. The state equation then reads

$$-\Delta y + b^T\nabla y - u - \xi = f$$

and the adjoint equation is

$$y - \max(0,\bar{\lambda}-\gamma y) - \Delta p - b^T\nabla p + r\xi = y_d.$$

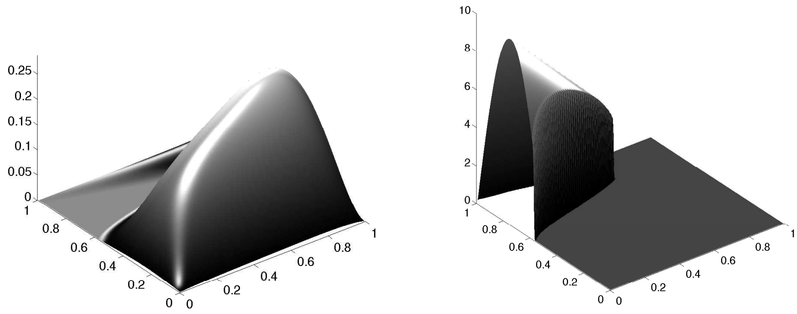

FIGURE 6.9. Optimal State y^* (left) and multiplier ξ^* (right) for Example 6.4.

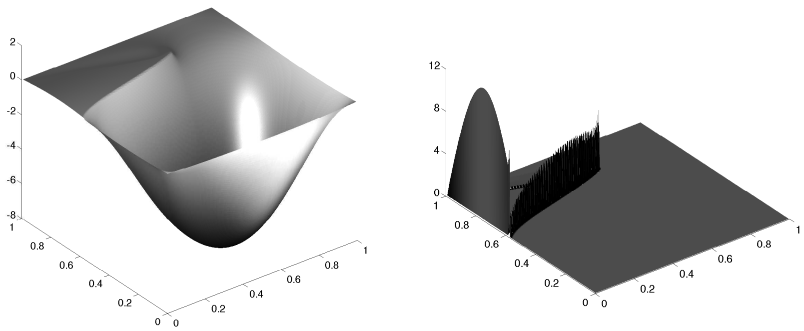

FIGURE 6.10. Optimal control u^* (left) and multiplier λ^* (right) for Example 6.4.

Note that $-\Delta + b^T\nabla : H_0^1(\Omega) \to H^{-1}(\Omega)$ as well as its adjoint operator $-\Delta - b^T\nabla : H_0^1(\Omega) \to H^{-1}(\Omega)$ are bounded and coercive. We specify the data for the test problem as

$$
\begin{aligned}
b &= (0.7, -0.7)^T, \\
f(\mathbf{x}_1, \mathbf{x}_2) &= 10\left(\sin(2\pi\mathbf{x}_2) + \mathbf{x}_1\right), \\
y_d(\mathbf{x}_1, \mathbf{x}_2) &= \min\left(\mathbf{x}_1(1-\mathbf{x}_1)\mathbf{x}_2(1-\mathbf{x}_2), 0.04\mathbf{x}_1 + 0.01\mathbf{x}_2\right)
\end{aligned}
$$

and $\nu = 0.001$. The optimal solutions are presented in figures 6.9-6.10. As in Example 6.2 the multiplier λ^* exhibits low regularity.

6.2. **Results.** Next we discuss some results obtained by our algorithm with Armijo-type line search.

Nested Grids. In the beginning of this section a nested iteration technique is proposed for Algorithm 1. Table 6.1 displays the number of iterations on the different grids for the various examples. It shows that most of the iterations are spent on the coarse meshes. This is a clear indication of the efficiency of nested grids when solving MPCCs in function space.

TABLE 6.1. Iterations on the different grids for Examples 6.1 - 6.4.

h	iter. pr. 6.1	iter. pr. 6.2	iter. pr. 6.3	iter. pr. 6.4
1/16	30	27	9	22
1/32	12	12	8	12
1/64	11	12	19	28
1/128	10	15	66	30
1/256	9	24	35	28
total	72	90	137	120

Rate of convergence. Tables 6.2- 6.3 display the convergence factors

$$\rho_k = \|y_\gamma^{k+1} - y_\gamma^*\|_{H_0^1} / \|y_\gamma^k - y_\gamma^*\|_{H_0^1}$$

of the state variable y in the H^1-norm over the iterates of the semismooth Newton algorithm for different values of γ. The problems were solved on a fixed grid with $h = 1/128$ up to a precision of tol = 1e-8. The exact solution y_γ^* was approximated by solving the corresponding problem to high accuracy (tol = 1e-12). When

TABLE 6.2. convergence factors for Example 6.1.

γ	1e3	1e4	1e5	1e6	1e7
ρ_k	2.1573e-2	4.4393e-1	9.3638e-1	9.5380e-1	9.8513e-1
	1.4600e-3	3.0727e-1	9.3873e-1	9.6137e-1	9.8870e-1
		1.6024e-1	...	...	...
		3.4093e-2	4.4082e-1	8.9863e-1	8.1664e-1
		1.0170e-2	2.8161e-1	1.1190e-1	7.6285e-1
			1.2539e-1	4.8308e-1	5.0669e-1
			2.1609e-2	3.3476e-1	1.1434e-5

TABLE 6.3. convergence factors for Example 6.2.

γ	1e3	1e4	1e5	1e6	1e7
ρ_k	1.7029e-1	5.3379e-1	9.3905e-1	9.5892e-1	9.9212e-1
	4.6349e-2	4.2591e-1	9.3019e-1	9.5761e-1	9.8332e-1
	4.3590e-4	2.9001e-1	...	...	...
		1.3731e-1	2.4973e-1	3.1833e-1	2.7615e-1
		2.2167e-2	1.9983e-1	1.5442e-1	8.9317e-1
			5.3870e-2	6.6321e-2	4.6193e-2
			1.4475e-2	7.5521e-6	2.0802e-6

observing the columns of tables 6.2 - 6.3 the local superlinear convergence as stated in Theorem 5.6 can be verified numerically. With increasing values of γ, as one would expect, the convergence radius of the semismooth Newton method becomes smaller. This is reflected in the increasing number of iterations as γ is increased. On the other hand we note that our nested iterations concept intertwined with a suitable γ-update strategy exhibits a rather stable convergence.

Stationarity. Lemma 4.9 gives conditions for the accumulation point of the algorithm to be almost C-stationary. As argued in (4.35) these conditions are satisfied if the solutions of the relaxed-regularized problems exhibit an approximation property of the quality

$$\|y_\gamma - y^*\| = \mathcal{O}(\gamma^{-1/2}). \tag{6.7}$$

This approximation order was verified for all of our numerical examples and it is shown exemplarily for Example 6.2 in figure 6.1. Although a rigorous error analysis is of interest, it is beyond the scope of this paper. In our numerical examples we could typically observe even strong stationarity.

Appendix A

In Corollary 4.5 we derived the first order optimality system of $(\mathcal{P}_{\alpha,\gamma})$ using results due to Zowe and Kurcyusz [44]. In this Appendix we briefly recall the main result of [44] and give the proof of Corollary 4.5. For this purpose consider the general mathematical programming problem

$$\min_{x\in X} F(x) \quad \text{s.t. } x \in C,\ g(x) \in K, \tag{A.1}$$

where F is a differentiable real functional defined on a real Banach space X, C is a non-empty closed convex subset of X, g is a continuously differentiable map from X into a real Banach space Y and K is a closed convex cone in Y with vertex at the origin. For fixed $x \in X$ and $y \in Y$ let $C(x)$ and $K(y)$ denote the conical hull of $C - \{x\}$ and $K - \{y\}$, respectively, i.e.,

$$C(x) := \{\lambda(c - x) : c \in C,\ \lambda \geq 0\},$$
$$K(y) := \{k - \lambda y : k \in K,\ \lambda \geq 0\}.$$

The quantity $y^* \in Y^*$ is called a Lagrange multiplier for problem (A.1) at an optimal point $\bar{x} \in X$, if

$$\begin{aligned} &(i) && y^* \in K^+,\\ &(ii) && \langle y^*, g(\bar{x})\rangle_{Y^*,Y} = 0,\\ &(iii) && F'(\bar{x}) - y^*(g'(\bar{x})) \in C(\bar{x})^+, \end{aligned}$$

where X^* and Y^* denote the topological duals of X and Y and for each subset M of X (or Y respectively), M^+ denotes its polar cone

$$M^+ := \{x^* \in X^* : \langle x^*, \mu\rangle_{X^*,X} \geq 0 \quad \text{for all } \mu \in M\}.$$

For an optimal point $\bar{x} \in X$ let $\Lambda(\bar{x})$ denote the set of Lagrange multipliers for problem (A.1) at $\bar{x}$. The main result in [44] is as follows:

Theorem A.1. *Let $\bar{x}$ be an optimal solution for problem* (A.1). *If $g'(\bar{x})C(\bar{x}) - K(g(\bar{x})) = Y$, then the set $\Lambda(\bar{x})$ of Lagrange multipliers for problem* (A.1) *at $\bar{x}$ is non-empty and bounded.*

In order to apply Theorem A.1 to problem $(\mathcal{P}_{\alpha,\gamma})$, we set

$$\begin{aligned} X := C &:= H_0^1(\Omega) \times L^2(\Omega) \times L^2(\Omega), \\ x &:= (y, u, \xi), \\ F(x) &:= \tilde{J}_\gamma(y, u, \xi), \\ Y &:= H^{-1}(\Omega) \times \mathbb{R} \times L^2(\Omega), \\ K &:= \{0\} \times \mathbb{R}_+ \times L^2(\Omega)_+ \ \text{ and} \\ g(x) &:= (-\mathcal{A}y + u + \xi + f, -(y, \xi) + \alpha, \xi), \end{aligned}$$

where $\mathbb{R}_+ = [0, \infty)$ and $L^2(\Omega)_+ = \{v \in L^2(\Omega) : v \geq 0 \text{ a.e. in } \Omega\}$. With this notation, problem (A.1) is equivalent to problem $(\mathcal{P}_{\alpha,\gamma})$. Now let $\bar{x} := (\bar{y}, \bar{u}, \bar{\xi})$ be an optimal solution for problem $(\mathcal{P}_{\alpha,\gamma})$. The constraint qualification $g'(\bar{x})C(\bar{x}) - K(g(\bar{x})) = Y$ in Theorem A.1 is then equivalent to the requirement that for every $(y_1, y_2, y_3) \in Y = H^{-1}(\Omega) \times \mathbb{R} \times L^2(\Omega)$ there must exist $(c_1, c_2, c_3) \in C(\bar{y}, \bar{u}, \bar{\xi})$, $\lambda \geq 0$ and $(k_1, k_2) \in \mathbb{R}_+ \times L^2(\Omega)_+$ such that

$$\begin{aligned} -\mathcal{A}c_1 + c_2 + c_3 &= y_1, \\ -(c_1, \bar{\xi}) - (\bar{y}, c_3) - (k_1 - \lambda(-(\bar{y}, \bar{\xi}) + \alpha)) &= y_2, \\ c_3 - (k_2 - \lambda\bar{\xi}) &= y_3. \end{aligned} \tag{A.2}$$

In our case the subspace C is the space X itself. Therefore we have $C(x) = X$ for every $x \in X$. Now let (y_1, y_2, y_3) be an arbitrary element of Y. If $\|\bar{\xi}\| > 0$, we set $\lambda := 1$, $k_2 := \bar{\xi}$ and $c_3 := y_3$ to satisfy the third equation. We further set $k_1 := \alpha - (\bar{y}, \bar{\xi})$ and define $\phi \in H_0^1(\Omega)$ such that $-\mathcal{A}\phi = y_1 \in H^{-1}(\Omega)$. As $\bar{\xi} \neq 0$ we can find an element $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$ such that $(\psi, \bar{\xi}) = -(\bar{y}, c_3) - y_2 - (\phi, \bar{\xi})$. If we set $c_1 := \phi + \psi$ and $c_2 := -c_3 + \mathcal{A}\psi \in L^2(\Omega)$, the system (A.2) is satisfied.

If $\bar{\xi} = 0$ a.e. in Ω, we choose $c_1 \in H_0^1(\Omega)$ so that $-\mathcal{A}c_1 = y_1$. We set $c_3 := y_3$, $c_2 := -c_3$, $\lambda := \frac{1}{\alpha} \max(0, (\bar{y}, c_3) + y_2)$, $k_1 := \max(0, -(\bar{y}, c_3) - y_2)$ and $k_2 := 0$. With these choices, the system (A.2) is again satisfied. Theorem A.1 then guarantees the existence of Lagrange multipliers $(p, r, \mu) \in Y^* = H_0^1(\Omega) \times \mathbb{R} \times L^2(\Omega)$ which satisfy

$$r \geq 0,\ \mu \in L^2(\Omega)_+, \tag{A.3a}$$

$$r\left((\bar{y}, \bar{\xi}) - \alpha\right) = 0,\ (\mu, \bar{\xi}) = 0, \tag{A.3b}$$

$$\begin{pmatrix} \bar{y} - y_d - \max(0, \bar{\lambda} - \gamma\bar{y}) + \mathcal{A}^*p + r\bar{\xi} \\ \nu\bar{u} - p \\ \kappa\bar{\xi} - p + r\bar{y} - \mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{A.3c}$$

It is easily shown that system (A.3), together with the constraint $g(x) \in K$, is equivalent to system (4.8) in Corollary 4.5, which completes the proof.

Appendix B

In this appendix we state a few auxiliary results. For $v \in H_0^1(\Omega)$ let v^+ denote the pointwise non-negative part of v, i.e., $v^+ := \max(0, v)$. The following lemma investigates the convergence behavior of y_γ^+.

Lemma B.1. *Let $\{v_\gamma\} \subset H_0^1(\Omega)$ be a sequence converging to some $\tilde{v}$ in $H_0^1(\Omega)$. Then also $v_\gamma^+ \to \tilde{v}^+$ strongly in $H_0^1(\Omega)$.*

Proof. Note that the strong L^2-convergence of $v_\gamma \to \tilde{v}$, together with the Lipschitz-property of the operator $\max(\cdot, 0) : L^2(\Omega) \to L^2(\Omega)$ yields strong convergence of v_γ^+ to $\tilde{v}^+$ in $L^2(\Omega)$. Furthermore for every $\gamma > 0$

$$\|\nabla v_\gamma^+\|_{L^2(\Omega)^n} \leq \|\nabla v_\gamma\|_{L^2(\Omega)^n} \leq C,$$

due to the strong convergence of $\{v_\gamma\}$ in $H_0^1(\Omega)$. Hence there exists $\beta = (\beta_1, \ldots, \beta_n) \in L^2(\Omega)^n$ such that on a suitable subsequence $\frac{\partial v_\gamma^+}{\partial x_i} \rightharpoonup \beta_i$ for all $1 \leq i \leq n$. We verify that indeed $\beta = \nabla \tilde{v}^+$ by multiplying by a test function $\varphi \in C_0^\infty(\Omega)$. Then for every $1 \leq i \leq n$ we find that

$$(\beta_i, \varphi) = \lim_{\gamma \to \infty} (\frac{\partial v_\gamma^+}{\partial x_i}, \varphi) = -\lim_{\gamma \to \infty} (v_\gamma^+, \frac{\partial \varphi}{\partial x_i}) = -(\tilde{v}^+, \frac{\partial \varphi}{\partial x_i}) = (\frac{\partial \tilde{v}^+}{\partial x_i}, \varphi).$$

As the weak limit of the subsequence is uniquely determined we find that in fact $\nabla v_\gamma^+ \rightharpoonup \nabla \tilde{v}^+$ on the whole sequence. Finally we show strong convergence of $\{\nabla v_\gamma^+\}$ by considering

$$\begin{aligned} \|\nabla v_\gamma^+ - \nabla \tilde{v}^+\|_{L^2(\Omega)^n}^2 &= (\nabla v_\gamma^+, \nabla v_\gamma^+) - 2(\nabla v_\gamma^+, \nabla \tilde{v}^+) + (\nabla \tilde{v}^+, \nabla \tilde{v}^+) \\ &= (\nabla v_\gamma^+, \nabla v_\gamma) - 2(\nabla v_\gamma^+, \nabla \tilde{v}^+) + (\nabla \tilde{v}^+, \nabla \tilde{v}^+) \to 0. \end{aligned}$$

□

In view of the requirements in connection with the optimality system defined by the function F in (5.3) we provide the following chain rule for semismooth and Fréchet differentiable functions.

Proposition B.2. *Let $F_1 : D \subset X \to Z$ be Newton-differentiable in the open subset $U \subset D$ with a generalized derivative G_1 such that $\{\|G_1(v)\|_{\mathcal{L}(X,Z)} : v \in U\}$ is bounded. Further let $F_2 : Y \to X$ be continuously Fréchet differentiable in $F_2^{-1}(U)$ with the derivative F_2'. Then $H := F_1 \circ F_2$ is Newton-differentiable with a generalized derivative $G := G_1(F_2)F_2' \in \mathcal{L}(Y, Z)$.*

Proof. The Newton-differentiability of F_1 in U implies that for all $u \in U$

$$F_1(u + h) - F_1(u) = G_1(u + h)h + \|h\|_X a(h) \tag{B.1}$$

with $a(h) \in Z$ such that $\|a(h)\|_Z \to 0$ as $\|h\|_X \to 0$. Similarly, the Fréchet differentiability of F_2 implies that for every $v \in F_2^{-1}(U)$

$$F_2(v + k) - F_2(v) = F_2'(v)k + \|k\|_Y b(k) \tag{B.2}$$

where $b(k) \in X$ such that $\|b(k)\|_X \to 0$ as $\|k\|_Y \to 0$. If we set $u := F_2(v)$ and $h := F_2(v + k) - F_2(v)$ in (B.1), then we obtain

$$F_1(F_2(v + k)) - F_1(F_2(v)) = G_1(F_2(v + k))F_2'(v + k)k + c(k)$$

with

$$\begin{aligned} c(k) = {} & G_1(F_2(v + k))\,(F_2'(v)k - F_2'(v + k)k) + \\ & \|k\| G_1(F_2(v + k))b(k) + \|F_2(v + k) - F_2(v)\| a(F_2(v + k) - F_2(v)). \end{aligned}$$

Dividing by the norm of k, we estimate

$$
\begin{aligned}
\frac{\|c(k)\|_Z}{\|k\|_Y} &\leq \|G_1(F_2(v+k))\|_{\mathcal{L}(X,Z)}\|F_2'(v)-F_2'(v+k)\|_{\mathcal{L}(Y,X)} \\
&\quad + \|G_1(F_2(v+k))\|_{\mathcal{L}(X,Z)}\|b(k)\|_X \\
&\quad + \frac{\|F_2(v+k)-F_2(v)\|_X}{\|k\|_Y}\|a(F_2(v+k)-F_2(v))\|_Z.
\end{aligned} \tag{B.3}
$$

Due to the boundedness of $\|G_1\|$ and the continuity of F_2', the expression on the right side of (B.3) tends to zero as $\|k\| \to 0$. This proves the assertion of the proposition. □

Acknowledgment

The authors would like to thank Daniel Ralph for the discussions and his input from which this paper has benefited.

References

[1] R. A. Adams, *Sobolev Spaces*, Pure and Applied Mathematics, Vol. 65, Academic Press, New York-London, 1975.

[2] H. W. Alt, *Lineare Funktionalanalysis*, 5th edition, Springer-Verlag, Berlin, Heidelberg, New York, 2006.

[3] V. Barbu, *Optimal Control of Variational Inequalities*, Research Notes in Mathematics, 100, Pitman Publishing, London, 1984.

[4] V. Barbu, P. Neittaanmäki and A. Niemistö, *Approximating optimal control problems governed by variational inequalities*, Numer. Funct. Anal. Optim., 15 (1994), pp. 489–502.

[5] M. Bergounioux, *Optimal control of an obstacle problem*, Appl. Math. Optim., 36 (1997), pp. 147–172.

[6] M. Bergounioux, *Optimal control of problems governed by abstract elliptic variational inequalities with state constraints*, SIAM J. Control Optim., 36 (1998), pp. 273–289.

[7] M. Bergounioux and H. Dietrich, *Optimal control of problems governed by obstacle type variational inequalities: a dual regularization-penalization approach*, J. Convex Anal., Vol. 5 (1998), No. 2, pp. 329–351.

[8] M. Bergounioux, K. Ito and K. Kunisch, *Primal-dual strategy for constrained optimal control problems*, SIAM J. Control Optim., 37 (1999), pp. 1176–1194.

[9] A. Bermudez and C. Saguez, *Optimal control of variational inequalities*, Control and Cybernetics, 14 (1985), pp. 9–30.

[10] H. Brézis, *Analyse fonctionnelle, Théorie et applications*, Dunod, Paris, 2005.

[11] H. Brézis and M. Sibony, *Équivalence de deux inéquations variationnelles et applications*, Arch. Rational Mech. Anal., 41 (1971), pp. 254–265.

[12] E. Casas, *Control of an elliptic problem with pointwise state constraints*, SIAM J. Control Optim., 24 (1986), pp. 1309–1318.

[13] X. Chen, Z. Nashed and L. Qi, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal., 38 (2000), pp. 1200–1216.

[14] F. Facchinei and J.-S. Pang, *Finite-dimensional variational inequalities and complementarity problems, Vol. II*, Springer Series in Operations Research, Springer-Verlag, New York, 2003.

[15] M. C. Ferris and T. S. Munson, *Interfaces to PATH 3.0: design, implementation and usage*, Comput. Optim. Appl., 12 (1999), pp. 207–227.

[16] A. Friedman, *Variational principles and free-boundary problems*, Pure and Applied Mathematics, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1982.

[17] A. Friedman, *Optimal control for variational inequalities*, SIAM J. Control Optim., 24 (1986), pp. 439–451.

[18] R. Glowinski, *Numerical methods for nonlinear variational problems*, Springer Series in Computational Physics, Springer-Verlag, New York, 1984.

[19] C. Großmann and H.-G. Roos, *Numerische Behandlung partieller Differentialgleichungen*, Teubner Studienbücher Mathematik (Teubner Mathematical Textbooks), third edition, B. G. Teubner, Stuttgart, 2005.
[20] W. Hackbusch, *Multigrid methods and applications*, Springer Series in Computational Mathematics, Vol. 4, Springer-Verlag, Berlin, 1985.
[21] W. Hackbusch, *Theorie und Numerik elliptischer Differentialgleichungen*, Teubner Studienbücher Mathematik, Stuttgart, 1996.
[22] A. Henrot and M. Pierre, *Variation et optimisation de formes. Une analyse géométrique*, Mathématiques et Applications, Vol. 48, Springer-Verlag, New York, 2005.
[23] M. Hintermüller, *Inverse coefficient problems for variational inequalities: optimality conditions and numerical realization*, M2AN Math. Model. Numer. Anal., 35 (2001), pp. 129–152.
[24] M. Hintermüller, K. Ito and K. Kunisch, *The primal-dual active set strategy as a semismooth Newton method*, SIAM J. Optim., 13 (2003), pp. 865–888.
[25] M. Hintermüller and K. Kunisch, *Path-following methods for a class of constrained minimization problems in function space*, SIAM J. Optim., 17 (2006), pp. 159–187.
[26] M. Hintermüller and K. Kunisch, *Feasible and noninterior path-following in constrained minimization with low multiplier regularity*, SIAM J. Control Optim., 45 (2006), pp. 1198–1221.
[27] K.-H. Hoffman and M. Kubo and N. Yamazaki, *Optimal control problems for elliptic-parabolic variational inequalities with time-dependent constraints*, Numer. Funct. Anal. Optim., 27 (2006), pp. 329–356.
[28] J. Jarušek and J. V. Outrata, *On sharp necessary optimality conditions in control of contact problems with strings*, Nonlinear Analysis, 67 (2007), pp. 1117–1128.
[29] A. M. Khludnev, *An optimal control problem for a fourth-order variational inequality*, Partial differential equations, Part 1, 2 (Warsaw, 1990), Banach Center Publ., 27, Part 1, vol. 2, pp. 225–231, 1992.
[30] D. Kinderlehrer and G. Stampacchia, *An introduction to variational inequalities and their applications*, Reprint of the 1980 original Classics in Applied Mathematics, 31. SIAM, Philadelphia, 2000.
[31] H. Lou, *An optimal control problem governed by quasi-linear variational inequalities*, SIAM J. Control Optim., 41 (2002), pp. 1229–1253.
[32] Z. Luo, J.-S. Pang and D. Ralph, *Mathematical Programs with Equilibrum Constraints*, Cambridge University Press, Cambridge, United Kingdom, 1996.
[33] R. Mifflin, *Semismooth and semiconvex functions in constrained optimization*, SIAM J. Control Optim., 15 (1977), pp. 959–972.
[34] F. Mignot and J.-P. Puel, *Optimal control in some variational inequalities*, SIAM J. Control Optim., 22 (1984), pp. 466–476.
[35] B. S. Mordukhovich, *Variational analysis and generalized differentiation, vol. I*, Springer-Verlag, New York, 2006.
[36] B. S. Mordukhovich, *Variational analysis and generalized differentiation, vol. II*, Springer-Verlag, New York, 2006.
[37] J. Outrata, M. Kočvara and J. Zowe, *Nonsmooth approach to optimization problems with equilibrium constraints*, Nonconvex Optimization and its Applications, 28, Kluwer Academic Publishers, Dordrecht, 1998.
[38] L. Qi and J. Sun, *A nonsmooth version of Newton's method*, Math. Programming, 58 (1993), pp. 353–367.
[39] D. Ralph, *Global convergence of damped Newton's method for nonsmooth equations via the path search*, Math. Oper. Res., 19 (1994), pp. 352–389.
[40] S. I. Repin, *Estimates of deviations from exact solutions of elliptic variational inequalities*, Journal of Mathematical Sciences, 115 (2003), pp. 2812–2819.
[41] H. Scheel and S. Scholtes, *Mathematical programs with complementarity constraints: stationarity, optimality, and sensitivity*, Math. Oper. Res., Vol. 25, No. 1 (2000), pp. 1–22.
[42] S. Scholtes, *Convergence properties of a regularization scheme for mathematical programs with complementarity constraints*, SIAM J. Optim., 11 (2001), pp. 918–936.
[43] J. J. Ye, *Necessary and sufficient optimality conditions for mathematical programs with equilibrium constraints*, J. Math. Anal. Appl., 307 (2005), pp. 350–369.
[44] J. Zowe and S. Kurcyusz, *Regularity and stability for the mathematical programming problem in Banach spaces*, Appl. Math. Optim., 5 (1979), pp. 49–62.

M. Hintermüller, Humboldt-University of Berlin, Department of Mathematics, Unter den Linden 6, D-10099 Berlin, Germany, and Karl-Franzens-University of Graz, Department of Mathematics and Scientific Computing, Heinrichstrasse 36, A-8010 Graz, Austria

E-mail address: michael.hintermueller@uni-graz.at

I. Kopacka, Karl-Franzens-University of Graz, Department of Mathematics and Scientific Computing, Heinrichstrasse 36, A-8010 Graz, Austria

E-mail address: ian.kopacka@uni-graz.at