Control and Cybernetics

vol. **39** (2010) No. 3

Local quadratic convergence of SQP for elliptic optimal control problems with mixed control-state constraints*

by

Roland Griesse[1], Nataliya Metla[2] and Arnd Rösch[3]

[1]TU Chemnitz, Faculty of Mathematics
D–09107 Chemnitz, Germany

[2]Johann Radon Institute for Computational and Applied Mathematics, (RICAM), Austrian Academy of Sciences
Altenbergerstraße 69, A–4040 Linz, Austria

[3]Universität Duisburg-Essen, Fachbereich Mathematik
Forsthausweg 2, D–47057 Duisburg, Germany

e-mail: roland.griesse@mathematik.tu-chemnitz.de, nataliya.metla@oeaw.ac.at, arnd.roesch@uni-due.de

Abstract: Semilinear elliptic optimal control problems with pointwise control and mixed control-state constraints are considered. Necessary and sufficient optimality conditions are given. The equivalence of the SQP method and Newton's method for a generalized equation is discussed. Local quadratic convergence of the SQP method is proved.

Keywords: optimal control, sequential quadratic programming, mixed control-state constraints, implicit function theorem, generalized equation.

1. Introduction

This paper is concerned with the local convergence analysis of the sequential quadratic programming (SQP) method for the following class of semilinear optimal control problems:

$$\text{Minimize } f(y,u) := \int_\Omega \phi(\xi, y(\xi), u(\xi))\, d\xi \qquad (\mathbf{P})$$

subject to $u \in L^\infty(\Omega)$ and the elliptic state equation

$$\begin{aligned} A\,y + d(\xi, y) &= u \quad \text{in } \Omega,\\ y &= 0 \quad \text{on } \partial\Omega, \end{aligned} \qquad (1.1)$$

*Submitted: April 2008; Accepted: July 2009.

as well as pointwise constraints

$$\begin{aligned} u \geqslant 0 \quad &\text{in } \Omega, \\ \varepsilon\, u + y \geqslant y_c \quad &\text{in } \Omega. \end{aligned} \tag{1.2}$$

Here and throughout, ξ denotes points in the bounded domain $\Omega \subset \mathbb{R}^N$, $N \in \{2,3\}$, which is convex or has a $C^{1,1}$ boundary $\partial\Omega$. In (1.1), A is an elliptic operator in $H_0^1(\Omega)$ specified below, and ε is a positive number. The bound y_c is a function in $L^\infty(\Omega)$.

Problems with mixed control-state constraints are important as Lavrientiev-type regularizations of pointwise state-constrained problems (Meyer, Prüfert and Tröltzsch, 2007; Meyer, Rösch and Tröltzsch, 2005; Meyer and Tröltzsch, 2006), but they are also interesting in their own right. Note that in addition to the mixed control-state constraint, a pure control constraint is present on the same domain. Since problem (**P**) is nonconvex, different local minima may occur.

SQP methods have proven to be fast solution methods for nonlinear programming problems. A large body of literature exists concerning the analysis of these methods for finite-dimensional problems. For a convergence analysis in a general Banach space setting with equality and inequality constraints, we refer to Alt (1990, 1994).

The main contribution of this paper is the proof of local quadratic convergence of the SQP method, applied to (**P**). To our knowledge, such convergence results in the context of PDE-constrained optimization are so far only available for purely control-constrained problems (Alt, Sontag and Tröltzsch, 1996; Heinkenschloss, 1998; Heinkenschloss and Tröltzsch, 1998; Tröltzsch, 1994, 1999; Tröltzsch and Volkwein, 2001). In the context of ordinary differential equations, the SQP method has been analyzed for instance in Alt and Malanowski (1993), Malanowski (1996, 2004), even in the presence of mixed control-state constraints and pure state constraints. Following Alt (1990, 1994), we exploit the equivalence between the SQP and the Lagrange-Newton methods, i.e., Newton's method, applied to a generalized (set-valued) equation representing necessary conditions of optimality. We concentrate on specific issues arising due to the semilinear state equation, e.g., the careful choice of suitable function spaces. An important step is the verification of the so-called strong regularity of the generalized equation, which is made difficult by the simultaneous presence of pure control and mixed control-state constraints (1.2). The key idea was recently developed in Alt et al. (2010), using concepts from Malanowski (2001).

We remark that strong regularity is known to be closely related to second-order sufficient conditions (SSC). For problems with pure control constraints, SSC are well understood and they are close to the necessary ones when so-called strongly active subsets are used, see, e.g., Tröltzsch (1999), Tröltzsch and Volkwein (2001), Tröltzsch and Wachsmuth (2006). However, the situation is more difficult for problems with mixed control-state constraints (Griesse and Wachsmuth, 2009; Rösch and Tröltzsch, 2003, 2006b) or even pure state constraints. In order to avoid a more technical discussion, we presently employ relatively strong SSC. We comment on the possibility of weakening these conditions in Section 8.

The material in this paper is organized as follows. In Section 2, we state our main assumptions and recall some properties concerning the state equation. Necessary and sufficient optimality conditions for problem (**P**) are stated in Section 3, and their reformulation as a generalized equation is given in Section 4. Section 5 addresses the equivalence of the SQP and Lagrange-Newton methods. Section 6 is devoted to the proof of strong regularity of the generalized equation. Finally, Section 7 completes the convergence analysis of the SQP method. A number of auxiliary results have been collected in the Appendix.

We denote by $L^p(\Omega)$ and $H^m(\Omega)$ the usual Lebesgue and Sobolev spaces (Adams, 1975), and $(\cdot,\cdot)$ is the scalar product in $L^2(\Omega)$ or $[L^2(\Omega)]^N$, respectively. $H_0^1(\Omega)$ is the subspace of $H^1(\Omega)$ with zero boundary traces, and $H^{-1}(\Omega)$ is its dual. The continuous embedding of a normed space X into a normed space Y is denoted by $X \hookrightarrow Y$. Throughout, we denote by $B_r^X(x)$ the open ball of radius r around x, in the topology of X. In particular, we write $B_r^\infty(x)$ for the open ball with respect to the $L^\infty(\Omega)$ norm. Throughout, c, c_1 etc. denote generic positive constants, whose value may change from instance to instance.

2. Assumptions and properties of the state equation

The following assumptions, (A1)–(A4), are taken to hold throughout the paper.

Assumptions

(A1) *Let Ω be a bounded domain in $\mathbb{R}^N$, $N \in \{2,3\}$ which is convex or has $C^{1,1}$ boundary $\partial\Omega$. The bound y_c is in $L^\infty(\Omega)$, and $\varepsilon > 0$.*

(A2) *The operator $A : H_0^1(\Omega) \to H^{-1}(\Omega)$ is defined as $A\,y(v) = a[y,v]$, where*

$$a[y,v] = \langle \nabla v, A_0 \nabla y\rangle_{H_0^1(\Omega),H^{-1}(\Omega)} + (c\,y, v).$$

A_0 is an $N \times N$ matrix with Lipschitz continuous entries on $\overline{\Omega}$ such that $\rho^\top A_0(\xi)\,\rho \geqslant m_0\,|\rho|^2$ holds with some $m_0 > 0$ for all $\rho \in \mathbb{R}^N$ and almost all $\xi \in \overline{\Omega}$. Moreover, $c \in L^\infty(\Omega)$ holds. The bilinear form $a[\cdot,\cdot]$ is not necessarily symmetric, but it is assumed to be continuous and coercive, i.e.,

$$a[y,v] \leqslant \overline{c}\,\|y\|_{H^1(\Omega)}\,\|v\|_{H^1(\Omega)}$$
$$a[y,y] \geqslant \underline{c}\,\|y\|^2_{H^1(\Omega)}$$

for all $y, v \in H_0^1(\Omega)$, with some positive constants $\overline{c}$ and $\underline{c}$. A simple example is $a[y,v] = (\nabla y, \nabla v)$, corresponding to $A = -\Delta$.

(A3) *$d(\xi, y)$ belongs to the C^2-class of functions with respect to y for almost all $\xi \in \Omega$. Moreover, d_{yy} is assumed be a locally bounded and locally Lipschitz continuous function with respect to y, i.e., the following conditions hold true: there exists $K > 0$ such that*

$$|d(\xi, 0)| + |d_y(\xi, 0)| + |d_{yy}(\xi, 0)| \leqslant K_d,$$

and for any $M > 0$, there exists $L_d(M) > 0$ such that

$$|d_{yy}(\xi, y_1) - d_{yy}(\xi, y_2)| \leqslant L_d(M)\, |y_1 - y_2| \quad \text{a.e. in } \Omega$$

for all $y_1, y_2 \in \mathbb{R}$ satisfying $|y_1|, |y_2| \leqslant M$.

Additionally, $d_y(\xi, y) \geqslant 0$ a.e. in Ω, for all $y \in \mathbb{R}$.

(A4) *The function $\phi = \phi(\xi, y, u)$ is measurable with respect to $\xi \in \Omega$ for each y and u, and of class C^2 with respect to y and u for almost all $\xi \in \Omega$. Moreover, the second derivatives are assumed to be locally bounded and locally Lipschitz continuous functions, i.e., the following conditions hold: there exist $K_y, K_u, K_{yu} > 0$ such that*

$$|\phi(\xi, 0, 0)| + |\phi_y(\xi, 0, 0)| + |\phi_{yy}(\xi, 0, 0)| \leqslant K_y, \quad |\phi_{yu}(\xi, 0, 0)| \leqslant K_{yu},$$
$$|\phi(\xi, 0, 0)| + |\phi_u(\xi, 0, 0)| + |\phi_{uu}(\xi, 0, 0)| \leqslant K_u.$$

Moreover, for any $M > 0$, there exists $L_\phi(M) > 0$ such that

$$|\phi_{yy}(\xi, y_1, u_1) - \phi_{yy}(\xi, y_2, u_2)| \leqslant L_\phi(M)\big(|y_1 - y_2| + |u_1 - u_2|\big),$$
$$|\phi_{yu}(\xi, y_1, u_1) - \phi_{yu}(\xi, y_2, u_2)| \leqslant L_\phi(M)\big(|y_1 - y_2| + |u_1 - u_2|\big),$$
$$|\phi_{uy}(\xi, y_1, u_1) - \phi_{uy}(\xi, y_2, u_2)| \leqslant L_\phi(M)\big(|y_1 - y_2| + |u_1 - u_2|\big),$$
$$|\phi_{uu}(\xi, y_1, u_1) - \phi_{uu}(\xi, y_2, u_2)| \leqslant L_\phi(M)\big(|y_1 - y_2| + |u_1 - u_2|\big)$$

for all $y_i, u_i \in \mathbb{R}$ satisfying $|y_i|, |u_i| \leqslant M$, $i = 1, 2$.

In addition, $\phi_{uu}(\xi, y, u) \geqslant m > 0$ a.e. in Ω, for all $(y, u) \in \mathbb{R}^2$.

In the sequel, we will simply write $d(y)$ instead of $d(\xi, y)$, etc. As a consequence of (A3)–(A4), the Nemyckii operators $d(\cdot)$ and $\phi(\cdot)$ are twice continuously Fréchet differentiable with respect to the $L^\infty(\Omega)$ norms, and their derivatives are locally Lipschitz continuous, see Lemma A.1.

The necessity of using $L^\infty(\Omega)$ norms for general nonlinearities d and ϕ motivates our choice

$$Y := H^2(\Omega) \cap H_0^1(\Omega)$$

as a state space, since $Y \hookrightarrow L^\infty(\Omega)$.

Remark 2.1 *In case Ω has only a Lipschitz boundary, our results remain true when Y is replaced by $H_0^1(\Omega) \cap L^\infty(\Omega)$.*

Recall that a function $y \in H_0^1(\Omega) \cap L^\infty(\Omega)$ is called a **weak solution** of (1.1) with $u \in L^2(\Omega)$ if $a[y,v] + (d(y), v) = (u, v)$ holds for all $v \in H_0^1(\Omega)$.

LEMMA 2.1 *Under assumptions* (A1)–(A3) *and for any given* $u \in L^2(\Omega)$*, the semilinear equation* (1.1) *possesses a unique weak solution* $y \in Y$*. It satisfies the a priori estimate*

$$\|y\|_{H^1(\Omega)} + \|y\|_{L^\infty(\Omega)} \leqslant C_\Omega \left(\|u\|_{L^2(\Omega)} + 1\right)$$

with a constant C_Ω *independent of* u*.*

Proof. The existence and uniqueness of a weak solution $y \in H_0^1(\Omega) \cap L^\infty(\Omega)$ is a standard result, Tröltzsch (2005, Theorem 4.8). It satisfies

$$\|y\|_{H^1(\Omega)} + \|y\|_{L^\infty(\Omega)} \leqslant C_\Omega \left(\|u\|_{L^2(\Omega)} + 1\right) =: M$$

with some constant C_Ω independent of u. Lemma A.1 (see the Appendix A) implies that $d(y) \in L^\infty(\Omega)$. Using the embedding $L^\infty(\Omega) \hookrightarrow L^2(\Omega)$, we conclude that the difference $u - d(y)$ belongs to $L^2(\Omega)$. Owing to assumption (A1), $y \in H^2(\Omega)$, see for instance Grisvard (1985, Theorem 2.2.2.3). ■

We will frequently also need the corresponding result for the linearized equation

$$\begin{aligned} A\,y + d_y(\overline{y})\,y &= u \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.1}$$

LEMMA 2.2 *Under assumptions* (A1)–(A3) *and given* $\overline{y} \in L^\infty(\Omega)$*, the linearized PDE* (2.1) *possesses a unique weak solution* $y \in Y$ *for any given* $u \in L^2(\Omega)$*. It satisfies the a priori estimate*

$$\|y\|_{H^2(\Omega)} \leqslant C_\Omega(\overline{y})\,\|u\|_{L^2(\Omega)}$$

with a constant $C_\Omega(\overline{y})$ *independent of* u*.*

Proof. According to (A3) and Lemma A.1, $d_y(\overline{y})$ is a nonnegative coefficient in $L^\infty(\Omega)$. The claim thus follows again from standard arguments, see, e.g., Grisvard (1985, Theorem 2.2.2.3). ■

EXAMPLE

We briefly comment on the existence of optimal controls for (**P**) in $L^\infty(\Omega)$, which we will suppose in the sequel. For the general objective function above, the existence of an optimal control in $L^2(\Omega)$ follows from the convexity assumption on ϕ in (A4), see Tröltzsch (2005, Theorem 4.13). However, our theory requires u^* to belong to $L^\infty(\Omega)$. Such a result typically follows from projection formulas, compare Rösch and Tröltzsch (2007). Projection formulas are derived from the first-order necessary optimality conditions, which, in turn, rely on the differentiability of the reduced objective. For general objective functions, this differentiability requires growth conditions.

The situation becomes easier for the following class of objective functions:

$$f(y,u) = \int_\Omega \psi(\xi, y(\xi))\, d\xi + \frac{\nu}{2} \|u\|^2_{L^2(\Omega)} .$$

The new function ψ is required to satisfy the same smoothness assumptions as in (A4). Clearly, $\phi_{uu} \equiv \nu > 0$ is satisfied. The reduced objective is twice differentiable w.r.t. $L^2(\Omega)$. The existence of Lagrange multipliers $\mu_i^* \in L^2(\Omega)$ follows even for $u^* \in L^2(\Omega)$. Using the projection formula (Griesse, Metla and Rösch, 2008, proof of Theorem 6.4), one can show that u^* and the Lagrange multipliers belong in fact to $L^\infty(\Omega)$. For examples of this type, all assumptions are satisfied, and only (**SSC**) and the separation assumption (A6) remain to be verified.

3. Necessary and sufficient optimality conditions

In this section, we introduce necessary and sufficient optimality conditions for problem (**P**). For convenience, we define the Lagrange functional

$$\mathcal{L} : Y \times L^\infty(\Omega) \times Y \times L^\infty(\Omega) \times L^\infty(\Omega) \to \mathbb{R}$$

as

$$\mathcal{L}(y,u,p,\mu_1,\mu_2) = f(y,u) + a[y,p] + (p, d(y) - u) - (\mu_1, u) - (\mu_2, \varepsilon\, u + y - y_c).$$

Here, μ_i are Lagrange multipliers associated to the inequality constraints, and p is the adjoint state. The existence of regular Lagrange multipliers $\mu_1, \mu_2 \in L^\infty(\Omega)$ was shown in Rösch and Tröltzsch (2006a, Theorem 7.3), which implies the following lemma:

LEMMA 3.1 *Suppose that $(y,u) \in Y \times L^\infty(\Omega)$ is a local optimal solution of* (**P**). *Then there exist regular Lagrange multipliers $\mu_1, \mu_2 \in L^\infty(\Omega)$ and an adjoint state $p \in Y$ such that the* **first-oder necessary optimality conditions**

$$\left.\begin{array}{rrr} \mathcal{L}_y(y,u,p,\mu_1,\mu_2) = 0, & \mathcal{L}_u(y,u,p,\mu_1,\mu_2) = 0, & \mathcal{L}_p(y,u,p,\mu_1,\mu_2) = 0, \\ u \geqslant 0, & \mu_1 \geqslant 0, & \mu_1 u = 0, \\ \varepsilon u + y - y_c \geqslant 0, & \mu_2 \geqslant 0, & \mu_2(\varepsilon u + y - y_c) = 0 \end{array}\right\} \quad (\mathbf{FON})$$

hold.

REMARK 3.1

1. *Note that due to the structure of the constraints, an additional regularity assumption such as the existence of a Slater point is not required.*

2. *The Lagrange multipliers and adjoint state associated to a local optimal solution of* (**P**) *need not be unique if the active sets* $\{\xi \in \Omega : u = 0\}$ *and* $\{\xi \in \Omega : \varepsilon\, u + y - y_c = 0\}$ *intersect nontrivially, see Alt et al., 2010, Remark 2.6. This situation will be excluded by Assumption* (A6) *below.*

Conditions (**FON**) are also stated in explicit form in (4.1) below. To guarantee that $x = (y, u)$ with associated multipliers $\lambda = (p, \mu_1, \mu_2)$ is a local solution of (**P**), we introduce the following **second-order sufficient optimality condition** (**SSC**):

There exists a constant $\alpha > 0$ such that

$$\mathcal{L}_{xx}(x, \lambda)(\delta x, \delta x) \geqslant \alpha \, \|\delta x\|^2_{[L^2(\Omega)]^2} \tag{3.1}$$

for all $\delta x = (\delta y, \delta u) \in Y \times L^\infty(\Omega)$ which satisfy the linearized equation

$$\begin{aligned} A\, \delta y + d_y(y)\, \delta y &= \delta u \quad \text{in } \Omega, \\ \delta y &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{3.2}$$

In (3.1), the Hessian of the Lagrange functional is given by

$$\mathcal{L}_{xx}(x, \lambda)(\delta x, \delta x) := \int_\Omega \begin{pmatrix} \delta y \\ \delta u \end{pmatrix}^\top \begin{pmatrix} \phi_{yy}(y,u) + d_{yy}(y)\, p & \phi_{yu}(y,u) \\ \phi_{uy}(y,u) & \phi_{uu}(y,u) \end{pmatrix} \begin{pmatrix} \delta y \\ \delta u \end{pmatrix} d\xi.$$

For convenience, we will use the abbreviation

$$X := Y \times L^\infty(\Omega) = H^2(\Omega) \cap H^1_0(\Omega) \times L^\infty(\Omega)$$

in the sequel.

ASSUMPTION

(A5) We assume that $x^* = (y^*, u^*) \in X$, together with associated Lagrange multipliers $\lambda^* = (p^*, \mu_1^*, \mu_2^*) \in Y \times [L^\infty(\Omega)]^2$, satisfies both (**FON**) and (**SSC**).

As mentioned in the introduction, the second-order sufficient conditions can be weakened by taking into account strongly active subsets. However, this would make the discussion more technical. Nevertheless, we comment on this possibility in Section 8.

DEFINITION 3.1

(a) A pair $x = (y, u) \in X$ *is called an* **admissible point** *if it satisfies* (1.1) *and* (1.2).

(b) A point $\bar{x} \in X$ *is called a* **strict local optimal solution in the sense of** $L^\infty(\Omega)$ *if there exists* $\varepsilon' > 0$ *such that the inequality* $f(\bar{x}) < f(x)$ *holds for all admissible* $x \in X \setminus \{\bar{x}\}$ *with* $\|x - \bar{x}\|_{[L^\infty(\Omega)]^2} \leqslant \varepsilon'$.

Theorem 3.1 *Under Assumptions* (A1)–(A5)*, there exists* $\beta > 0$ *and* $\varepsilon' > 0$ *such that*

$$f(x) \geqslant f(x^*) + \beta \, \|x - x^*\|^2_{[L^2(\Omega)]^2}$$

holds for all admissible $x \in X$ *with* $\|x - x^*\|_{[L^\infty(\Omega)]^2} \leqslant \varepsilon'$*. In particular,* x^* *is a strict local optimal solution in the sense of* $L^\infty(\Omega)$*.*

Proof. The proof can be done along the lines of Maurer (1981, Theorem 3.5). It has to observe the two-norm discrepancy principle and uses Lemma A.2. ∎

4. Generalized equation

We recall the necessary optimality conditions (**FON**) for problem (**P**), which read in the explicit form

$$\left.\begin{aligned}
a[v,p] + (d_y(y)p, v) + (\phi_y(y,u), v) - (\mu_2, v) = 0, &\quad \forall v \in H^1_0(\Omega)\\
(\phi_u(y,u), v) - (p, v) - (\mu_1, v) - (\varepsilon\mu_2, v) = 0, &\quad \forall v \in L^2(\Omega)\\
a[y,v] + (d(y), v) - (u, v) = 0, &\quad \forall v \in H^1_0(\Omega)\\
\left.\begin{array}{lll}\mu_1 \geqslant 0, & u \geqslant 0, & \mu_1 u = 0\\ \mu_2 \geqslant 0, & \varepsilon\, u + y - y_c \geqslant 0, & \mu_2(\varepsilon\, u + y - y_c) = 0\end{array}\right\} &\quad \text{a.e. in } \Omega.
\end{aligned}\right\} \tag{4.1}$$

As mentioned in the introduction, the local convergence analysis of SQP is based on its interpretation as Newton's method for a **generalized (set-valued) equation**

$$0 \in F(y, u, p, \mu_1, \mu_2) + N(y, u, p, \mu_1, \mu_2) \tag{4.2}$$

equivalent to (4.1). We define

$$K := \{\mu \in L^\infty(\Omega) : \mu \geqslant 0 \quad \text{a.e. in } \Omega\},$$

the cone of nonnegative functions in $L^\infty(\Omega)$, and the **dual cone** $N_1 : L^\infty(\Omega) \longrightarrow P(L^\infty(\Omega))$,

$$N_1(\mu) := \begin{cases} \{z \in L^\infty(\Omega) : (z, \mu - \nu) \geqslant 0 \quad \forall \nu \in K\} & \text{if } \mu \in K,\\ \emptyset & \text{if } \mu \notin K.\end{cases}$$

Here, $P(L^\infty)$ denotes the **power set** of $L^\infty(\Omega)$, i.e., the set of all subsets of $L^\infty(\Omega)$. In (4.2), F contains the single-valued part of (4.1), i.e.,

$$F(y, u, p, \mu_1, \mu_2) = \begin{pmatrix} A^\star p + d_y(y)\, p + \phi_y(y,u) - \mu_2\\ \phi_u(y,u) - p - \mu_1 - \varepsilon\mu_2\\ A\, y + d(y) - u\\ u\\ \varepsilon\, u + y - y_c \end{pmatrix}.$$

Both A and its formal adjoint $A^\star$ are considered here as operators from Y to $L^2(\Omega)$, i.e., $A\,y = -\text{div}\,(A_0 \nabla y) + c\,y$ and $A^\star p = -\text{div}\,(A_0^\top \nabla p) + c\,p$ hold. Moreover, N is the set-valued function

$$N(y, u, p, \mu_1, \mu_2) = \big(\{0\}, \{0\}, \{0\}, N_1(\mu_1), N_1(\mu_2)\big)^\top.$$

Note that the generalized equation (4.2) is nonlinear, since it contains the nonlinear functions d, d_y, ϕ_y and ϕ_u.

Remark 4.1 *Let*

$$\begin{aligned} W &:= Y \times L^\infty(\Omega) \times Y \times L^\infty(\Omega) \times L^\infty(\Omega), \\ Z &:= L^2(\Omega) \times L^\infty(\Omega) \times L^2(\Omega) \times L^\infty(\Omega) \times L^\infty(\Omega). \end{aligned}$$

Then $F : W \longrightarrow Z$ and $N : W \longrightarrow P(Z)$. Owing to Assumptions (A3) *and* (A4)*, F is continuously Fréchet differentiable with respect to the $L^\infty(\Omega)$ norms, see Lemma A.1.*

Lemma 4.1 *The first-order necessary conditions* (4.1) *and the generalized equation* (4.2) *are equivalent.*

Proof. (4.2) $\Rightarrow$ (4.1): This is immediate for the first three components. For the fourth component we have

$$\begin{aligned} & -u \in N_1(\mu_1) \\ \Rightarrow\quad & \mu_1 \in K \quad\text{and}\quad (-u, \mu_1 - \nu) \geqslant 0 \quad \text{for all } \nu \in K \\ \Rightarrow\quad & \mu_1(\xi) \geqslant 0 \quad\text{and}\quad -u(\xi)(\mu_1(\xi) - \nu) \geqslant 0 \quad \text{for all } \nu \geqslant 0, \quad \text{a.e. in } \Omega. \end{aligned}$$

This implies

$$\begin{aligned} \mu_1(\xi) = 0 \quad &\Rightarrow \quad u(\xi) \geqslant 0 \\ \mu_1(\xi) > 0 \quad &\Rightarrow \quad u(\xi) = 0, \end{aligned}$$

which shows the first complementarity system in (4.1). The second follows analogously.

(4.1) $\Rightarrow$ (4.2): This is again immediate for the first three components. From the first complementarity system in (4.1) we infer that

$$\begin{aligned} & u(\xi)\,\nu \geqslant 0 \quad \text{for all } \nu \geqslant 0, \quad \text{a.e. in } \Omega \\ \Rightarrow\quad & -u(\xi)(\mu_1(\xi) - \nu) \geqslant 0 \quad \text{for all } \nu \geqslant 0, \quad \text{a.e. in } \Omega \\ \Rightarrow\quad & -(u, \mu_1 - \nu) \geqslant 0 \quad \text{for all } \nu \in K. \end{aligned}$$

In view of $\mu_1 \in K$, this implies $-u \in N_1(\mu_1)$. Again, $-(\varepsilon\,u + y - y_c) \in N_1(\mu_2)$ follows analogously. ∎

5. The SQP method

In this section we briefly recall the SQP (sequential quadratic programming) method for the solution of problem (**P**). We also discuss its equivalence with Newton's method, applied to the generalized equation (4.2), which is often called the Lagrange-Newton approach. Throughout the rest of the paper we use the notation

$$w^k := (x^k, \lambda^k) = (y^k, u^k, p^k, \mu_1^k, \mu_2^k) \in W$$

to denote an iterate of either method. SQP methods break down the solution of (**P**) into a sequence of quadratic programming problems. At any given iterate w^k, one solves

$$\text{Minimize} \quad f_x(x^k)(x - x^k) + \frac{1}{2}\mathcal{L}_{xx}(x^k, \lambda^k)(x - x^k, x - x^k) \qquad (\mathbf{QP}_k)$$

subject to $x = (y, u) \in Y \times L^\infty(\Omega)$, the linear state equation

$$\begin{aligned} A\,y + d(y^k) + d_y(y^k)(y - y^k) &= u \quad \text{in } \Omega, \\ y &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{5.1}$$

and inequality constraints

$$\begin{aligned} u &\geqslant 0 \quad \text{in } \Omega, \\ \varepsilon\, u + y - y_c &\geqslant 0 \quad \text{in } \Omega. \end{aligned} \tag{5.2}$$

The solution (which needs to be shown to exist)

$$x = (y, u) \in Y \times L^\infty(\Omega),$$

together with the adjoint state and Lagrange multipliers

$$\lambda = (p, \mu_1, \mu_2) \in Y \times L^\infty(\Omega) \times L^\infty(\Omega),$$

will serve as the next iterate w^{k+1}.

Lemma 5.1 *There exists $R > 0$ such that $(\mathbf{QP}_k)$ has a unique global solution $x = (y, u) \in X$, provided that $(x^k, p^k) \in B_R^\infty(x^*, p^*)$.*

Proof. We have to verify that the feasible set

$$M^k := \{x = (y, u) \in Y \times L^2(\Omega) \text{ satisfying (5.1) and (5.2)}\}.$$

is nonempty. This follows from Alt et al. (2010, Lemma 2.3) using $\delta_3 = -d(y^k) + d_y(y^k)\,y^k$, whose proof uses the maximum principle for the differential operator $Ay + d_y(y^k)\,y$. Lemma A.3 allows to show the uniform convexity of the objective. The existence of a unique global optimal solution now follows from standard arguments. ∎

The solution (y,u) of $(\mathbf{QP}_k)$ and its Lagrange multipliers (p,μ_1,μ_2) are characterized by the first order optimality system (compare Alt et al., 2010, Lemma 2.5):

$$\left.\begin{array}{r} a[v,p]+(d_y(y^k)\,p,v)+(\phi_y(y^k,u^k),v)+(\phi_{yu}(y^k,u^k)(u-u^k),v) \\ +\big((\phi_{yy}(y^k,u^k)+d_{yy}(y^k)\,p^k)(y-y^k),v\big)-(\mu_2,v)=0, \quad \forall v\in H_0^1(\Omega) \\ (\phi_u(y^k,u^k),v)+(\phi_{uu}(y^k,u^k)(u-u^k),v) \\ +(\phi_{uy}(y^k,u^k)(y-y^k),v)-(p,v)-(\mu_1,v)-(\varepsilon\mu_2,v)=0, \quad \forall v\in L^2(\Omega) \\ a[y,v]+(d(y^k),v)+(d_y(y^k)(y-y^k),v)-(u,v)=0, \quad \forall v\in H_0^1(\Omega) \\ \left.\begin{array}{lll} \mu_1\geqslant 0, & u\geqslant 0, & \mu_1 u=0 \\ \mu_2\geqslant 0, & \varepsilon\,u+y-y_c\geqslant 0, & \mu_2(\varepsilon\,u+y-y_c)=0 \end{array}\right\} \quad \text{a.e. in } \Omega. \end{array}\right\} \tag{5.3}$$

Note that due to the convexity of the cost functional, (5.3) is both necessary and sufficient for optimality, provided that $(x^k,p^k)\in B_R^\infty(x^*,p^*)$.

REMARK 5.1 *The Lagrange multipliers* (μ_1,μ_2) *and the adjoint state* p *in* (5.3) *need not be unique, compare Alt et al. (2010, Remark 2.6). Non-uniqueness can occur only if* μ_1 *and* μ_2 *are simultaneously nonzero on a set of positive measure.*

We recall for convenience the generalized equation (4.2),

$$0\in F(w)+N(w). \tag{5.4}$$

Given the iterate w^k, Newton's method yields the next iterate w^{k+1} as the solution of the linearized generalized equation

$$0\in F(w^k)+F'(w^k)(w-w^k)+N(w). \tag{5.5}$$

Analogously to Lemma 4.1, one can show:

LEMMA 5.2 *System* (5.3) *and the linearized generalized equation* (5.5) *are equivalent.*

6. Strong regularity and implicit function theorem

The local convergence analysis of Newton's method (5.5) for the solution of (5.4) is based on a perturbation argument using

$$\delta\in F(w^*)+F'(w^*)(w-w^*)+N(w), \tag{6.1}$$

where w^* is a solution of (5.4), see for instance Alt (1994). We briefly sketch it here to make the necessary auxiliary results more apparent. The main ingredient in the convergence proof is the strong regularity of (5.4), see Definition 6.1 and Theorem 6.1 below. This property implies that (6.1) has a locally unique solution $w(\delta)$ for small δ, which depends Lipschitz continuously on δ.

The Newton step (5.5) can be equivalently expressed as

$$\delta^{k+1} \in F(w^*) + F'(w^*)(w^{k+1} - w^*) + N(w^{k+1}) \tag{6.2}$$

where

$$\delta^{k+1} := F(w^*) - F(w^k) + F'(w^*)(w^{k+1} - w^*) - F'(w^k)(w^{k+1} - w^k).$$

Since δ^{k+1} itself depends on the unknown solution w^{k+1}, we employ an implicit function theorem due to Dontchev (1995, Theorem 2.4) to get existence and local uniqueness of a solution to the Newton step (6.2) or (5.5), see Theorem 6.2. Note that in contrast to Lemma 5.1, this approach allows us to conclude local uniqueness also for the dual variables. A straightforward estimate for δ^{k+1} then implies the quadratic local convergence (Theorem 7.1). Throughout, the parameter δ belongs to the image space of F, i.e.,

$$Z := L^2(\Omega) \times L^\infty(\Omega) \times L^2(\Omega) \times L^\infty(\Omega) \times L^\infty(\Omega),$$

see Remark 4.1. Note that w^* is a solution of both (5.4) and (5.5) for $w^k = w^*$.

Definition 6.1 (*see Robinson, 1980*) *The generalized equation* (5.4) *is called* **strongly regular** *at w^* if there exist radii $r_1 > 0$, $r_2 > 0$ and a positive constant L_δ such that for all perturbations $\delta \in B^Z_{r_1}(0)$, the following hold:*

1. *the linearized equation* (6.1) *has in $B^W_{r_2}(w^*)$ a unique solution $w_\delta = w(\delta)$*
2. *w_δ satisfies the Lipschitz condition*

$$\|w_\delta - w_{\delta'}\|_W \leqslant L_\delta \|\delta - \delta'\|_Z \qquad \text{for all } \delta, \delta' \in B^Z_{r_1}(0).$$

The verification of strong regularity is based on the interpretation of (6.1) as the optimality system of the following QP problem, which depends on the perturbation δ:

$$\text{Minimize } f_x(x^*)(x - x^*) + \frac{1}{2}\mathcal{L}_{xx}(x^*, \lambda^*)\big(x - x^*, x - x^*\big) - \big([\delta_1, \delta_2], x - x^*\big) \tag{$\mathbf{LQP}(\delta)$}$$

subject to $x = (y, u) \in Y \times L^\infty(\Omega)$, the linear state equation

$$\begin{aligned} A\,y + d(y^*) + d_y(y^*)(y - y^*) &= u + \delta_3 \quad \text{in } \Omega, \\ y &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{6.3}$$

and inequality constraints

$$\begin{aligned} u &\geqslant \delta_4 \quad \text{in } \Omega, \\ \varepsilon\, u + y - y_c &\geqslant \delta_5 \quad \text{in } \Omega. \end{aligned} \tag{6.4}$$

As before, it is easy to check that the necessary optimality conditions of $(\mathbf{LQP}(\delta))$ are equivalent to (6.1).

Lemma 6.1 *For any $\delta \in Z$, problem $(\mathbf{LQP}(\delta))$ possesses a unique global solution $x_\delta = (y_\delta, u_\delta) \in X$. If $\lambda_\delta = (p_\delta, \mu_{1,\delta}, \mu_{2,\delta}) \in Y \times L^\infty(\Omega) \times L^\infty(\Omega)$ are associated Lagrange multipliers, then $(x_\delta, \lambda_\delta)$ satisfies (6.1). On the other hand, if any $(x_\delta, \lambda_\delta) \in W$ satisfies (6.1), then x_δ is the unique global solution of $(\mathbf{LQP}(\delta))$, and λ_δ are associated adjoint state and Lagrange multipliers.*

Proof. For any given $\delta \in Z$, let us denote by M_δ the set of all $x = (y, u) \in Y \times L^2(\Omega)$ satisfying (6.3) and (6.4). Then, M_δ is nonempty (as can be shown along the lines of Alt et al., 2010, Lemma 2.3), convex and closed. Moreover, (A5) implies that the cost functional $f_\delta(x)$ of $(\mathbf{LQP}(\delta))$ satisfies

$$f_\delta(x) \geqslant \frac{\alpha}{2} \|x\|^2_{[L^2(\Omega)]^2} + \text{ linear terms in } x$$

for all x satisfying (6.3). As in the proof of Lemma 5.1, we conclude that $(\mathbf{LQP}(\delta))$ has a unique solution $x_\delta = (y_\delta, u_\delta) \in X$.

Suppose that $\lambda_\delta = (p_\delta, \mu_{1,\delta}, \mu_{2,\delta}) \in Y \times L^\infty(\Omega) \times L^\infty(\Omega)$ are associated Lagrange multipliers, i.e., the necessary optimality conditions of $(\mathbf{LQP}(\delta))$ are satisfied. As argued above, it is easy to check that then (6.1) holds. On the other hand, suppose that any $(x_\delta, \lambda_\delta) \in W$ satisfies (6.1), i.e., the necessary optimality conditions of $(\mathbf{LQP}(\delta))$. As f_δ is strictly convex, these conditions are, likewise, sufficient for optimality, and the minimizer x_δ is unique. ∎

The proof of Lipschitz stability of solutions for problems of type $(\mathbf{LQP}(\delta))$ has recently been achieved in Alt et al. (2010). The main difficulty consisted in overcoming the non-uniqueness of the associated adjoint state and Lagrange multipliers. We follow the same technique here.

Definition 6.2 *Let $\sigma > 0$ be real number. We define two subsets of Ω,*

$$S_1^\sigma = \{\xi \in \Omega : 0 \leqslant u^*(\xi) \leqslant \sigma\}$$
$$S_2^\sigma = \{\xi \in \Omega : 0 \leqslant \varepsilon u^*(\xi) + y^*(\xi) - y_c(\xi) \leqslant \sigma\},$$

called the **security sets** *of level σ for* $(\mathbf{P})$.

Assumption
(A6) We require that $S_1^\sigma \cap S_2^\sigma = \emptyset$ for some fixed $\sigma > 0$.

From now on, we suppose (A1)–(A6) to hold. Assumption (A6) implies that the active sets

$$\mathcal{A}_1^* = \{\xi \in \Omega : u^*(\xi) = 0\}$$
$$\mathcal{A}_2^* = \{\xi \in \Omega : \varepsilon\, u^*(\xi) + y^*(\xi) - y_c(\xi) = 0\}$$

are well separated. This, in turn, implies the uniqueness of the Lagrange multipliers and adjoint state (p^*, μ_1^*, μ_2^*), see Alt et al. (2010, Lemma 3.1). Due to a continuity argument, the same conclusions hold for the solution and Lagrange multipliers of $(\mathbf{LQP}(\delta))$ for sufficiently small δ, as stated in the following theorem.

THEOREM 6.1 *There exist $G > 0$ and $L_\delta > 0$ such that $\|\delta\|_Z \leqslant G\,\sigma$ implies:*

1. *The Lagrange multipliers $\lambda_\delta = (p_\delta, \mu_{1,\delta}, \mu_{2,\delta})$ for $(\mathbf{LQP}(\delta))$ are unique.*
2. *For any such δ and δ', the corresponding solutions and Lagrange multipliers of $(\mathbf{LQP}(\delta))$ satisfy*

$$\|x_{\delta'} - x_\delta\|_{Y\times L^\infty(\Omega)} + \|\lambda_{\delta'} - \lambda_\delta\|_{Y\times L^\infty(\Omega)\times L^\infty(\Omega)} \leqslant L_\delta\,\|\delta' - \delta\|_Z. \quad (6.5)$$

The technique of proof was introduced in Alt et al. (2010), using ideas of Malanowski (2001). We only sketch the arguments and refer to the extended preprint by Griesse, Metla and Rösch (2008, Theorem 6.4) for the complete proof. One considers an auxiliary problem where the inequality constraints are restricted to the disjoint security sets. For this problem, the adjoint variables are unique, and one can show a stability estimate w.r.t. L^2. Using a projection formula (compare Alt et al., 2010, Lemma 2.7), the stability estimate can be lifted to L^∞ for the Lagrange multipliers and the control. Finally, one shows that the solution and Lagrange multipliers (extended by zero outside the security sets) of the auxiliary problem coincides with the solution of $(\mathbf{LQP}(\delta))$. A similar proof for a problem with quadratic objective function can be found in Griesse and Wachsmuth (2009), Proposition 3.3.

REMARK 6.1 *Theorem 6.1, together with Lemma 6.1, proves strong regularity of (5.4) at w^*.*

In order to apply the implicit function theorem (Dontchev, 1995, Theorem 2.4), we verify the following Lipschitz property for F:

LEMMA 6.2 *For any radii $r_3 > 0$, $r_4 > 0$ there exists $L > 0$ such that for any $\eta_1, \eta_2 \in B^W_{r_3}(w^*)$ and for all $w \in B^W_{r_4}(w^*)$ there holds the Lipschitz condition*

$$\|F(\eta_1) + F'(\eta_1)(w-\eta_1) - F(\eta_2) - F'(\eta_2)(w-\eta_2)\|_Z \leqslant L\,\|\eta_1 - \eta_2\|_W\,. \quad (6.6)$$

Proof. Let us denote $\eta_i = (y_i, u_i, p_i, \mu_1^i, \mu_2^i) \in B^W_{r_3}(w^*)$ and $w = (y, u, p, \mu_1, \mu_2) \in B^W_{r_4}(w^*)$, with $r_3, r_4 > 0$ arbitrary. A simple calculation shows that

$$\begin{aligned} &F(\eta_1) + F'(\eta_1)(w-\eta_1) - F(\eta_2) - F'(\eta_2)(w-\eta_2)\\ &\quad = (f_1(y_1,u_1,p_1) - f_1(y_2,u_2,p_2), f_2(y_1,u_1) - f_2(y_2,u_2), f_3(y_1) - f_3(y_2), 0, 0)^\top, \end{aligned}$$

where

$$\begin{aligned} f_1(y_i,u_i,p_i) &= d_y(y_i)\,p + \phi_y(y_i,u_i) + [\phi_{yy}(y_i,u_i) + d_{yy}(y_i)\,p_i](y-y_i)\\ &\quad + \phi_{yu}(y_i,u_i)(u-u_i)\\ f_2(y_i,u_i) &= \phi_u(y_i,u_i) + \phi_{uy}(y_i,u_i)(y-y_i) + \phi_{uu}(y_i,u_i)(u-u_i)\\ f_3(y_i) &= d(y_i) + d_y(y_i)(y-y_i). \end{aligned}$$

We consider only the Lipschitz condition for f_3, the rest follows analogously. Using the triangle inequality, we obtain

$$\begin{aligned}\|f_3(y_1) - f_3(y_2)\|_{L^2(\Omega)} &\leqslant \|d(y_1) - d(y_2)\|_{L^2(\Omega)} + \|d_y(y_1)(y_2 - y_1)\|_{L^2(\Omega)} \\ &\quad + \|(d_y(y_1) - d_y(y_2))(y - y_2)\|_{L^2(\Omega)} \\ &\leqslant \|d(y_1) - d(y_2)\|_{L^2(\Omega)} + \|d_y(y_1)\|_{L^\infty(\Omega)} \|y_2 - y_1\|_{L^2(\Omega)} \\ &\quad + \|d_y(y_1) - d_y(y_2)\|_{L^\infty(\Omega)} \|y - y_2\|_{L^2(\Omega)} .\end{aligned}$$

The properties of d, see Lemma A.1, imply that $\|d_y(y_1)\|_{L^\infty(\Omega)}$ is uniformly bounded for all $y_1 \in B^\infty_{r_3}(y^*)$. Moreover, $\|y - y_2\|_{L^2(\Omega)} \leqslant \|y - y^*\|_{L^2(\Omega)} + \|y^* - y_2\|_{L^2(\Omega)} \leqslant c\,(r_3 + r_4)$ holds. Together with the Lipschitz properties of d and d_y, see again Lemma A.1, we obtain

$$\|f_3(y_1) - f_3(y_2)\|_{L^2(\Omega)} \leqslant L\, \|y_1 - y_2\|_{L^\infty(\Omega)}$$

for some constant $L > 0$. ■

Now we apply Dontchev's implicit function theorem (Dontchev, 1995, Theorem 2.4 and Remark 2.5). Lemma 6.2 verifies assumption (*i*) of this theorem, and the strong regularity (Theorem 6.1 together with Lemma 6.1) corresponds to assumption (*iii*). We recall that we use this implicit function theorem to establish the (locally unique) solvability of the Newton step (5.5) or equivalently (6.2), in particular with regard to the dual variables. It is not needed to show the quadratic order of convergence.

THEOREM 6.2 *There exist radii $r_5 > 0$, $r_6 > 0$ such that for any $w^k \in B^W_{r_5}(w^*)$, there exists a solution $w^{k+1} \in B^W_{r_6}(w^*)$ of* (5.5)*, which is unique in this neighborhood.*

7. Local convergence analysis of SQP

This section is devoted to the local quadratic convergence analysis of the SQP method. As was shown in Section 5, the SQP method is equivalent to Newton's method (5.5), applied to the generalized equation (5.4). We recall the function spaces

$$\begin{aligned}&W := Y \times L^\infty(\Omega) \times Y \times L^\infty(\Omega) \times L^\infty(\Omega), \quad Y := H^2(\Omega) \cap H^1_0(\Omega) \\ &Z := L^2(\Omega) \times L^\infty(\Omega) \times L^2(\Omega) \times L^\infty(\Omega) \times L^\infty(\Omega).\end{aligned}$$

THEOREM 7.1 *There exists a radius $r > 0$ and a constant $C_{SQP} > 0$ such that for each starting point $w^0 \in B^W_r(w^*)$, every Newton step* (5.5) *has a unique solution in $B^W_r(w^*)$. The generated sequence satisfies*

$$\left\|w^{k+1} - w^*\right\|_W \leqslant C_{SQP} \left\|w^k - w^*\right\|^2_W .$$

Proof. The proof relies on standard arguments, see, e.g., Alt (1994), Dokov and Dontchev (1998) and it is therefore omitted here. We refer to the preprint Griesse, Metla and Rösch (2008, Theorem 7.1) for details. ∎

Clearly, Theorem 7.1 implies the local quadratic convergence of the SQP method. Recall that the iterates w^k are defined by means of Theorem 6.2, as *locally unique* solutions, Lagrange multipliers and adjoint states of $(\mathbf{QP}_k)$. Indeed, we can now prove that $w^{k+1} = (x^{k+1}, \lambda^{k+1})$ is *globally unique*, provided that w^k is already sufficiently close to w^*. For the primal variables x^{k+1}, this was already shown in Lemma 5.1.

Corollary 7.1 *There exists a radius $r' > 0$ such that $w^k \in B^W_{r'}(w^*)$ implies that $(\mathbf{QP}_k)$ has a unique global solution x^{k+1}. The associated Lagrange multipliers and adjoint state $\lambda^{k+1} = (p^{k+1}, \mu_1^{k+1}, \mu_2^{k+1})$ are also unique. The iterate w^{k+1} lies again in $B^W_{r'}(x^*, \lambda^*)$.*

Proof. We first observe that Theorem 7.1 remains valid (with the same constant C_{SQP}) if r is taken to be smaller than chosen in the proof. Here, we set

$$r' = \min\Big\{\sigma, \frac{\sigma}{c_\infty + \varepsilon}, R, r\Big\},$$

where R and r are the radii from Lemma 5.1 and Theorem 7.1, respectively, and c_∞ is the embedding constant of $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$.

Suppose that $w^k \in B^W_{r'}(w^*)$ holds. Then, Lemma 5.1 implies that $(\mathbf{QP}_k)$ possesses a globally unique solution $x^{k+1} \in Y \times L^\infty(\Omega)$. The corresponding active sets are defined by

$$\begin{aligned}\mathcal{A}_1^{k+1} &:= \{\xi \in \Omega : u^{k+1}(\xi) = 0\}\\ \mathcal{A}_2^{k+1} &:= \{\xi \in \Omega : \varepsilon\, u^{k+1}(\xi) + y^{k+1}(\xi) - y_c(\xi) = 0\}.\end{aligned}$$

We show that $\mathcal{A}_1^{k+1} \subset S_1^\sigma$ and $\mathcal{A}_2^{k+1} \subset S_2^\sigma$. For almost every $\xi \in \mathcal{A}_1^{k+1}$, we have

$$u^*(\xi) = u^*(\xi) - u^{k+1}(\xi) \leqslant \big\|u^* - u^{k+1}\big\|_{L^\infty(\Omega)} \leqslant r' \leqslant \sigma,$$

since Theorem 7.1 implies that $w^{k+1} \in B^W_{r'}(w^*)$ and thus, in particular, $u^{k+1} \in B^\infty_{r'}(u^*)$. By the same argument, for almost every $\xi \in \mathcal{A}_2^{k+1}$ we obtain

$$\begin{aligned} y^*(\xi) + \varepsilon\, u^*(\xi) - y_c(\xi) &= y^*(\xi) + \varepsilon\, u^*(\xi) - y^{k+1}(\xi) - \varepsilon\, u^{k+1}(\xi)\\ &\leqslant \big\|y^* - y^{k+1}\big\|_{L^\infty(\Omega)} + \varepsilon \big\|u^* - u^{k+1}\big\|_{L^\infty(\Omega)}\\ &\leqslant (c_\infty + \varepsilon)\, r' \leqslant \sigma.\end{aligned}$$

Owing to Assumption (A6), the active sets $\mathcal{A}_1^{k+1}$ and $\mathcal{A}_2^{k+1}$ are disjoint, and one can show as in Alt et al. (2010, Lemma 3.1) that the Lagrange multipliers μ_1^{k+1}, μ_2^{k+1} and adjoint state p^{k+1} are unique. ∎

8. Remark on second-order sufficient conditions

Finally, we comment on the possibility of weakening of the strong second-order sufficient conditions (**SSC**). It is enough to require the coercivity condition (3.1) on the critical subspace C_τ for some $\tau > 0$, defined by

$$C_\tau = \big\{\delta x = (\delta y, \delta u) \in Y \times L^\infty(\Omega) \text{ which satisfy the linearized equation (3.2)}, \\ \text{and } \delta u = 0 \text{ on } \mathcal{A}_1^\tau, \quad \varepsilon\, \delta u + \delta y = 0 \text{ on } \mathcal{A}_2^\tau\big\},$$

where

$$\mathcal{A}_1^\tau = \{\xi \in \Omega : \mu_1^*(\xi) \geqslant \tau\}, \quad \mathcal{A}_2^\tau = \{\xi \in \Omega : \mu_2^*(\xi) \geqslant \tau\}$$

are the strongly active subsets of level τ. It was shown in Meyer and Tröltzsch (2006) for a quadratic objective that the quadratic growth (Theorem 3.1) continues to hold for these weaker conditions. The result can be extended for our more general objective function. A further necessary modification concerns the proof of strong regularity (Theorem 6.1). This can be done along the lines of Griesse and Wachsmuth, 2009, Proposition 3.3. The remaining results carry over without change.

Appendix A. Auxiliary results

In this appendix we collect some auxiliary results. We begin with a standard result for the Nemyckii operators $d(\cdot)$ and $\phi(\cdot)$, whose proof can be found, e.g., in Tröltzsch (2005, Lemma 4.10, Satz 4.20). Throughout, we impose Assumptions (A1)–(A5).

LEMMA A.1 *The Nemyckii operator $d(\cdot)$ maps $L^\infty(\Omega)$ into $L^\infty(\Omega)$ and it is twice continuously differentiable in these spaces. For arbitrary $M > 0$, the Lipschitz condition*

$$\|d_{yy}(y_1) - d_{yy}(y_2)\|_{L^\infty(\Omega)} \leqslant L_d(M)\, \|y_1 - y_2\|_{L^\infty(\Omega)}$$

holds for all $y_i \in L^\infty(\Omega)$ such that $\|y_i\|_{L^\infty(\Omega)} \leqslant M$, $i = 1, 2$. In particular,

$$\|d_{yy}(y)\|_{L^\infty(\Omega)} \leqslant K_d + L_d(M)\, M$$

holds for all $y \in L^\infty(\Omega)$ such that $\|y\|_{L^\infty(\Omega)} \leqslant M$. The same properties, with different constants, are valid for $d_y(\cdot)$ and $d(\cdot)$. Analogous results hold for ϕ and its derivatives up to second-order, for all $(y, u) \in [L^\infty(\Omega)]^2$ such that $\|y_i\|_{L^\infty(\Omega)} + \|u_i\|_{L^\infty(\Omega)} \leqslant M$.

The remaining results address the coercivity of the second derivative of the Lagrangian, considered at different linearization points and for perturbed PDEs. Recall that $(x^*, \lambda^*) \in W$ satisfies the second-order sufficient conditions (**SSC**) with coercivity constant $\alpha > 0$, see (3.1).

Lemma A.2 *There exists $\varepsilon > 0$ and $\alpha' > 0$ such that*

$$\mathcal{L}_{xx}(x^*, \lambda^*)(x - x^*, x - x^*) \geqslant \alpha' \, \|x - x^*\|^2_{[L^2(\Omega)]^2} \tag{8.1}$$

holds for all $x = (y, u) \in Y \times L^\infty(\Omega)$ which satisfy the semilinear PDE (1.1) *and $\|x - x^*\|_{[L^\infty(\Omega)]^2} \leqslant \varepsilon$.*

Proof. Let $x = (y, u)$ satisfy (1.1). We define $\delta u = u - u^*$ and $\delta x = (\delta y, \delta u) \in Y \times L^\infty(\Omega)$ by

$$A \, \delta y + d_y(y^*) \, \delta y = \delta u \quad \text{on } \Omega$$

with homogeneous Dirichlet boundary conditions. Then the error $e := y^* - y - \delta y$ satisfies the linear PDE

$$A \, e + d_y(y^*) \, e = f \quad \text{on } \Omega \tag{8.2}$$

with homogeneous Dirichlet boundary conditions and

$$f := d(y) - d(y^*) - d_y(y^*)(y - y^*).$$

We estimate

$$\begin{aligned} \|f\|_{L^2(\Omega)} &= \left\| \int_0^1 \big[d_y(y^* + s(y - y^*)) - d_y(y^*) \big] \, ds \, (y - y^*) \right\|_{L^2(\Omega)} \\ &\leqslant L \int_0^1 s \, ds \, \|y - y^*\|_{L^\infty(\Omega)} \, \|y - y^*\|_{L^2(\Omega)} \\ &\leqslant \frac{L}{2} \, \|y - y^*\|_{L^\infty(\Omega)} \big(\|\delta y\|_{L^2(\Omega)} + \|e\|_{L^2(\Omega)} \big). \end{aligned}$$

In view of Lemma A.1, $d_y(y^*) \in L^\infty(\Omega)$ holds and it is a standard result that the unique solution e of (8.2) satisfies an a priori estimate

$$\|e\|_{L^\infty(\Omega)} \leqslant c \, \|f\|_{L^2(\Omega)} \, .$$

In view of the embedding $L^\infty(\Omega) \hookrightarrow L^2(\Omega)$, we obtain

$$\|e\|_{L^2(\Omega)} \leqslant c' \, \frac{L\varepsilon}{2} \big(\|\delta y\|_{L^2(\Omega)} + \|e\|_{L^2(\Omega)} \big).$$

For sufficiently small $\varepsilon > 0$, we can absorb the last term in the left hand side and obtain

$$\|e\|_{L^2(\Omega)} \leqslant c''(\varepsilon) \, \|\delta y\|_{L^2(\Omega)}$$

where $c''(\varepsilon) \searrow 0$ as $\varepsilon \searrow 0$. A straightforward application of Maurer and Zowe (1979, Lemma 5.5) concludes the proof. ∎

LEMMA A.3 *There exists $R > 0$ and $\alpha'' > 0$ such that*

$$\mathcal{L}_{xx}(x^k, \lambda^k)(x, x) \geqslant \alpha'' \|x\|^2_{[L^2(\Omega)]^2}$$

holds for all $(y, u) \in Y \times L^2(\Omega)$:

$$\begin{aligned} A\,y + d_y(y^k)\,y &= u \qquad \text{in } \Omega \\ y &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{8.3}$$

provided that $\|x^k - x^*\|_{[L^\infty(\Omega)]^2} + \|p^k - p^*\|_{L^\infty(\Omega)} < R$.

Proof. Let (y, u) be an arbitrary pair satisfying (8.3) and define $\hat{y} \in Y$ as the unique solution of

$$\begin{aligned} A\,\hat{y} + d_y(y^*)\,\hat{y} &= u \quad \text{in } \Omega \\ \hat{y} &= 0 \quad \text{on } \partial\Omega, \end{aligned}$$

for the same control u as above. Then $\delta y := y - \hat{y}$ satisfies

$$A\,\delta y + d_y(y^*)\,\delta y = \big(d_y(y^*) - d_y(y^k)\big)\,y \quad \text{in } \Omega$$

with homogeneous boundary conditions. A standard a priori estimate and the triangle inequality yield

$$\begin{aligned} \|\delta y\|_{L^2(\Omega)} &\leqslant \|d_y(y^*) - d_y(y^k)\|_{L^\infty(\Omega)} \|y\|_{L^2(\Omega)} \\ &\leqslant \|d_y(y^*) - d_y(y^k)\|_{L^\infty(\Omega)} \big(\|\hat{y}\|_{L^2(\Omega)} + \|\delta y\|_{L^2(\Omega)} \big). \end{aligned}$$

Due to the Lipschitz property of $d_y(\cdot)$ with respect to $L^\infty(\Omega)$, there exists a function $c(R)$ tending to 0 as $R \to 0$, such that $\|d_y(y^*) - d_y(y^k)\|_{L^\infty(\Omega)} \leqslant c(R)$, provided that $\|y^k - y^*\|_{L^\infty(\Omega)} < R$. For sufficiently small R, the term $\|\delta y\|_{L^2(\Omega)}$ can be absorbed in the left hand side, and we obtain

$$\|\delta y\|_{L^2(\Omega)} \leqslant c'(R)\,\|\hat{y}\|_{L^2(\Omega)}\,,$$

where $c'(R)$ has the same property as $c(R)$. Again, Maurer and Zowe (1979, Lemma 5.5) implies that there exists $\alpha_0 > 0$ and $R > 0$ such that

$$\mathcal{L}_{xx}(x^*, \lambda^*)(x, x) \geqslant \alpha_0 \|x\|^2_{[L^2(\Omega)]^2}\,,$$

provided that $\|y^k - y^*\|_{L^\infty(\Omega)} < R$.

Note that $\mathcal{L}_{xx}$ depends only on x and the adjoint state p. Owing to its Lipschitz property, we further conclude that

$$\begin{aligned} \mathcal{L}_{xx}(x^k, \lambda^k)(x, x) &= \mathcal{L}_{xx}(x^*, \lambda^*)(x, x) + \big[\mathcal{L}_{xx}(x^k, \lambda^k) - \mathcal{L}_{xx}(x^*, \lambda^*)\big](x, x) \\ &\geqslant \alpha_0 \|x\|^2_{[L^2(\Omega)]^2} - L\,\|(x^k, p^k) - (x^*, p^*)\|_{[L^\infty(\Omega)]^3} \|x\|^2_{[L^2(\Omega)]^2} \\ &\geqslant \big(\alpha_0 - L\,R\big) \|x\|^2_{[L^2(\Omega)]^2} =: \alpha'' \|x\|^2_{[L^2(\Omega)]^2}\,, \end{aligned}$$

given that $(x^k, p^k) \in B^\infty_R(x^*, p^*)$ and $\|x^k - x^*\|_{[L^\infty(\Omega)]^2} + \|p^k - p^*\|_{L^\infty(\Omega)} < R$. For sufficiently small R, we obtain $\alpha'' > 0$, which completes the proof. ■

Acknowledgement

This work was supported by the Austrian Science Fund FWF under project number P18056-N12.

References

Adams, R. (1975) *Sobolev Spaces.* Pure and Applied Mathematics **65**. Academic Press, New York-London.

Alt, W. (1990) The Lagrange-Newton method for infinite-dimensional optimization problems. *Numerical Functional Analysis and Optimization*, **11**, 201–224.

Alt, W. (1994) Local convergence of the Lagrange-Newton method with applications to optimal control. *Control & Cybernetics*, **23**(1–2), 87–105.

Alt, W., Griesse, R., Metla, N. and Rösch, A. (2010) Lipschitz stability for elliptic optimal control problems with mixed control-state constraints. *Optimization*, **59**(6), 833–849.

Alt, W. and Malanowski, K. (1993) The Lagrange-Newton method for nonlinear optimal control problems. *Computational Optimization and Applications. An International Journal*, **2**(1), 77–100.

Alt, W., Sontag, R. and Tröltzsch, F. (1996) An SQP method for optimal control of weakly singular Hammerstein integral equations. *Appl. Math. Optim.*, **33**(3), 227–252.

Dokov, S.P. and Dontchev, A.L. (1998) Robinson's strong regularity implies robust local convergence of Newton's method. In: *Optimal control (Gainesville, FL, 1997)*, *Appl. Optim.*, Kluwer Acad. Publ., Dordrecht, **15**, 116–129.

Dontchev, A. (1995) Implicit function theorems for generalized equations. *Mathematical Programming*, **70**, 91–106.

Griesse, R., Metla, N. and Rösch, A. (2008) Local quadratic convergence of SQP for elliptic optimal control problems with mixed control-state constraints. RICAM Report 2008–21, Johann Radon Institute for Computational and Applied Mathematics (RICAM), Austrian Academy of Sciences, Linz, Austria. `http://www.ricam.oeaw.ac.at/publications/reports/08/rep08-21.pdf`.

Griesse, R. and Wachsmuth, D. (2009) Sensitivity analysis and the adjoint update strategy for optimal control problems with mixed control-state constraints. *Computational Optimization and Applications*, **44**(1), 57–81.

Grisvard, P. (1985) *Elliptic Problems in Nonsmooth Domains.* Pitman, Boston.

Heinkenschloss, M. (1998) Formulation and analysis of a sequential quadratic programming method for the optimal Dirichlet boundary control of Navier-Stokes flow. In: *Optimal Control (Gainesville, FL, 1997)*, *Applied Optimization*, Kluwer Academic Publishers, Boston, **15**, 178–203.

HEINKENSCHLOSS, M. and TRÖLTZSCH, F. (1998) Analysis of the Lagrange-SQP-Newton Method for the Control of a Phase-Field Equation. *Control & Cybernetics*, **28**, 177–211.

MALANOWSKI, K. (1996) Convergence of Lagrange-Newton method for control-state and pure state constrained optimal control problems. In: *System modelling and optimization (Prague, 1995).* Chapman & Hall, London, 419–426.

MALANOWSKI, K. (2001) Stability and sensitivity analysis for optimal control problems with control-state constraints. *Dissertationes Mathematicae (Rozprawy Matematyczne)*, 394.

MALANOWSKI, K. (2004) Convergence of the Lagrange-Newton method for optimal control problems. *International Journal of Applied Mathematics and Computer Science*, **14**(4), 531–540.

MAURER, H. (1981) First and second order sufficient optimality conditions in mathematical programming and optimal control. *Mathematical Programming Study*, **14**, 163–177.

MAURER, H. and ZOWE, J. (1979) First and second order necessary and sufficient optimality conditions for infinite-dimensional programming problems. *Mathematical Programming*, **16**, 98–110.

MEYER, C., PRÜFERT, U. and TRÖLTZSCH, F. (2007) On two numerical methods for state-constrained elliptic control problems. *Optimization Methods and Software*, **22**(6), 871–899.

MEYER, C., RÖSCH, A. and TRÖLTZSCH, F. (2005) Optimal control of PDEs with regularized pointwise state constraints. *Computational Optimization and Applications*, **33**(2–3), 209–228.

MEYER, C. and TRÖLTZSCH, F. (2006) On an elliptic optimal control problem with pointwise mixed control-state constraints. In: A. Seeger, ed., *Recent Advances in Optimization. Proceedings of the 12th French-German-Spanish Conference on Optimization.* **LNEMS 563**, Springer, New York, 187–204.

ROBINSON, S. (1980) Strongly regular generalized equations. *Mathematics of Operations Research*, **5**(1), 43–62.

RÖSCH, A. and TRÖLTZSCH, F. (2003) Sufficient second-order optimality conditions for a parabolic optimal control problem with pointwise control-state constraints. *SIAM Journal on Control and Optimization*, **42**(1), 138–154.

RÖSCH, A. and TRÖLTZSCH, F. (2006a) Existence of regular Lagrange multipliers for elliptic optimal control problems with pointwise control-state constraints. *SIAM Journal on Control and Optimization*, **45**(2), 548–564.

RÖSCH, A. and TRÖLTZSCH, F. (2006b) Sufficient second-order optimality conditions for an elliptic optimal control problem with pointwise control-state constraints. *SIAM Journal on Optimization*, **17**(3), 776–794.

RÖSCH, A. and TRÖLTZSCH, F. (2007) On regularity of solutions and Lagrange multipliers of optimal control problems for semilinear equations with mixed pointwise control-state constraints. *SIAM Journal on Control and Optimization*, **46**(3), 1098–1115.

Tröltzsch, F. (1994) An SQP method for optimal control of a nonlinear heat equation. *Control & Cybernetics*, **23**, 267–288.

Tröltzsch, F. (1999) On the Lagrange-Newton-SQP method for the optimal control of semilinear parabolic equations. *SIAM Journal on Control and Optimization*, **38**(1), 294–312.

Tröltzsch, F. (2005) *Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen.* Vieweg, Wiesbaden.

Tröltzsch, F. and Volkwein, S. (2001) The SQP method for control constrained optimal control of the Burgers equation. *ESAIM: Control, Optimisation and Calculus of Variations*, **6**, 649–674.

Tröltzsch, F. and Wachsmuth, D. (2006) Second-order sufficient optimality conditions for the optimal control of Navier-Stokes equations. *ESAIM: Control, Optimisation and Calculus of Variations*, **12**(1), 93–119.