Control and Cybernetics

vol. **38** (2009) No. 4A

Optimal control of partial differential equations with affine control constraints*

by

Juan Carlos De Los Reyes[1] **and Karl Kunisch**[2]

[1]Departmento de Matemática, EPN Quito, Ecuador
[2]Institute for Mathematics and Scientific Computing, University of Graz, Austria

Abstract: Numerical solution of PDE optimal control problems involving affine pointwise control constraints is investigated. Optimality conditions are derived and a semi-smooth Newton method is presented. Global and local superlinear convergence of the method are obtained for linear problems. Differently from box constraints, in the case of general affine constraints a proper weighting of the control costs is essential for superlinear convergence of semi-smooth Newton methods. This is also demonstrated numerically by controlling the two-dimensional Stokes equations with different kinds of affine constraints.

Keywords: optimal control, affine control constraints, semi-smooth Newton methods.

1. Introduction

Numerical solution of optimal control problems in presence of control constraints has become an active research field in recent years. The presence of pointwise constraints on the control adds new analytical and numerical difficulties to the already challenging unconstrained control problems. The importance of control constraints becomes clear if e.g. technological or financial restrictions are given. In recent years the case of box constraints has received a considerable amount of attention, both analytically and numerically. We refer to e.g. Bergounioux (1999), Ito and Kunisch (2004), Kunisch and Rösch (2002), Casas and Tröltzsch (2002), Weiser (2005), and the literature cited there.

Methods that are being investigated include the primal-dual active set strategy, semi-smooth Newton methods, Lavrentiev regularization and interior point methods. A special feature of the primal-dual active set strategy and the semi-smooth Newton methods lies in the fact that for box constrained optimal control problems it can be shown for a rather general class of problems that these

*Submitted: February 2009; Accepted: June 2009.

methods do not require an additional regularization parameter for obtaining local super-linear convergence. For certain classes of problems it was shown, moreover, that the primal-dual active set method and the semi-smooth Newton method coincide, Hintermüller et al. (2002), and that convergence from arbitrary initial data holds, Ito and Kunisch (2004). Such methods have been utilized for the control of a variety of phenomena including fluid flow, De Los Reyes and Kunisch (2005), Hintermüller and Hinze (2006), reaction diffusion systems, Griesse and Volkwein (2005), or crystal growth processes, Meyer et al. (2006).

While box constrained problems are fairly well understood, relatively little research was directed towards devising and analyzing efficient second order type methods for more general constraints. This paper is intended to make a step in this direction by considering constraints of affine type. While for the derivation of optimality conditions we consider a quite general class of affine constraints, in the numerical part of this paper we restrict ourselves to the case where the matrix, characterizing the inequality constraints, is surjective. The non-surjective case appears to require a different treatment.

After deriving optimality conditions, the algorithm is analyzed, showing its equivalence to primal-dual iterations and proving its local superlinear convergence for linear state equations. Subsequently, a certain class of optimal control problems with nonlinear dynamics is treated as well. First order necessary and second order sufficient optimality conditions are derived. Differently from previous contributions (see Bonnans, 1998; Dunn, 1995; Wachsmuth, 2006), the second order sufficient optimality condition avoids the so called two-norm discrepancy by using a contradiction argument and exploiting the structure of the cost functional.

To obtain the desired superlinear convergence rate for the semi-smooth Newton method it is necessary that the control cost term is put into correspondence to the type of constraints that are imposed. In particular, the angle between the affine manifolds characterizing the control constraints influences the numerical behavior of the algorithm. Numerical examples involving the optimal control of the Stokes equations will illustrate the practical importance of this issue.

The outline of the paper is as follows. In Section 2 the optimal control problem with affine state equations is studied and an optimality system is derived. For the solution of this system, a semi-smooth Newton method is proposed in Section 3. Local superlinear convergence of the method is proved and its equivalence to the primal-dual active set strategy shown. In Section 4 a global convergence result for the semi-smooth Newton algorithm is presented. Nonlinear control problems are studied in Section 5. Optimality conditions of first and second order are obtained and sufficient conditions for local superlinear convergence of the semi-smooth Newton method are proved. Finally, in Section 6, numerical experiments illustrate the importance of using a proper weight in the control costs.

2. Optimal control problem

Let Ω be a bounded domain of $\mathbb{R}^n$. We consider the following optimal control problem:

$$\begin{cases} \min \; J(y) + \frac{\alpha}{2}\,|Cu|^2_{L^2(\hat{\Omega},\mathbb{R}^l)} + \frac{\alpha}{2}\,|Pu|^2_U \\ \text{subject to} \\ e(y,u) = f \\ Cu \le \psi \quad \text{a.e.,} \end{cases} \tag{1}$$

where $\alpha > 0$, $C \in \mathbb{R}^{l\times m}$, $\psi \in L^2(\hat{\Omega},\mathbb{R}^l)$ and $P : \mathbb{R}^m \to \mathbb{R}^m$ is the orthogonal projection onto $\ker(C)$. The operator $e : Y \times U \to Y'$, with Y, U Hilbert spaces, is assumed to be linear and continuous in both variables, i.e.

$$e(y,u) = e_y(y) + e_u(u),$$

with $e_y \in \mathcal{L}(Y,Y')$, $e_u \in \mathcal{L}(U,Y')$. The Hilbert spaces are $\mathbb{R}^n$-valued function spaces over a bounded domain $\Omega \subset \mathbb{R}^n$, such that $Y = H^1(\Omega,\mathbb{R}^n)$. Throughout, the space of controls is

$$U = L^2(\hat{\Omega},\mathbb{R}^m),\ \hat{\Omega} \subset \Omega \subset \mathbb{R}^n.$$

Further, we choose $f \in Y'$ and J as

$$J(y) = \frac{1}{2}(y,Qy)_Y + (q,y)_Y,$$

where $Q \in \mathcal{L}(Y,Y)$, Q is positive semi-definite and $q \in Y$.

Assumption 1 *The operator e_y is continuously invertible.*

In particular, this implies that for every $u \in U$, there exists a unique $y = y(u) \in Y$ such that $e(y,u) = f$ and the mapping $u \mapsto y(u)$ is a continuous affine operator from U to Y.

Example 1 *Consider the Stokes equations in $\Omega \subset \mathbb{R}^n$*

$$\begin{aligned} -\nu\Delta y + \nabla p &= u && \text{in } \Omega \\ \operatorname{div} y &= 0 && \text{in } \Omega \\ y &= 0 && \text{on } \partial\Omega, \end{aligned}$$

where ν stands for the viscosity coefficient of the fluid, y for the velocity vector field, p for the scalar pressure and u for a distributed body force. Introducing the spaces $Y = \{H^1_0(\Omega,\mathbb{R}^n) : \operatorname{div} y = 0\}$, $U = L^2(\Omega,\mathbb{R}^n)$ and the operator

$$e : Y \to Y' \tag{2}$$

$$y \mapsto \nu(\nabla y, \nabla\cdot)_U - (u,\cdot)_U, \tag{3}$$

the equations can be formulated in weak form as

$$e(y, u) = 0 \text{ in } Y',$$

and satisfy Assumption 1.

Throughout the paper we will use the following assumption with respect to the restriction matrix:

ASSUMPTION 2 *For any $v \in \mathbb{R}^m$ satisfying $Cv \leq \psi(x)$ and for a.e. $x \in \hat{\Omega}$ the rows $\{C_i\}_{i \in \mathcal{A}(v,x)}$ are linearly independent in $\mathbb{R}^m$, where $\mathcal{A}(v, x) := \{i : (Cv)_i = \psi_i(x)\}$.*

EXAMPLE 2 *For $U = L^2(\hat{\Omega}, \mathbb{R}^2)$, the constraints $u_1 \leq \psi_1, u_2 \leq \psi_2$ result in $C = I, P = 0$, which was considered in previous work (see De Los Reyes, 2006; De Los Reyes and Kunisch, 2005).*

EXAMPLE 3 *The case $U = L^2(\Omega, \mathbb{R}^m)$, $u_i \leq 0$, $-1 \leq \sum_{i=1}^m u_i$ results in $l = m + 1$, $C = \begin{pmatrix} I \\ -e \end{pmatrix}$, where $e = (1, \ldots, 1) \in \mathbb{R}^m$, I is the $m \times m$ identity matrix, $\psi = (0, \ldots, 0, 1)$ and $P = 0$. Here, Assumption 2 is satisfied.*

EXAMPLE 4 *The case $U = L^2(\Omega, \mathbb{R}^m)$ with bilateral constraints $\psi_1 \leq u_1 \leq \psi_2$ a.e. in Ω results in $C = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ -1 & 0 & \cdots & 0 \end{pmatrix}$, which also satisfies the requirements of Assumption 2. The techniques of Ito and Kunisch (2004) are applicable in this case.*

THEOREM 1 *If there exists a feasible control $u \in L^2(\hat{\Omega}, \mathbb{R}^m)$ satisfying $Cu \leq \psi$ and Assumption 1 holds, then there exists a unique solution $(y^*, u^*) = (y(u^*), u^*)$ to problem* (1).

Proof. Let $(y(u_n), u_n)$ be a minimizing sequence. From the structure of the cost functional the sequence $\{u_n\}_{n=1}^\infty$ is bounded in U and hence there exists a weakly convergent subsequence, denoted again by $\{u_n\}$, and a limit $u^* \in U$ such that $u_n \rightharpoonup u^*$ in U. We have $Cu^*(x) \leq \psi(x)$ a.e. and, since a bounded linear operator maps weakly convergent sequences into weakly convergent sequences, $y(u_n) \rightharpoonup y^* = y(u^*)$.

Moreover, since the cost functional

$$\mathcal{J}(y, u) = J(y) + \frac{\alpha}{2} |Cu|^2_{L^2(\hat{\Omega}, \mathbb{R}^l)} + \frac{\alpha}{2} |Pu|^2_U$$

is weakly lower semi-continuous, it follows that

$$\mathcal{J}(y^*, u^*) \leq \liminf \mathcal{J}(y_n, u_n),$$

and, therefore, (y^*, u^*) is a solution to (1). Uniqueness follows from the strict convexity of the cost functional and the affine properties of the constraints. ∎

Next, existence of Lagrange multipliers for (1) is verified and a first order optimality system is derived. We set

$$\mathcal{A} = \bigcup_{i=1}^{l} \mathcal{A}_i$$

where

$$\mathcal{A}_i = \{x \in \hat{\Omega} : C_i u(x) = \psi_i(x)\},$$

and define the inactive set $\mathcal{I} = \Omega \backslash \mathcal{A}$.

THEOREM 2 *Let Assumptions 1 and 2 hold.Then there exist multipliers $p \in Y$ and $\lambda \in L^2(\hat{\Omega}, \mathbb{R}^l)$ such that*

$$e(y^*, u^*) = f \tag{4}$$

$$e_y^\star p = -J'(y^*) \tag{5}$$

$$\alpha C^T C u^* + \alpha P u^* + C^T \lambda + e_u^\star p = 0 \tag{6}$$

$$C u^* \leq \psi,\ \lambda \geq 0,\ \lambda^T (C u^* - \psi) = 0 \ \textit{a.e. in}\ \hat{\Omega}. \tag{7}$$

Proof. The first order necessary and sufficient optimality condition satisfied by u^* is given by

$$(\alpha C^T C u^* + \alpha P u + e_u^\star p, u - u^*)_{L^2(\hat{\Omega})} \geq 0, \tag{8}$$

for all $u \in L^2(\hat{\Omega}, \mathbb{R}^m)$ satisfying $Cu \leq \psi$, where

$$\begin{cases} e(y^*, u^*) = f \\ e_y^\star p = -J'(y^*). \end{cases} \tag{9}$$

To obtain from (8) the pointwise almost everywhere relations given in (7) it will be convenient to introduce an auxiliary control problem for which the pointwise control constraints are imposed only on the active set. For that purpose we define a partitioning of the active set next.

Note that by Assumption 2 at most m constraints can be active simultaneously at a.e. $x \in \hat{\Omega}$. Let $\mathcal{P}$ be the set of all subsets of $\{1, ..., l\}$ of cardinality $\leq m$ and set for $I \in \mathcal{P}$

$$\Omega_I = \{x \in \hat{\Omega} : C_j u^*(x) = \psi_j(x),\ \text{for all } j \in I\}.$$

Then $\hat{\Omega} = \biguplus_{I \subset \mathcal{P}} \Omega_I \uplus \mathcal{I}$ and we have

$$\mathcal{A}_i = \{x \in \Omega_I : C_i u(x) = \psi_i(x) \text{ for some } I \in \mathcal{P}\}.$$

We consider the auxiliary problem

$$\begin{cases} \min_{u\in L^2(\hat\Omega,\mathbb{R}^m)} J(y) + |Cu|^2_{L^2(\hat\Omega,\mathbb{R}^l)} + \frac{\alpha}{2}\,|Pu|^2_U \\ \text{subject to:} \\ e(y,u) = f \\ C_i u \le \psi_i \text{ on } \Omega_I \text{ for } i \in I \text{ and all } I \in \mathcal{P}. \end{cases} \tag{P_{aux}}$$

Note that the inequality constraints in (P_{aux}) can equivalently be expressed as $C_i u \le \psi_i$ on $\mathcal{A}_i$, for $i = 1,\dots,l$. Clearly, (P_{aux}) admits a unique solution $\hat u \in U$. Associated to (P_{aux}) we introduce the Lagrangian $\mathcal{L} : Y\times U\times Y\times Z \to \mathbb{R}$, where $Z = \bigotimes_{I\in\mathcal{P}} L^2(\Omega_I,\mathbb{R}^{\#(I)})$, with $\#(I)$ the cardinality of I,

$$\mathcal{L}(y,u,p,\tilde\lambda) = J(y) + |Cu|^2_{L^2(\hat\Omega,\mathbb{R}^l)} + \frac{\alpha}{2}\,|Pu|^2_U + \langle p, e(y,u) - f\rangle_{Y,Y'} + \sum_{I\in\mathcal{P}}\sum_{i\in I}(\lambda_i^I, C_i\,u - \psi_i)_{L^2(\Omega_I,\mathbb{R})}.$$

By Assumptions 1 and 2 the linearized constraints

$$(e, \{(C_i)_{i\in I} : I \in \mathcal{P}\}) : Y\times U \to Y' \times Z$$

are surjective. Here we identify U with $\bigotimes_{I\in\mathcal{P}} L^2(\Omega_I,\mathbb{R}^l) \times L^2(\mathcal{I},\mathbb{R}^l)$. Hence, there exists $(p, \{\lambda^I\}_{I\in\mathcal{P}}) \in Y\times Z$, which is a Lagrange multiplier for (P_{aux}), i.e.:

$$\begin{cases} e(\hat y,\hat u) = f \\ e_y^\star p = -J'(\hat y) \\ \alpha C^T C\hat u + \alpha P\hat u + e_u^\star p + \sum_{I\in\mathcal{P}}\sum_{i\in I} C_i^T \lambda_i^I \chi_{\Omega_I} = 0 \\ C\hat u \le \psi, \text{ in } \hat\Omega \\ \lambda_i^I \ge 0,\ \lambda_i(C_i\hat u - \psi_i) = 0, i\in I,\ I\in\mathcal{P}. \end{cases} \tag{10}$$

Defining $\lambda \in L^2(\hat\Omega,\mathbb{R}^l)$ by setting

$$\lambda_i = \lambda_i^I \text{ for } i\in I, \text{ and } \lambda_i = 0 \text{ for } i \notin I, \qquad \text{for any } \ I\in\mathcal{P},\ x\in\Omega_I;$$

$$\lambda_i = 0 \text{ on } \mathcal{I},$$

(10) can equivalently expressed as

$$\begin{cases} e(\hat y,\hat u) = f \\ e_y^\star p = -J'(\hat y) \\ \alpha C^T C\hat u + \alpha P\hat u + e_u^\star p + C^T\lambda = 0 \\ \lambda \ge 0,\ C\hat u \le \psi,\ (\lambda, C\hat u - \psi)_{L^2(\hat\Omega,\mathbb{R}^l)} = 0. \end{cases} \tag{11}$$

From (11) we obtain for $Cu \le \psi$,

$$(\alpha C^T C\hat u + \alpha P\hat u + e_u^\star p, u - \hat u) = (\lambda, C\hat u - Cu) = (\lambda, \psi - Cu) \ge 0.$$

Hence, $\hat{u}$ satisfies the first order condition (8) and therefore $\hat{u} = u^*$. System (4)-(7) follows from (11). ∎

Remark 1 *Note that from equation* (6) *we have* $\alpha P u^* + P e_u^\star p = 0$.

3. Semi-smooth Newton method

The complementarity condition (7) can be reformulated as the following operator equation

$$\lambda = \max(0, \lambda + c(Cu^* - \psi)), \tag{12}$$

for any $c > 0$. Here *max* is interpreted componentwise and (12) must be interpreted in the a.e. in $\hat{\Omega}$ sense. Throughout the remainder of this section we assume that

$$C \text{ is surjective.} \tag{13}$$

Then C^T is injective and (6) can equivalently be expressed as

$$\begin{cases} \alpha C u^* + \lambda + D^{-1} C P_{R(C^T)} e_u^\star p = 0 \\ \alpha P u^* + P e_u^\star p = 0, \end{cases} \tag{14}$$

where

$$D = CC^T \in \mathbb{R}^{l \times l}$$

and $P_{R(C^T)}$ denotes the projection onto $range(C^T)$. Since C^T is injective, $range(C^T) = \mathbb{R}^m$ and we may replace the operator $P_{R(C^T)}$ by the identity matrix. Choosing $c = \alpha$ in (12) results in

$$-\alpha C u - D^{-1} C e_u^\star p = \max(0, -D^{-1} C e_u^\star p - \alpha\psi). \tag{15}$$

Considering p as a function of u given by equations in (4)-(5), the optimality system can equivalently be expressed as

$$F(u) = 0, \tag{16}$$

where $F : L^2(\hat{\Omega}; \mathbb{R}^m) \to L^2(\hat{\Omega}; \mathbb{R}^l) \times L^2(\hat{\Omega}; \mathbb{R}^m)$ is defined by

$$F(u) = \begin{pmatrix} \alpha C u + D^{-1} C e_u^\star p + \max(0, -D^{-1} C e_u^\star p - \alpha\psi) \\ \alpha P u + P e_u^* p \end{pmatrix}, \tag{17}$$

and $p = p(u)$. We shall use a semi-smooth Newton approach to solve (17). For this purpose it is convenient to recall the following definition and superlinear convergence result from Hintermüller, Ito and Kunisch (2002).

Definition 1 *Let X and Z be Banach spaces and $D \subset X$ an open subset. The mapping $F : D \to Z$ is called Newton differentiable on the open subset $U \subset D$ if there exists a generalized derivative $G : U \to L(X, Z)$ such that*

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

for every $x \in U$.

Proposition 1 *If x^* is a solution of $F(x) = 0$, F is Newton differentiable in an open neighborhood U containing x^* with generalized derivative G. If $\{|G(y)^{-1}|_{\mathcal{L}(Z,X)} : y \in U\}$ is bounded, then the Newton iterations*

$$x_{k+1} = x_k - G(x_k)^{-1} F(x_k)$$

converge superlinearly to x^*, provided that $|x_0 - x^*|_X$ is sufficiently small.

We shall apply Proposition 1 with $X = L^2(\hat{\Omega}, \mathbb{R}^m)$ and $Z = L^2(\hat{\Omega}, \mathbb{R}^l) \times L^2(\hat{\Omega}, \ker C)$. To define a generalized derivative of F in the sense of Definition 1 we first introduce a generalized derivative for $\max : L^2(\hat{\Omega}, \mathbb{R}^l) \mapsto L^2(\hat{\Omega}, \mathbb{R}^l)$ by setting

$$(G_m \varphi(x))_i = \begin{cases} 1 & \text{if } \varphi(x)_i > 0 \\ 0 & \text{if } \varphi(x)_i \leq 0. \end{cases} \tag{18}$$

From Hintermüller, Ito and Kunisch (2002) it is known that $\max : L^q(\hat{\Omega}, \mathbb{R}^l) \mapsto L^2(\hat{\Omega}, \mathbb{R}^l)$ is Newton differentiable with generalized derivative given by (18) if $q > 2$. We henceforth assume that

$$h \to p(h) \text{ is continuous from } L^2(\hat{\Omega}; \mathbb{R}^m) \text{ to } L^q(\Omega; \mathbb{R}^n), \text{ for some } q > 2. \tag{19}$$

For Example 1, $p \in H^1_0(\Omega, \mathbb{R}^m)$, which embeds continuously into $L^q(\Omega, \mathbb{R}^m)$ for $q \leq \frac{2m}{m-2}$, and hence (19) is satisfied for any dimension of Ω.

Note that $h \to p(h)$ is affine and is given as the solution to

$$e(y, h) = f \qquad \text{and} \qquad e_y^\star p = -\mathcal{J}'(y). \tag{20}$$

As a generalized derivative for F we choose

$$G_F(u)h = \begin{cases} \alpha Ch + D^{-1}Ce_u^\star p'(h) - G_m(-D^{-1}Ce_u^\star p(u) - \alpha\psi)D^{-1}Ce_u^\star p'(h) \\ \alpha Ph + Pe_u^\star p'(h), \end{cases} \tag{21}$$

where $G_F \in \mathcal{L}(L^2(\hat{\Omega}; \mathbb{R}^m), L^2(\hat{\Omega}; \mathbb{R}^l) \times L^2(\hat{\Omega}; \mathbb{R}^m))$, and $p'(h)$ is the solution to

$$\begin{aligned} &e(v, h) = 0, \\ &e_y^\star p'(h) = -\,Qv. \end{aligned} \tag{22}$$

Utilizing (19) and a chain rule argument it is by now quite standard that $G_F(u)$ defines a Newton derivative for F on $L^2(\hat{\Omega};\mathbb{R}^m)$, Ito and Kunisch (2004).

To show local superlinear convergence of the semi-smooth Newton method it remains to argue that $G_F \in \mathcal{L}(L^2(\hat{\Omega},\mathbb{R}^m), L^2(\hat{\Omega},\mathbb{R}^l) \times L^2(\hat{\Omega},\mathbb{R}^m))$ admits a uniformly bounded inverse for all u in a neighborhood $U(u^*)$ of u^*.

This will imply the following theorem.

THEOREM 3 *Let Assumptions 1 and 2 hold and let $C : \mathbb{R}^m \to \mathbb{R}^l$ be surjective. Then the semi-smooth Newton method applied to $F(u) = 0$, with F given in* (17) *and generalized derivative $G_F(u)$ as in* (21) *converges locally superlinearly.*

Proof. It remains to analyze uniform bounded invertibility of $G_F(u)$. We define for $i = 1, \dots, l$

$$\tilde{\mathcal{A}}_i = \{x \in \hat{\Omega} : (-D^{-1}Ce_u^\star p(u) - \alpha\psi)_i(x) > 0\} \text{ and } \tilde{\mathcal{I}}_i = \hat{\Omega}\backslash\tilde{\mathcal{A}}_i.$$

and the diagonal matrix valued function $\chi_{\tilde{\mathcal{A}}} \in L^2(\hat{\Omega},\mathbb{R}^{m\times n})$ with

$$(\chi_{\tilde{\mathcal{A}}})_{i,i} = \chi_{\tilde{\mathcal{A}}i} \text{ for } i = 1, \dots, l, \text{ and } (\chi_{\tilde{\mathcal{A}}})_{i,j} = 0, \text{ for } i \neq j,$$

and analogously for $\chi_{\tilde{\mathcal{I}}}$. Then, for $u \in L^2(\hat{\Omega},\mathbb{R}^m)$ and $(f_1, f_2) \in L^2(\hat{\Omega},\mathbb{R}^l) \times L^2(\hat{\Omega}, ker(C))$ the equation

$$G_F(u)h = (f_1, f_2),$$

for $h \in L^2(\hat{\Omega},\mathbb{R}^m)$ can be expressed as

$$\begin{cases} \alpha\,\chi_{\tilde{\mathcal{A}}}\,Ch = \chi_{\tilde{\mathcal{A}}}\,f_1 \\ \alpha\,\chi_{\tilde{\mathcal{I}}}\,Ch + \chi_{\tilde{\mathcal{I}}}D^{-1}Ce_u^\star p'(h) = \chi_{\tilde{\mathcal{I}}}\,f_1 \\ \alpha\,Ph + Pe_u^\star p'(h) = f_2. \end{cases} \tag{23}$$

Observe that the dependence of this equation on u appears through $\tilde{\mathcal{A}}_i$ and $\tilde{\mathcal{I}}_i$.

To argue the existence of a solution h to (23) and continuous dependence on (f_1, f_2) we consider the following auxiliary problem

$$\begin{cases} \min\ \mathcal{J}_a(v,h) = \frac{1}{2}(v,Qv)_Y + \frac{\alpha}{2}|\chi_{\tilde{\mathcal{I}}}(C\,h - g_1)|^2_{L^2(\tilde{\mathcal{I}},\mathbb{R}^l)} + \frac{\alpha}{2}|Ph - g_2|^2_U \\ \text{subject to:} \\ e(v,h) = 0 \\ \chi_{\tilde{\mathcal{A}}}Ch = \chi_{\tilde{\mathcal{A}}}\,g_1 \end{cases} \tag{24}$$

where $(g_1, g_2) = (\frac{1}{\alpha}f_1, \frac{1}{\alpha}f_2) \in L^2(\hat{\Omega},\mathbb{R}^l) \times L^2(\hat{\Omega}, \ker(C))$, where $\ker(C) \subset \mathbb{R}^m$ and $\dim\ker(C) = m - l$. This is a linear quadratic optimization problem of the form

$$\begin{cases} \min\limits_{(v,h)\in Y\times U} \mathcal{J}_a(v,h) \\ \text{subject to:} \\ \mathcal{E}(v,h) = 0, \end{cases} \tag{25}$$

with equality constraint $\mathcal{E} : Y \times U \to Y' \times \bigotimes_{i=1}^{l} L^2(\tilde{\mathcal{A}}_i, \mathbb{R})$ given by

$$\mathcal{E}(v,h) = \begin{pmatrix} e(v,h) \\ \chi_{\tilde{\mathcal{A}}}(C\,h - \,g_1) \end{pmatrix}.$$

The kernel of the linearized equality constraints $\mathcal{E}'$ is given by

$$\ker(\mathcal{E}') = \{(v,h) \in Y \times U : \chi_{\tilde{\mathcal{A}}} Ch = 0, \quad e(v,h) = 0\}.$$

For the Hessian of $\mathcal{J}_a$ we find

$$\mathcal{J}_a''(\delta v, \delta h)^2 = (\delta v, Q\,\delta v)_Y + \alpha|\chi_{\tilde{\mathcal{I}}} C\,\delta h|^2_{L^2(\tilde{\mathcal{I}},\mathbb{R}^l)} + \alpha|P\,\delta h|_U^2$$

and hence for $(\delta v, \delta h) \in \ker(\mathcal{E}')$

$$\mathcal{J}_a''(\delta v, \delta h)^2 \geq \alpha \int_{\hat{\Omega}} |C\,\delta h|^2 + \alpha \int_{\hat{\Omega}} |P\,\delta h|^2.$$

Since C is surjective, there exists a constant $\bar{K} > 0$, such that

$$\mathcal{J}_a''(\delta v, \delta h)^2 \geq \alpha \bar{K} |\delta h|_U^2 \text{ for all } (\delta v, \delta h) \in \ker(\mathcal{E}') \tag{26}$$

independently of u. By Assumption 1, there exists $K \geq 1$ such that

$$|\delta v|_Y \leq \sqrt{K}\,|\delta h|_U$$

for all $(\delta v, \delta h)$ satisfying $e(\delta v, \delta h) = 0$. Combined with (26) this implies that

$$\mathcal{J}_a''(\delta v, \delta h)^2 \geq \frac{\alpha \bar{K}}{2K} |(\delta v, \delta h)|^2_{Y \times U} \text{ for all } (\delta v, \delta h) \in \ker(\mathcal{E}'), \tag{27}$$

where, without loss of generality, we assume that $K \geq 1$.

By Assumption 1 and surjectivity of C the linearization $\mathcal{E}' : Y \times U \to Y' \times \bigotimes_{i=1}^{l} L^2(\tilde{\mathcal{A}}_i, \mathbb{R})$ is surjective. These properties of $\mathcal{J}$ and $\mathcal{E}$ imply the existence of a unique solution $(v^*, h^*) \in Y \times U$ of (24) as well as of (p^*, μ^*), such that the Lagrangian

$$\mathcal{L} : Y \times U \times Y \times \bigotimes_{i=1}^{l} L^2(\tilde{\mathcal{A}}_i, \mathbb{R}),$$

given by

$$\mathcal{L}(v,h,p,\mu) = \mathcal{J}_a(v,h) + \langle p, e(v,h)\rangle_{Y,Y'} + (\mu, \chi_{\tilde{\mathcal{A}}}(C\,h - g_1))_{L^2(\tilde{\mathcal{A}},\mathbb{R}^l)}$$

with $L^2(\tilde{\mathcal{A}}, \mathbb{R}^l) := \bigotimes_{i=1}^{l} L^2(\tilde{\mathcal{A}}_i, \mathbb{R})$, is stationary at (v^*, h^*, p^*, μ^*), i.e.

$$\begin{cases} \mathcal{L}'(v^*, h^*, p^*, \mu^*) = 0 \\ \mathcal{E}(v^*, h^*) = 0, \end{cases} \tag{28}$$

where $\mathcal{L}'$ denotes the derivative of $\mathcal{L}$ with respect to (v, h). This is equivalent to

$$\begin{cases} e(v^*, h^*) = 0, \\ e_y^\star p = -\mathcal{J}_a'(v^*), \\ \chi_{\tilde{\mathcal{A}}}(C\,h^* - g_1) = 0, \\ \alpha C^T \chi_{\tilde{\mathcal{I}}}(C\,h^* - g_1) + \alpha(Ph^* - g_2) + e_u^\star p^* + C^T \chi_{\tilde{\mathcal{A}}}\,\mu^* = 0. \end{cases} \tag{29}$$

Projecting the last equation in (29) with respect to P and $I-P = C^T(C\,C^T)^{-1}C$, we obtain the equivalent system

$$\begin{cases} e(v^*, h^*) = 0, \\ e_y^\star p^* = -\mathcal{J}_a'(v^*) \\ \chi_{\tilde{\mathcal{A}}}(\,C\,h^* - g_1) = 0, \\ \alpha\chi_{\tilde{\mathcal{I}}}\,Ch^* + \chi_{\tilde{\mathcal{I}}}\,D^{-1}Ce_u^\star\,p^* = \alpha\chi_{\tilde{\mathcal{I}}}\,g_1 \\ \chi_{\tilde{\mathcal{A}}}\mu^* + \chi_{\tilde{\mathcal{A}}}D^{-1}Ce_u^\star p^* = 0 \\ \alpha\,Ph^* + Pe_u^\star p^* = \alpha\,g_2. \end{cases} \tag{30}$$

Using (22) this implies the existence of a solution to (23), with $(g_1, g_2) = (\frac{1}{\alpha}f_1, \frac{1}{\alpha}f_2)$. The solution (v^*, h^*, p^*, μ^*) to (30) depends linearly on (g_1, g_2). We next argue that this dependence is continuous, independently of $\tilde{\mathcal{A}}$, $\tilde{\mathcal{I}}$, i.e. independently of $u \in U$.

Continuous dependence of the solutions to

$$\begin{cases} \mathcal{L}'(v, h, p, \mu) = 0, \\ \mathcal{E}(v, h) = 0 \end{cases} \tag{31}$$

or equivalently of the solution to (30) with respect to $(g_1, g_2) \in L^2(\hat{\Omega}, \mathbb{R}^l) \times L^2(\hat{\Omega}, \ker(C))$ can be argued with techniques which are quite standard by now, see e.g. Alt (1983), Ito and Kunisch (1992). From these general results it is not immediate to conclude the required uniformity of the bound for the inverse of $G_F(u)$ with respect to u, however. For this reason and for the sake of completeness, we include the proof.

System (31) can be equivalently expressed as

$$\begin{cases} \mathcal{J}_a''(v, h) + (\mathcal{E}')^\star(p, \mu) = (0, \alpha\,C^T\,\chi_{\tilde{\mathcal{I}}}\,g_1 + \alpha\,g_2)^T_{Y' \times L^2(\hat{\Omega}, \mathbb{R}^m)} \\ \mathcal{E}'(v, h) = (0, \chi_{\tilde{\mathcal{A}}}\,g_1)^T. \end{cases} \tag{32}$$

Since neither $\mathcal{J}_a''$ nor $\mathcal{E}'$ depend on the point where the derivatives are taken, we only indicate the direction, in which $\mathcal{J}_a''$ and $\mathcal{E}'$ are evaluated. By Assumption 1 and surjectivity of C, the operator $\mathcal{E}' : Y \times U \to Y^* \times \bigotimes_{i=1}^m L^2(\tilde{\mathcal{A}}_i, \mathbb{R})$ given by

$$\mathcal{E}'(v, h) = \begin{pmatrix} e(v, h) \\ \chi_{\tilde{\mathcal{A}}} C\,h \end{pmatrix}$$

is surjective. Hence $range((\mathcal{E}')^\star)$ is closed and the decomposition

$$(v,h) = (v_k,h_k) + (v_r,h_r) \in ker(\mathcal{E}') \oplus range((\mathcal{E}')^\star)$$

is well defined. Throughout the remainder of the proof K_i denote constants, which are independent of $u \in U$, $\alpha > 0$ and $(g_1,g_2) \in L^2(\hat{\Omega},\mathbb{R}^l)\times L^2(\hat{\Omega}, ker\,(C))$. From the second equation in (32) we obtain that $\chi_{\mathcal{A}} C h_r = \chi_{\mathcal{A}} g_1$, which implies that $|h_r|_{L^2} \le K_2|g_1|$ (since $\chi_{\mathcal{A}} C$ is invertible on range$((\mathcal{E}')^\star)$). Also since (v_r,h_r) satisfies equation $e(v_r,h_r)=0$, we obtain from Assumption 1 that there exists K_1 such that

$$|(v_r,h_r)|_{Y\times U} \le K_1|g_1|_{L^2(\hat{\Omega},\mathbb{R}^l)}. \tag{33}$$

Since $\mathcal{E}'$ is surjective, $(\mathcal{E}')^\star$ is continuously invertible on its range. Hence the first equation in (32) implies that for some K_2

$$|(p,\mu)|_{Y\times L^2(\tilde{\mathcal{A}},\mathbb{R}^l)} \le K_2(|(v,\alpha h)|_{Y\times U} + \alpha|(g_1,g_2)|_{L^2(\hat{\Omega},\mathbb{R}^l)\times L^2(\hat{\Omega},ker\,(C))}). \tag{34}$$

Using (27) and (32) we find

$$\begin{aligned}
&\frac{\alpha\bar{K}}{2K}|(v_k,h_k)|^2_{Y\times U} \le \langle \mathcal{J}_a''(v_k,h_k),(v_k,h_k)\rangle \\
&= \langle \mathcal{J}_a''(v,h),(v,h)\rangle - 2\langle \mathcal{J}_a''(v_k,h_k),(v_r,h_r)\rangle - \langle \mathcal{J}_a''(v_r,h_r),(v_r,h_r)\rangle \\
&= \alpha(\chi_{\tilde{\mathcal{I}}}\, g_1, \chi_{\tilde{\mathcal{I}}}\, Ch)_{L^2(\hat{\Omega},\mathbb{R}^l)} + \alpha(g_2,h)_{L^2(\hat{\Omega},\mathbb{R}^m)} \\
&\quad -(\mu,\chi_{\tilde{\mathcal{A}}}\, g_1)_{L^2(\tilde{\mathcal{A}},\mathbb{R}^l)} - 2\langle \mathcal{J}_a''(v_k,h_k),(v_r,h_r)\rangle - \langle \mathcal{J}_a''(v_r,h_r),(v_r,h_r)\rangle \\
&\le \alpha(\chi_{\tilde{\mathcal{I}}}\, g_1, \chi_{\tilde{\mathcal{I}}}\, Ch)_{L^2(\hat{\Omega},\mathbb{R}^l)} + \alpha(g_2,h_k)_{L^2(\hat{\Omega},\mathbb{R}^m)} + \alpha(g_2,h_r)_{L^2(\hat{\Omega},\mathbb{R}^m)} \\
&\quad -(\mu,\chi_{\tilde{\mathcal{A}}} g_1)_{L^2(\tilde{\mathcal{A}},\mathbb{R}^l)} - 2\langle \mathcal{J}_a''(v_k,h_k),(v_r,h_r)\rangle.
\end{aligned} \tag{35}$$

From (33) and (34) we deduce that there exist K_3 and K_4 such that

$$(\mu,\chi_{\tilde{\mathcal{A}}}\, g_1)_{L^2(\tilde{\mathcal{A}},\mathbb{R}^l)} \le K_3\alpha(|g_1|^2 + |g_2|^2 + \frac{1}{\alpha^2}|g_1|^2) + \frac{\alpha\bar{K}}{8K}|h_k|^2$$

and

$$2\langle \mathcal{J}_a''(v_k,h_k),(v_r,h_r)\rangle \le K_4\alpha(|g_1|^2 + \frac{1}{\alpha^2}|g_1|^2) + \frac{\alpha\bar{K}}{8K}|v_k|^2.$$

These estimates imply

$$\begin{aligned}
&\frac{\alpha\bar{K}}{2K}|(v_k,h_k)|^2_{Y\times U} \le \alpha|C|_{\mathbb{R}^{l\times m}}\,|g_1|(|h_k|+|h_r|) + \alpha|g_2||h_k| \\
&\quad + \alpha|g_2||h_r| + \alpha(K_3+K_4)(|g_1|^2+|g_2|^2+\frac{1}{\alpha^2}|g_1|^2) + \frac{\alpha\bar{K}}{4K}|(v_k,h_k)|^2_{Y\times U}.
\end{aligned}$$

Taking also into account (33) we conclude that there exists K_5 such that

$$|(v_k, h_k)|^2_{Y\times U} \leq K_5(|g_1|^2 + |g_2|^2 + \frac{1}{\alpha^2}|g_1|^2),$$

and therefore

$$|(v, h)|_{Y\times U} \leq \frac{1}{\alpha} K_6(|f_1|_{L^2(\hat\Omega,\mathbb{R}^l)} + \frac{1}{\alpha}|f_1|_{L^2(\hat\Omega,\mathbb{R}^l)} + |f_2|_{L^2(\hat\Omega,\ker C)}).$$

This estimate implies the desired a-priori bound on the inverse of $G_F(u)$ uniformly. Moreover, this bound decreases as α increases. ∎

As announced in the introduction, we now turn to the semi-smooth Newton iteration applied to $F(u) = 0$ and explain its relationship to the primal-dual active set method.

Algorithm 1 (Semi-smooth Newton method)

1. *Initialize* u_0, *set* $k = 0$
2. *Solve* $G_F(u_k)\delta u_k = -F(u_k)$.
3. *Set* $u_{k+1} = u_k + \delta u_k$.
3. *Solve* $e(y, u_{k+1}) = f$ *for* y_{k+1}.
4. *Solve* $e^\star_y p = -\mathcal{J}'(y_{k+1})$ *for* $p_{k+1} = p(u_{k+1})$.
5. *Stop or set* $k = k + 1$, *goto 2.*

Note that $G_F(u_k)$ depends, besides $p(u_k)$, also on $p'(\delta u_k)$, which is the solution to

$$e(\delta y, \delta u_k) = 0, \qquad e^\star_y p'(\delta u_k) = -\mathcal{J}'(\delta y). \tag{36}$$

Setting

$$\mathcal{A}^k_i = \{x \in \hat\Omega : -(D^{-1}Ce^*_u p_k + \alpha\psi)_i > 0\},\ \mathcal{I}^k_i = \hat\Omega\backslash\mathcal{A}^k_i,$$

step 2. can equivalently be expressed as

$$\begin{cases} \alpha\chi_{\mathcal{A}^k} C\,\delta u_k = -\alpha\chi_{\mathcal{A}^k}(Cu_k - \psi), \\ \alpha\chi_{\mathcal{I}^k} C\,\delta u_k + \chi_{\mathcal{I}^k} D^{-1}Ce^\star_u p'(\delta u_k) = -\alpha\chi_{\mathcal{I}^k} Cu_k - \chi_{\mathcal{I}^k} D^{-1}Ce^\star_u p(u_k) \\ \alpha P\delta u_k + Pe^\star_u p'(\delta u_k) = -\alpha P_{u_k} - Pe^\star_u p(u_k). \end{cases} \tag{37}$$

Utilizing (36) and $y_{k+1} = y_k + \delta y_k$, $p_{k+1} = p_k + p'(\delta u_k)$ the Newton step (37) can equivalently be expressed as

$$\begin{cases} e(y_{k+1}, u_{k+1}) = f, \quad e^\star_y p_{k+1} = -\mathcal{I}_Y \mathcal{J}'(y_{k+1}), \\ \chi_{\mathcal{A}^k} Cu_{k+1} = \chi_{\mathcal{A}^k}\psi, \\ \alpha\chi_{\mathcal{I}^k} Cu_{k+1} + \chi_{\mathcal{I}^k} D^{-1}Ce^\star_u p_{k+1} = 0 \\ \alpha Pu_{k+1} + Pe^\star_u p_{k+1} = 0. \end{cases} \tag{38}$$

Note that (38) is the necessary and sufficient optimality system for

$$\begin{cases} \min \ J(y) + \frac{\alpha}{2}\,|Cu|^2_{L^2(\hat\Omega,\mathbb{R}^l)} + \frac{\alpha}{2}\,|Pu|^2_U \\ \text{subject to} \\ e(y,u) = f \\ \chi_{A_k} Cu = \chi_{A_k}\psi \quad \text{a.e.} \end{cases} \tag{39}$$

If we introduce a Lagrange multiplier $\lambda_k \in L^2(\mathcal{A}_k, \mathbb{R}^l)$ associated to the constraint $\chi_{A_k} Cu = \chi_{A_k}\psi$, then line 3 of the optimality system can be replaced by

$$\alpha Cu_{k+1} + D^{-1}Ce_u^{\star} p_{k+1} + \chi_{\mathcal{A}^k}\lambda_{k+1} = 0. \tag{40}$$

Let us extend $(\lambda_{k+1})_i$ by 0 on $(\mathcal{I}_k)_i$. Then

$$\lambda_{k+1} + \alpha(Cu_{k+1} - \psi) = -D^{-1}Ce_u^{\star} p_{k+1} - \alpha\psi.$$

The expression on the right hand side determines the active/inactive sets in (38). The expression on the left is used in the primal-dual active set method, see Hintermüller, Ito and Kunisch (2002), with α replaced by any $c > 0$, where the active sets are determined on the basis of

$$\hat{\mathcal{A}}_i^k = \{x \in \hat\Omega : (\lambda_{k+1} + \alpha(Cu_{k+1} - \psi))_i > 0\}.$$

We have just seen that $\hat{\mathcal{A}}_i^k$ and $\mathcal{A}_i^k$ coincide for the choice $\alpha = c$, and it can be argued as in Bergounioux, Ito and Kunisch (1999) that this is also the case for arbitrary $c \neq \alpha$, after the initialization phase, with $k \geq 1$.

Algorithm 2 (Primal dual-active set strategy)

1. *Initialize* p_0 *and set* $k = 0$.
2. *Set* $\mathcal{A}_k^i = \{x : -(D^{-1}Ce_u^{\star} p_k + \alpha\psi)_i(x) > 0\}$ *and* $\mathcal{I}_k^i = \hat\Omega \backslash \mathcal{A}_k^i$.
3. *Solve* (39) *(or equivalently* (38)*) for* $(y_{k+1}, u_{k+1}, p_{k+1})$.
4. *Update* $\lambda_{k+1} = \begin{cases} 0 & \text{on } \mathcal{I}_k \\ -\alpha\psi - D^{-1}Ce_u^{\star}\, p_{k+1} & \text{on } \mathcal{A}_k \end{cases}$.
5. *Stop or set* $k = k + 1$*, goto 2.*

The $\lambda-$update is not explicitly needed for Algorithm 2 but it will be convenient for later reference. The initialization p_0 need not correspond to the adjoint state at u_0. If it does, the Algorithms 1 and 2 coincide.

4. Global convergence

In this section, we give a sufficient condition for the convergence of Algorithm 2, respectively Algorithm 1 from arbitrary initial data.

THEOREM 4 *Suppose that Assumption 1 holds and that C is surjective. If α is sufficiently large, then the iterates (y_k, u_k, p_k) converge in $Y \times U \times Y$ to $(y(u^*), u^*, p(u^*))$ as $k \to \infty$, for arbitrary initialization $p_0 \in Y$.*

Proof. We can follow a procedure developed in Ito and Kunisch (2004), adapted to the present situation. Let $k \geq 1$. For $v \in L^2(\hat{\Omega}, \mathbb{R}^l)$, $w \in L^2(\hat{\Omega}, \mathbb{R}^l)$ we use the notation

$$v = w \text{ on } \mathcal{A}^{k-1} \cap \mathcal{A}^k$$

if $v_i(x) = w_i(x)$ for all $x \in \mathcal{A}_i^{k-1} \cap \mathcal{A}_i^k$, $i = 1, \dots, l$, and analogously for other combinations of active and inactive sets. For two consecutive iterates of controls we have

$$Cu_{k+1} - Cu_k = \begin{cases} 0 & \text{on } \mathcal{A}^{k-1} \cap \mathcal{A}^k \\ \psi - Cu_k & \text{on } \mathcal{I}^{k-1} \cap \mathcal{A}^k \\ -\frac{1}{\alpha} D^{-1} C e_u^\star (p_{k+1} - p_k) + \frac{1}{\alpha} \lambda_k & \text{on } \mathcal{A}^{k-1} \cap \mathcal{I}^k \\ -\frac{1}{\alpha} D^{-1} C e_u^\star (p_{k+1} - p_k) & \text{on } \mathcal{I}^{k-1} \cap \mathcal{I}^k, \end{cases}$$

where, as above $p_k = p(u_k)$. Therefore

$$Cu_{k+1} - Cu_k = -\frac{1}{\alpha} \chi_{\mathcal{I}_k} D^{-1} C e_u^\star (p_{k+1} - p_k) + R_k \tag{41}$$

where

$$R_k = \begin{cases} 0 & \text{on } \mathcal{A}^{k-1} \cap \mathcal{A}^k \\ \psi - Cu_k \leq 0 & \text{on } \mathcal{I}^{k-1} \cap \mathcal{A}^k \\ \frac{1}{\alpha} \lambda_k \leq 0 & \text{on } \mathcal{A}^{k-1} \cap \mathcal{I}^k \\ 0 & \text{on } \mathcal{I}^{k-1} \cap \mathcal{I}^k. \end{cases} \tag{42}$$

We next estimate two consecutive adjoint states:

$$|p'(u_{k+1} - u_k)|_Y \leq |e_u^\star|_{\mathcal{L}(Y,U)} |(e_y^\star)^{-1}|_{\mathcal{L}(Y^*,Y)} |Q|_{\mathcal{L}(Y)} |y_{k+1} - y_k|_Y,$$

where $y_k = y(u_k)$. Consequently

$$|p'(u_{k+1} - u_k)|_Y \leq K_1 |u_{k+1} - u_k|_U,$$

where

$$K_1 = |e_u|_{\mathcal{L}(U,Y^*)} |e_y^{-1}|^2_{\mathcal{L}(Y^*,Y)} |Q|_{\mathcal{L}(Y)}. \tag{43}$$

Hence there exists a constant $K_2 = K_2(K_1, C)$ such that

$$\begin{aligned} |p_{k+1} - p_k|_Y = |p'(u_{k+1} - u_k)|_Y &\leq K_2 \left(|Cu_{k+1} - Cu_k|_{L^2(\hat{\Omega},\mathbb{R}^l)} + |P(u_{k+1} - u_k)|_U \right) \\ &\leq \frac{K_2}{\alpha} |e_u|_{\mathcal{L}(U,Y^*)} (|D^{-1}C|_{\mathbb{R}^{l\times m}} + 1) |p_{k+1} - p_k|_Y + K_2 |R_k|_{L^2(\hat{\Omega},\mathbb{R}^l)}. \end{aligned}$$

Let $K_3 = K_2 |e_u|_{\mathcal{L}(U,Y^*)}(|D^{-1}C|_{\mathbb{R}^{l\times m}} + 1)$ and assume that

$$\alpha \geq 2\, K_3. \tag{44}$$

Then we have the estimate

$$|p'(u_{k+1} - u_k)|_Y \leq \frac{\alpha K_2}{\alpha - K_3} |R_k|_{L^2(\hat{\Omega},\mathbb{R}^l)} \leq 2K_2 |R_k|_{L^2(\hat{\Omega},\mathbb{R}^l)}. \tag{45}$$

Let us define $M : U \times L^2(\hat{\Omega}, \mathbb{R}^l) \to \mathbb{R}$ by

$$M(u, \lambda) = \alpha^2 \int_{\hat{\Omega}} |\max(0, Cu - \psi)|^2 dx + \int_{\hat{\Omega}} |\min(0, \lambda)|^2 dx.$$

By (42) and (45) we have

$$|p'(u_{k+1} - u_k)|_Y^2 \leq \frac{4K_2^2}{\alpha^2} M(u_k, \lambda_k). \tag{46}$$

To estimate $M(u_k, \lambda_k)$ note that on $\mathcal{A}^k$

$$\lambda_{k+1} = -D^{-1}Ce_u^\star p'(u_{k+1} - u_k) - D^{-1}Ce_u^\star p(u_k) - \alpha\psi$$

and hence by (38)

$$\lambda_{k+1} = -D^{-1}Ce_u^\star p'(u_{k+1} - u_k) + \begin{cases} \lambda_k > 0 & \text{on } \mathcal{A}^{k-1} \cap \mathcal{A}^k \\ \alpha(Cu_k - \psi) > 0 & \text{on } \mathcal{I}^{k-1} \cap \mathcal{A}^k. \end{cases}$$

Since $\lambda_{k+1}(x) = 0$ on $\mathcal{I}^k$ we find

$$\int_{\hat{\Omega}} |\min(0, \lambda_{k+1})|^2 dx \leq |D^{-1}C|^2_{\mathbb{R}^{l\times m}} |e_u|^2_{\mathcal{L}(U,Y^*)} |p'(u_{k+1} - u_k)|_Y^2. \tag{47}$$

Similarly on $\mathcal{I}^k$

$$Cu_{k+1} - \psi = -\frac{1}{\alpha} D^{-1}Ce_u^\star p'(u_{k+1} - u_k) - \frac{1}{\alpha} D^{-1}Ce_u^\star p_k - \psi$$

and thus

$$Cu_{k+1} - \psi = -\frac{1}{\alpha} D^{-1}Ce_u^\star p'(u_{k+1} - u_k) + \begin{cases} \frac{1}{\alpha}\lambda_k \leq 0 & \text{on } \mathcal{A}^{k-1} \cap \mathcal{I}^k \\ Cu_k - \psi \leq 0 & \text{on } \mathcal{I}^{k-1} \cap \mathcal{I}^k. \end{cases}$$

Since $Cu_{k+1} = \psi$ on $\mathcal{A}^k$ we have

$$\int_{\hat{\Omega}} |\max(0, Cu_{k+1} - \psi)|^2 dx \leq \frac{1}{\alpha^2} |D^{-1}C|^2_{\mathbb{R}^{l\times m}} |e_u|^2_{\mathcal{L}(U,Y^*)} |p'(u_{k+1} - u_k)|_Y^2,$$

and combined with (47)

$$M(u_{k+1},\lambda_{k+1}) \le 2|D^{-1}C|^2_{\mathbb{R}^{l\times m}}|e_u|^2_{\mathcal{L}(U,Y^*)}|p'(u_{k+1}-u_k)|^2_Y,$$

and (46)

$$M(u_{k+1},\lambda_{k+1}) \le \frac{\rho^2}{\alpha^2}M(u_k,\lambda_k), \qquad k=1,2,\dots \tag{48}$$

where $\rho = \sqrt{8}|D^{-1}C|_{\mathbb{R}^{l\times m}}|e_u|_{\mathcal{L}(U,Y^*)}K_2$. Then

$$M(u_{k+1},\lambda_{k+1}) \le \left(\frac{\rho}{\alpha}\right)^{2k} M(u_1,\lambda_1),$$

and hence $\{p_k\}_{k=1}^{\infty}$ is a Cauchy sequence in Y, provided that $\alpha > \rho$. Thus, there exists $p^* \in Y$ such that $\lim_{k\to\infty} p_k = p^*$ in Y and by the last equation in (38) there exists $u^*_{\ker} \in L^2(\hat{\Omega}, \ker C)$ such that

$$\lim_{k\to\infty} Pu_k = u^*_{\ker} \text{ in } U.$$

Since $\mathcal{A}^k_i = \{x : (-D^{-1}Ce^\star_u p_k - \alpha\psi)_i > 0\}$, $i=1,\dots,l$,

$$\lambda_{k+1} = \max(0, -D^{-1}Ce^\star_u p_k - \alpha\psi) + \chi_{\mathcal{A}^k}D^{-1}Ce^\star_u(p_{k+1}-p_k).$$

We conclude that $\{\lambda_k\}$ converges in $L^2(\hat{\Omega},\mathbb{R}^l)$ to some $\lambda^* \in L^2(\hat{\Omega},\mathbb{R}^l)$ and

$$\lambda^* = \max(0, -D^{-1}Ce^\star_u p^* - \alpha\psi). \tag{49}$$

Since $\alpha Cu_k = -D^{-1}Ce^\star_u p_k - \lambda_k$ for all k and since the expression on the right hand side converges in $L^2(\hat{\Omega},\mathbb{R}^l)$, there exists $u_{\ker^\perp} \in L^2(\hat{\Omega}, \ker C^\perp)$ such that

$$\lim_{k\to\infty}(I-P)u_k = u_{\ker^\perp}$$

and

$$\alpha Cu_{\ker^\perp} + D^{-1}Ce^\star_u p^* + \lambda^* = 0.$$

Moreover, $\lim_{k\to\infty} u_k = u_{\ker} + u_{\ker^\perp}$.

For $y^* = \lim_{k\to\infty} y(u_k)$ it follows that $(y^*, u^*, p^*, \lambda^*)$ is the unique solution to the optimality system (4)–(7) ■

5. Nonlinear case

In this section we consider some cases in which the operator $e : Y \times U \to Y'$ is not necessarily linear in both variables and turn to nonlinear operators of the form:

$$e(y,u) = e_1(y) + e_2 u, \tag{50}$$

with e_2 a compact linear operator from U to Y' and $e_1 : Y \to Y'$ satisfies the following conditions.

Assumption 3 *There exists a neighborhood $V(y^*)$ of the optimal state y^* such that:*

a) $e_1 : Y \to Y'$ *is twice Frechet differentiable in* $V(y^*)$.

b) $e_1'(y)$ *is continuously invertible for each* $y \in V(y^*)$.

c) e_1'' *is Lipschitz continuous in* $V(y^*)$, *i.e., there exists a constant* $L > 0$ *such that*

$$|e_1''(\bar{y}) - e_1''(y^*)|_{\mathcal{L}(Y\times Y,Y')} \leq L|\bar{y} - y^*|_Y, \quad \text{for } \bar{y} \in V(y^*). \tag{51}$$

These regularity requirements are needed for first and second order optimality conditions as well as for the convergence analysis of the semi-smooth Newton method.

Example 5 *Let* $\Omega \subset \mathbb{R}^m$, $m \leq 3$, *be a bounded domain. Consider the stationary Navier-Stokes equations*

$$\begin{aligned} -\nu\Delta y + (y\cdot\nabla)y + \nabla p = u \qquad & \text{in } \Omega \\ \operatorname{div} y = 0 \qquad & \text{in } \Omega \\ y = 0 \qquad & \text{on } \partial\Omega, \end{aligned}$$

with $(y\cdot\nabla)y = \sum_{i=1}^m y_i \partial_i y$, $Y = \{H_0^1(\Omega,\mathbb{R}^m) : \operatorname{div} y = 0\}$, $U = L^2(\Omega,\mathbb{R}^m)$ *and the remaining data as in Example 1. The operator* e_1 *is given by*

$$e_1 : Y \to Y' \tag{52}$$

$$y \mapsto \nu(\nabla y, \nabla\cdot)_U - ((y\cdot\nabla)y, \cdot)_U. \tag{53}$$

It can be easily verified that the operator e_1 *is twice Frechet differentiable with its first and second derivatives given by*

$$e_1'(y)w = \nu(\nabla w, \nabla\cdot)_U + ((w\cdot\nabla)y + (y\cdot\nabla)w, \cdot)_U \qquad \text{and} \qquad e_1''(y)[w]^2 = (2(w\cdot\nabla)w, \cdot)_U,$$

respectively. Condition (51) *follows immediately from the expression for the second derivative. To verify the surjectivity* $e_1'(y)$ *let us consider the linearized equation*

$$e_1'(y)w = \nu(\nabla w, \nabla\cdot)_U + ((w\cdot\nabla)y + (y\cdot\nabla)w, \cdot)_U = \langle g, \cdot\rangle_{Y',Y}, \tag{54}$$

with $g \in Y'$. *We assume that* ν *is sufficiently large so that*

$$\nu > \mathcal{M}(y^*) := \sup_{v\in Y} \frac{\left|((v\cdot\nabla y^*), v)_{\mathbf{L}^2(\Omega)}\right|}{|v|_Y^2}.$$

It can be argued that there exists a neighborhood $V(y^*) \subset Y$ *of* y^* *such that this inequality remains correct with* y^* *replaced by* $y \in V(y^*)$. *Then there exists a unique solution* w_g *to the linearized equation* (54) *associated with* g *for each* $y \in V(y^*)$. *From the bijectivity of* $e_1'(y)$ *the continuous invertibility follows. Summarizing, Assumption 3 holds for this problem.*

Similar to the linear case we assume that for every $u \in U$, there exists a locally unique $y = y(u) \in Y$ such that $e(y,u) = 0$. Moreover, we assume that the corresponding optimal control problem

$$\begin{cases} \min \ J(y) + \frac{\alpha}{2}\,|Cu|^2_{L^2(\hat{\Omega},\mathbb{R}^l)} + \frac{\alpha}{2}\,|Pu|^2_U \\ \text{subject to} \\ e_1(y) + e_2 u = 0 \\ Cu \le \psi \quad \text{a.e.}, \end{cases} \tag{55}$$

has a local solution $(y^*, u^*) \in Y \times U$. The differentiability of the control to state mapping in a neighborhood of the optimal solution follows from Assumption 3 and the implicit function theorem. From hypothesis b) in Assumption 3 the surjectivity of $e_1'(y)$ follows. Therefore a first order necessary condition for (55) is given by

$$(\alpha C^T C u^* + \alpha P u + e_u^\star p, u - u^*) \ge 0, \tag{56}$$

for all $u \in L^2(\hat{\Omega}, \mathbb{R}^m)$ satisfying $Cu \le \psi$, where

$$\begin{cases} e(y^*, u^*) = 0 \\ (e_1'(y^*))^\star p = -J'(y^*). \end{cases} \tag{57}$$

Proceeding as in the proof of Theorem 2, it can be shown that the first order necessary condition can be equivalently expressed as the following optimality system,

$$e(y^*, u^*) = 0 \tag{58}$$

$$(e_1'(y^*))^\star p = -J'(y^*) \tag{59}$$

$$\alpha C^T C u^* + \alpha P u^* + C^T \lambda + e_u^\star p = 0 \tag{60}$$

$$Cu^* \le \psi,\ \lambda \ge 0,\ \lambda^T (Cu^* - \psi) = 0 \text{ a.e. in } \hat{\Omega}. \tag{61}$$

For the subsequent analysis let us next introduce the Lagrangian

$$\mathcal{L}(y, u, p) = \mathcal{J}(y, u) + \langle p, e(y, u)\rangle_{Y,Y'}.$$

We consider the following cone of critical directions

$$K(u^*) = \left\{ v \in U : (C_j v)(x) \begin{cases} = 0 & \text{if } \lambda_j(x) \ne 0 \\ \le 0 & \text{if } (Cu^*)_j = \psi_j \text{ and } \lambda_j(x) = 0 \end{cases} \right\}.$$

A second order sufficient optimality condition is stated next. The result utilizes the critical cone $K(u^*)$, which does not involve strongly active constraints. Moreover, sufficient optimality is obtained without the use of a two-norm discrepancy argument. Rather a technique based solely on a second order optimality condition and the structure of the cost functional is used. The technique was previously applied in Casas, Mateos and Raymond (2007) to the optimal control of the Navier-Stokes equations with box constraints and in Casas, De Los Reyes and Tröltzsch (2008) to semilinear state constrained optimal control problems.

For some work concerning second order conditions for control problems with special kinds of control constraints we refer to Bonnans (1998), Dunn (1995). In the cited papers, constraints of the type $u(x) \in U$, with U independent of x and polygonal, are considered.

In Wachsmuth (2006) second order sufficient conditions for control problems with quadratic cost functionals and more general convex control constraints are studied. The result is based on the direct approach used in, e.g., Tröltzsch (2005), which includes strongly active constraints in the definition of the critical cone. Moreover, the result involves the classical two-norm discrepancy resulting from the residuum estimates.

In our case, due to the contradiction argument (see Casas et al., 2007, 2008), the strongly active constraints may be avoided. Additionally, the proof technique, together with the quadratic structure of the cost functional, allows us to obtain a sufficient optimality condition without the two-norm discrepancy. The complete proof of the following result is given in the Appendix.

THEOREM 5 *Suppose that Assumption 3 holds and let (y^*, u^*, p^*) be a solution of the necessary condition* (56)-(57). *Suppose that there exists a constant $\kappa > 0$ such that*

$$\alpha \int_{\hat{\Omega}} |C\,h|^2 + \alpha \int_{\hat{\Omega}} |P\,h|^2 + (v, Qv) + (p^*, e_1''(y^*)[v]^2) \geq \kappa\, |h|_U^2 \qquad \text{(SSC)}$$

holds for every pair $(v, h) \in Y \times K(u^*)$, $(v, h) \neq (0, 0)$ that solves the linearized equation

$$e_1'(y^*)v + e_u h = 0. \qquad (62)$$

Then there exist $\varepsilon > 0$ and $\delta > 0$ such that

$$J(y^*, u^*) + \frac{\delta}{2}|u - u^*|_U^2 \leq J(y, u),$$

for every feasible pair (y, u) such that $|u - u^*|_U \leq \varepsilon$.

Let us now turn to the analysis of semi-smooth Newton methods applied to (58)-(61). Let us hereafter assume that C is surjective. From the optimality system we obtain that

$$\lambda(x) = -\left(\alpha C u^* + D^{-1} C e_u^\star p\right)(x).$$

The system can then be expressed as the operator equation

$$F(u) = \begin{pmatrix} \alpha C u + D^{-1} C e_u^\star p(u) + \max(0, -D^{-1} C e_u^\star p(u) - \alpha\psi) \\ \alpha P u + P e_u^\star p(u) \end{pmatrix} = 0. \quad (63)$$

The generalized derivative of F at u is given by $G_F \in \mathcal{L}(L^2(\hat{\Omega};\mathbb{R}^m), L^2(\hat{\Omega};\mathbb{R}^l) \times L^2(\hat{\Omega};\mathbb{R}^m))$ with

$$G_F(u)h = \begin{pmatrix} \alpha Ch + D^{-1}Ce_u^\star p'(u;h) - G_m(-D^{-1}Ce_u^\star p(u) - \alpha\psi)D^{-1}Ce_u^\star p'(u;h) \\ \alpha Ph + Pe_u^* p'(u;h) \end{pmatrix}, \tag{64}$$

where $p'(u;h)$ is solution of

$$\begin{cases} e_1'(y)y' + e_u h = 0 \\ e_y^\star(y)p'(u;h) = -\mathcal{J}'(y') - ((e_1')^\star)'(p(u), y'). \end{cases} \tag{65}$$

For the subsequent analysis the following additional hypotheses are used:

$u \to e_u^* p(u)$ is Frechet differentiable from $L^2(\hat{\Omega};\mathbb{R}^m)$ to $L^q(\Omega;\mathbb{R}^n)$, for some $q > 2$. (H1)

$e_1'(y)^\star$ is uniformly continuously invertible in $V(y^*)$. (H2)

$$\begin{cases} \text{any solution } (v,h) \in Y \times U \text{ of the linearized equation} \\ e_1'(y)v + e_2 h = 0 \text{ satisfies, for } y \in V(y^*), \text{ the estimate} \\ |v|_Y \le \sqrt{K}|h|_U, \text{ with } K \text{ independent of } y. \end{cases} \tag{H3}$$

Hypothesis (H1) guarantees Newton differentiability of the operator equation (63).

The following stronger second order condition is also assumed to hold: there exists a constant $\kappa > 0$ such that

$$\alpha \int_{\hat{\Omega}} |C\,h|^2 + \alpha \int_{\hat{\Omega}} |P\,h|^2 + (v, Qv) + (p(u^*), e_1''(y^*)[v]^2) \ge \kappa |h|_U^2 \tag{SSC'}$$

holds for every pair $(v,h) \in Y \times U$ that solves the linearized equation (62).

THEOREM 6 *Let $C : \mathbb{R}^m \to \mathbb{R}^l$ be surjective and let Assumption 3,* (H1), (H2), (H3) *and* (SSC') *hold. Then the semi-smooth Newton method applied to $F(u) = 0$, with F given in* (63) *and generalized derivative $G_F(u)$ as in* (64) *converges locally superlinearly.*

Proof. The proof is given in the Appendix. ■

A complete semi-smooth Newton step for problem (55) is then given by the following algorithm.

ALGORITHM 3 (SEMI-SMOOTH NEWTON METHOD FOR NONLINEAR PROBLEMS)

1. *Initialize,* $u_0, k = 0$
2. *Solve* $G_F(u_k)\delta u_k = -F(u_k)$.
3. *Set* $u_{k+1} = u_k + \delta u_k$.
3. *Solve* $e(y, u_{k+1}) = f$ *for* y_{k+1}.
4. *Solve* $(e_y(y_k))^\star p = -\mathcal{J}'(y_{k+1})$ *for* $p_{k+1} = p(u_{k+1})$.
5. *Stop or set* $k = k + 1$, *goto 2.*

6. Numerical experiments

In this section we test the efficiency of Algorithm 2 for solving optimal control problems governed by the Stokes equations in the presence of different affine constraints. The domain $\Omega = (0,1)^2$ was discretized using a uniform triangular mesh. Boundary conditions of Dirichlet type were imposed. On the upper boundary the horizontal velocity takes the value one, while the vertical component is zero. On the remaining boundary the condition is of no slip type. This problem is referred to as "driven cavity flow".

For the numerical solution of the state and adjoint equations a finite element method was used. Taylor-Hood elements with quadratic basis functions for the velocity and linear functions for the pressure were employed on a uniform triangular mesh. For this type of elements, the following estimates are known to hold, Gunzburger (2000): if $y \in \mathbf{H}^m(\Omega) \cap \mathbf{H}_0^1(\Omega)$ and $p \in H^{m-1}(\Omega) \cap L_0^2(\Omega)$, $m = 2, 3$, then

$$|y - y_h|_{\mathbf{H}_0^1} = O(h^{m-1}), \qquad |y - y_h|_{\mathbf{L}^2} = O(h^m), \qquad |p - p_h|_{L^2} = O(h^{m-1}).$$

For the solution of the discretized systems appearing in each semi-smooth Newton step a penalty method was applied (see Gunzburger, 2000, p.125). This method considers, for $0 < \varepsilon << 1$, the modified Stokes system

$$\begin{pmatrix} A & B^T \\ B & \varepsilon I \end{pmatrix} \begin{pmatrix} \vec{Y} \\ \vec{P} \end{pmatrix} = \begin{pmatrix} F \\ 0 \end{pmatrix},$$

where A and B are the matrices resulting from the discretization of the Stokes equations, I is the identity matrix and $\vec{Y}, \vec{P}$ are the solution vectors for the velocity and pressure, respectively. A similar penalty scheme was used for the adjoint equations. For convergence results on this approach we refer to Gunzburger (2000).

The semi-smooth Newton algorithm stops if the L^2-residuum of the discretized control is lower than a given tolerance, typically set as 10^{-5}. The method is initialized setting the controls equal to 0 and solving successively the Stokes and the adjoint equations. With this values at hand, the active and inactive sets are determined for the first iteration.

We introduce the quantities

$$\varrho_k = |u_k - u_{k-1}|_{\mathbf{L}_h^2}, \qquad \vartheta_k = \frac{|u_k - u_{k-1}|_{\mathbf{L}_h^2}}{|u_{k-1} - u_{k-2}|_{\mathbf{L}_h^2}}$$

for the evaluation of the increment and the convergence rate, respectively. For the discrete cost functional evaluation, the mass matrix from the finite element discretization is used.

The resulting linear systems in each SSN iteration were solved using MATLAB exact solver. All algorithms were implemented in MATLAB 6.3 and run on a Pentium 5 machine with a precision of $eps = 2.2204e - 016$.

6.1. Example 1

First, we consider the optimal control problem (1) with

$$C = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \qquad \text{and} \qquad \psi = \begin{pmatrix} 0.05 \\ 0.02 \end{pmatrix}$$

as constraint matrix and vector respectively. The correspondent projection matrix is given by $P = 0$ and the desired state by $z_d \equiv 0$. The remaining parameter values are $\alpha = 0.01$ and $\varepsilon = \sqrt{eps}$.

With a mesh step size $h = 1/400$, the algorithm stops after 4 iterations. The sizes of the resulting active sets are 648 and 598 for the first and the second constraints, respectively. The constraints and the correspondent multipliers for the optimal solution are depicted in Fig. 1. From the graphics the complementarity condition can be verified by inspection.

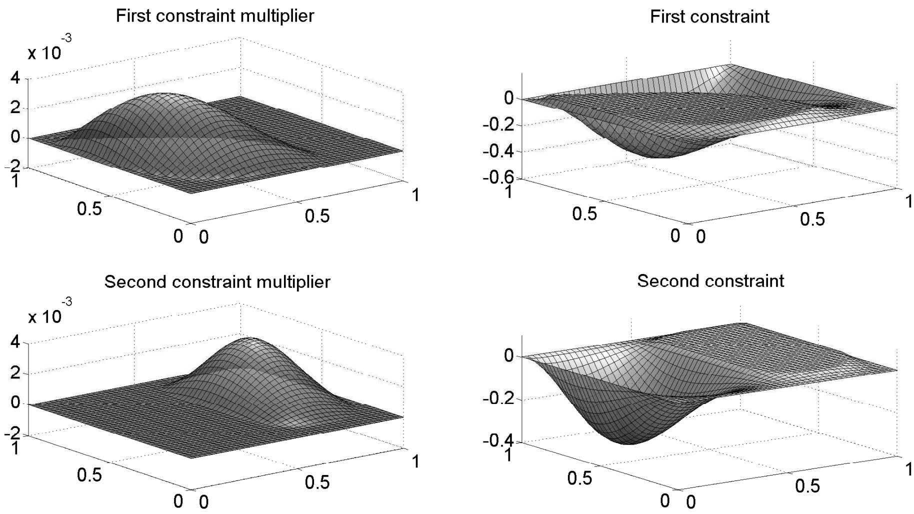

Figure 1. Polygonal constraints and their multipliers

In Table 1 the convergence history is documented. From the data, superlinear rate of convergence can be inferred. Also a monotonic decrease of the cost functional value can be observed.

The sensitivity of the semi-smooth Newton method with respect to changes in the matrix coefficients was also tested. We considered the parameter dependent matrix

$$C = \begin{pmatrix} -1 & 0 \\ 1 & \gamma \end{pmatrix}$$

Table 1. Example 1, $\alpha = 0.01$, 400 mesh points.

Iteration	$\mid \mathcal{A}_k^1 \mid$	$\mid \mathcal{A}_k^2 \mid$	$J(y,u)$	ϱ_k	ϑ_k
1	0	0	0.032628	-	-
2	883	755	0.032362	0.283736	-
3	647	598	0.032361	0.015507	0.0546537
4	648	598	0.032361	5.7025e-7	0.0000367

Table 2. Example 1, $\alpha = 0.01$, 225 mesh nodes.

γ	it. with step 2.	it. with step 2'.
1000	4	4
100	4	4
10	4	no convergence
1	4	no convergence
0.1	5	no convergence
0.01	8	no convergence
0.001	11	no convergence

and studied the behavior of the algorithm with respect to changes of the parameter γ. As $\gamma \to 0^+$ the opening angle of the cone of admissible directions tends to zero.

In the second column of Table 2 the number of iterations for different γ values is tabulated. As γ decreases a moderate increase of the number of iterations is required to achieve convergence.

In order to realize the importance of using matrix C in the control cost term, a semi-smooth Newton algorithm with the cost functional

$$\widehat{\mathcal{J}}(y,u) = J(y) + \frac{\alpha}{2} |u|^2_{L^2(\hat{\Omega},\mathbb{R}^m)}$$

was also implemented. In this case, the optimality condition is given by

$$\alpha u + C^T \lambda + e_u^\star p = 0,$$

which implies that

$$\alpha C u + D\lambda + C e_u^\star p = 0.$$

Consequently,

$$\lambda + \alpha(Cu - \psi) = -\alpha D^{-1} C u - D^{-1} C e_u^\star p + \alpha(Cu - \psi),$$

and, therefore, step 2. in Algorithm 2 has to be replaced by

2'. Set $\mathcal{A}_{k+1}^i = \{x : [-\alpha D^{-1}Cu_k - D^{-1}Ce_u^\star p_k + \alpha(Cu_u - \psi)]_i(x) > 0\}$ and $\mathcal{I}_{k+1}^i = \hat{\Omega}\backslash\mathcal{A}_{k+1}^i$.

Without C in the cost functional the algorithm diverges unless γ is large enough. This phenomenon can be noted from the last column in Table 2. Note that the large values of γ correspond to small values of the the angle between the two inequality constraints.

6.2. Example 2

In this example we consider the matrix $C = (1, 1)$ and the right hand side bound $\psi = 0.05$, which corresponds to the constraint $u_1 + u_2 \leq 0.05$. For this problem, the projection matrix is given by

$$P = \frac{1}{2}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}.$$

The constraint, evaluated at the optimal control, and its multiplier, are depicted in Fig. 2. The regularization parameter $\alpha = 0.01$ and the desired state $z_d \equiv 0$ were utilized in this case.

In Table 3 the number of iterations and the value of the cost functional for different values of α are presented. More iterations are needed as α becomes smaller. The mesh independence behavior of the method was also tested. From the data in Table 4 a mesh independent behavior of the method can be observed.

Table 3. Example 2, 225 mesh points.

α	it.	$J(y, u)$	Active points
0.1	4	0.0323742	0
0.01	4	0.0320779	288
0.001	4	0.0298277	440
0.0001	5	0.0225571	455
0.00001	7	0.0155221	439

Table 4. Example 2: number of iterations vs. mesh size; $\alpha = 0.01$.

$1/h$	5	10	15	20	25
it.	4	4	4	4	4

To verify the global convergence of the method, tests with different initial values were carried out. Apart from the initial control value $(u_1^0, u_2^0) = (0, 0)$ we initialized the algorithms with the control values $(0, 0.05)$, $(-1, -1)$ and $(-10, 10)$. The number of iterations for each initialization are recorded in Table 5. Convergence to the same solution is obtained for each initial value. Moreover, the numbers of iterations do not differ for these particular choices.

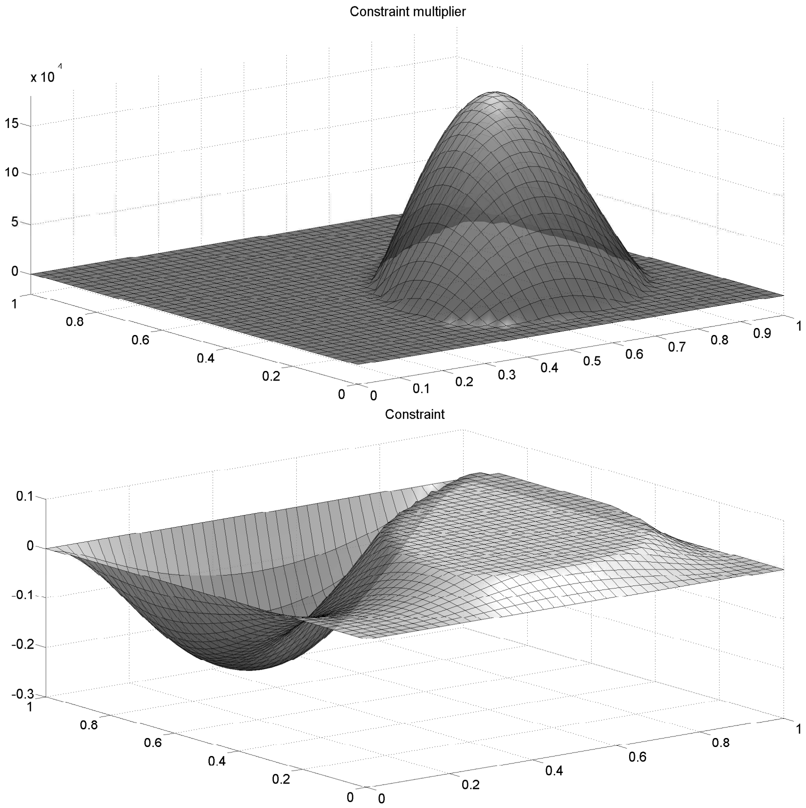

Figure 2. Example 2: affine constraint and its multipliers

Table 5. Example 2, number of iterations for different initial values, $h = 1/121$.

$u_0(x)$	(0,0)	(0,0.05)	(-1,-1)	(-10,10)
iter.	4	4	4	4

7. Appendix

Proof. (of **Theorem 5**) Let us suppose that u^* does not satisfy the quadratic growth condition. Then there exists a feasible sequence $\{u_k\}_{k=1}^{\infty} \subset U$ such that

$$|u_k - u^*|_U < \frac{1}{k^2} \tag{66}$$

and

$$J(y^*, u^*) + \frac{1}{k}|u_k - u^*|_U^2 > J(y_k, u_k) = \mathcal{L}(y_k, u_k, p^*) \quad \text{for all } k, \tag{67}$$

where y_k denotes the unique solution of (58) associated with u_k. As we define

$$\rho_k = |u_k - u^*|_U \quad \text{and} \quad h_k = \frac{1}{\rho_k}(u_k - u^*),$$

it follows that $|h_k|_U = 1$ and, therefore, we may extract a subsequence, denoted also by $\{h_k\}$, such that $h_k \rightharpoonup h$ weakly in U. The proof is now given in four steps.

Step 1: ($\frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h = 0$). From the mean value theorem it follows that

$$\begin{aligned} &\mathcal{L}(y_k, u_k, p^*) + \frac{\partial \mathcal{L}}{\partial y}(z_k, u_k, p^*)(y^* - y_k) = \mathcal{L}(y^*, u_k, p^*) \\ &= \mathcal{L}(y^*, u^*, p^*) + \rho_k \frac{\partial \mathcal{L}}{\partial u}(y^*, w_k, p^*)h_k, \end{aligned}$$

where w_k and z_k are points between u^* and u_k and y^* and y_k, respectively. By (67) it follows that

$$\frac{\partial \mathcal{L}}{\partial u}(y^*, w_k, p^*)h_k < \frac{1}{k}|u_k - u^*|_U + \frac{1}{\rho_k}\frac{\partial \mathcal{L}}{\partial y}(z_k, u_k, p^*)(y^* - y_k). \tag{68}$$

Working on the last term we obtain

$$\begin{aligned} &\frac{\partial \mathcal{L}}{\partial y}(z_k, u_k, p^*)\,(y^* - y_k) = J'(z_k)(y^* - y_k) + \langle p^*, e_y(z_k)(y^* - y_k)\rangle_{Y,Y'} \\ &= J'(z_k)(y^* - y_k) + \langle p^*, e_1'(y^*)(y^* - y_k)\rangle_{Y,Y'} \\ &\qquad\qquad + \langle p^*, e_1''(y^*)(z_k - y^*)(y^* - y_k)\rangle_{Y,Y'} \\ &\qquad\qquad + \langle p^*, (e_1''(\zeta_k) - e_1''(y^*))(z_k - y^*)(y^* - y_k)\rangle_{Y,Y'}, \end{aligned}$$

with $\zeta_k = y^* + \xi(z_k - y^*)$, for some $\xi \in [0, 1]$. From the optimality system and Assumption 3 we get that

$$\begin{aligned} &\left|\frac{\partial \mathcal{L}}{\partial y}(z_k, u_k, p^*)\,(y^* - y_k)\right| \le |J'(z_k) - J'(y^*)|_{Y'}|y^* - y_k|_Y \\ &\quad + |p^*|_Y |e_1''(y^*)|_{\mathcal{L}(Y\times Y, Y')}|z_k - y^*|_Y |y^* - y_k|_Y + L|p^*|_Y |z_k - y^*|_Y^2 |y^* - y_k|_Y. \end{aligned}$$

Due to the quadratic nature of J and since $h_k \rightharpoonup h$ weakly in U, $w_k \to u^*$ in U and $y_k \to y^*$ in Y, we obtain from (68) that

$$\frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h = \lim_{k\to\infty} \frac{\partial \mathcal{L}}{\partial u}(y^*, w_k, p^*)h_k \le 0. \tag{69}$$

On the other hand, we know that $Cu_k(x) \le \psi(x)$ a.e. in Ω, which implies that

$$\frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h_k = \rho_k \frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)(u_k - u^*) \ge 0, \tag{70}$$

and consequently

$$\frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h = \lim_{k\to\infty} \frac{\partial \mathcal{L}}{\partial u}(y^*, w_k, p^*)h_k \geq 0.$$

Altogether we obtain that

$$\frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h = 0. \tag{71}$$

Step 2: $(h \in K(u^*))$. The set

$$\{v \in U : (C_j v)(x) \leq 0, \text{ if } (C_j u^*) = \psi_j, \ \lambda_j(x) = 0, \ j = 1, \ldots, l\}$$

is closed and convex and, therefore, it is weakly sequentially closed. Since each h_k belongs to this set, then h also does. From the optimality condition, it follows that $-\lambda_j(x) C_j h(x) \geq 0$ for all j, a.e. in Ω, which implies that

$$\begin{aligned} 0 = \frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h &= (\alpha C^* C u^* + e_u^* p^*, h)_U \\ &= -\sum_{i=1}^{l} \int_\Omega \lambda_j(x) \ C_j h(x)\, dx = \sum_{i=1}^{l} \int_\Omega |\lambda_j(x) \ C_j h(x)|\, dx. \end{aligned}$$

Consequently, $C_j h(x) = 0$ if $\lambda_j(x) \neq 0$ and, therefore, $h \in K(u^*)$.

Step 3: $(h = 0)$. From condition (SSC) it suffices to show that

$$\begin{aligned} &\frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*)h + \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*)v \\ &\qquad = \alpha \int_{\hat{\Omega}} |C\, h|^2 + \alpha \int_{\hat{\Omega}} |P\, h|^2 + (v, Qv) + (p^*, e_1''(y^*)[v]^2) \leq 0. \end{aligned} \tag{72}$$

Using a Taylor expansion of the Lagrangian we get that

$$\begin{aligned} \mathcal{L}(y_k, u_k, p^*) = \mathcal{L}(y^*, u^*, p^*) &+ \rho_k \frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h_k \\ &+ \frac{\rho_k^2}{2} \frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*)h_k^2 + \frac{1}{2} \frac{\partial^2 \mathcal{L}}{\partial y^2}(z_k, u^*, p^*)(y_k - y^*)^2, \end{aligned} \tag{73}$$

with z_k an intermediate point between y^* and y_k. We therefore get that

$$\begin{aligned} \rho_k \frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*)h_k &+ \frac{\rho_k^2}{2} \frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*)h_k^2 + \frac{\rho_k^2}{2} \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(\frac{y_k - y^*}{\rho_k} \right)^2 \\ &= \mathcal{L}(y_k, u_k, p^*) - \mathcal{L}(y^*, u^*, p^*) \\ &\quad + \frac{\rho_k^2}{2} \left[\frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) - \frac{\partial^2 \mathcal{L}}{\partial y^2}(z_k, u^*, p^*) \right] \left(\frac{y_k - y^*}{\rho_k} \right)^2. \end{aligned} \tag{74}$$

Additionally, by (67),

$$\mathcal{L}(y_k, u_k, p^*) - \mathcal{L}(y^*, u^*, p^*) \leq \frac{\rho_k^2}{k}. \tag{75}$$

Since $u_k \to u^*$ in U and $|h_k|_U = 1$, we obtain from (51) that

$$\left| \left[\frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) - \frac{\partial^2 \mathcal{L}}{\partial y^2}(z_k, u^*, p^*) \right] \left(\frac{y_k - y^*}{\rho_k} \right)^2 \right|$$
$$\leq |p^*|_Y |e_1''(y^*) - e_1''(y_k)|_{\mathcal{L}(Y^2, Y')} \left| \frac{y_k - y^*}{\rho_k} \right|^2 \to 0 \quad \text{when} \quad k \to \infty. \tag{76}$$

For the latter we used the fact that, due to the differentiability of the control to state mapping, $\left| \frac{y_k - y^*}{\rho_k} \right|_Y$ is bounded.

Consequently by (74),

$$\lim_{k\to\infty} \inf \frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*) h_k^2 + \lim_{k\to\infty} \inf \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(\frac{y_k - y^*}{\rho_k} \right)^2$$
$$\leq 2 \lim_{k\to\infty} \sup \frac{1}{\rho_k^2} (\mathcal{L}(y_k, u_k, p^*) - \mathcal{L}(y^*, u^*, p^*)) - 2 \lim_{k\to\infty} \inf \frac{1}{\rho_k} \frac{\partial \mathcal{L}}{\partial u}(y^*, u^*, p^*) h_k,$$

which implies, since $\frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*)$ is w.l.s.c. and thanks to (70), (75), that

$$\frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*) h^2 + \lim_{k\to\infty} \inf \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(\frac{y_k - y^*}{\rho_k} \right)^2 \leq 2 \lim \frac{1}{k} = 0.$$

Additionally,

$$\frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(\frac{y_k - y^*}{\rho_k} \right)^2 = \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(\frac{y_k - y^*}{\rho_k} - v_{h_k} \right)^2$$
$$+ 2 \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(\frac{y_k - y^*}{\rho_k} - v_{h_k}, v_{h_k} \right) + \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) \left(v_{h_k} \right)^2,$$

where v_{h_k} is the solution to (62) associated to h_k, which also corresponds to the derivative of the control-to-state mapping at u^* in direction h_k. Due to the differentiability of this mapping, continuity of the bilinear form $\frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*)$, and since $v_{h_k} \to v_h$ strongly in Y (by the compactness of e_2), we obtain that

$$\frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*) h^2 + \frac{\partial^2 \mathcal{L}}{\partial y^2}(y^*, u^*, p^*) v_h^2 \leq 2 \lim \frac{1}{k} = 0.$$

Since $h \in K(u^*)$, it follows by (SSC) that $(v_h, h) = (0, 0)$.

Step 4: ($h_k \to 0$ strongly in U). Since $h_k \rightharpoonup 0$ weakly in U, it follows that $v_{h_k} \to 0$ strongly in Y. By (70), (74), (75), (76) and due to (SSC), we obtain that

$$\kappa \lim_{k\to\infty} \sup |h_k|_U^2 \leq \lim_{k\to\infty} \sup \frac{\partial^2 \mathcal{L}}{\partial u^2}(y^*, u^*, p^*) h_k^2 \leq 2 \lim \frac{1}{k} = 0.$$

Thus, h_k converges to 0 strongly. Since $|h_k|_U = 1$, a contradiction is obtained. ■

Proof. (of **Theorem 6**) We need to verify the assumption of Proposition 1. Newton differentiability of F follows from (H1), Newton differentiability of the max-function from $L^q(\hat{\Omega})$ to $L^2(\hat{\Omega})$ and the chain rule for Newton-differentiable functions, see Ito and Kunisch (2004). It remains to argue uniform boundedness of the inverse of the generalized derivative.

We proceed as in Theorem 3 and only indicate the necessary changes. We have to analyze the equation

$$G_F(u)h = (f_1, f_2),$$

for $(f_1, f_2) \in L^2(\hat{\Omega}; \mathbb{R}^l) \times L^2(\hat{\Omega}; \ker C)$ and $h \in L^2(\hat{\Omega}; \mathbb{R}^m)$, which can also be written as

$$\begin{cases} \alpha\, \chi_{\tilde{\mathcal{A}}}\, Ch = \chi_{\tilde{\mathcal{A}}}\, f_1 \\ \alpha\, \chi_{\tilde{\mathcal{I}}}\, Ch + \chi_{\tilde{\mathcal{I}}} D^{-1} C e_u^{\star} p'(u; h) = \chi_{\tilde{\mathcal{I}}}\, f_1 \\ \alpha\, Ph + P e_u^{\star} p'(u; h) = f_2. \end{cases} \tag{77}$$

To argue existence we consider the auxiliary problem:

$$\begin{cases} \min\ \mathcal{J}_a(v, h) = \frac{1}{2}(v, Qv)_Y + \frac{\alpha}{2}|\chi_{\tilde{\mathcal{I}}}(C\,h - g_1)|^2_{L^2(\tilde{\mathcal{I}}, \mathbb{R}^l)} \\ \qquad\qquad + \frac{\alpha}{2}|Ph - g_2|_U^2 + \frac{1}{2}\langle p, e_1''(y)[v]^2\rangle_{Y,Y'} \\ \text{subject to:} \\ e_1'(y)v + e_u h = 0 \\ \chi_{\tilde{\mathcal{A}}} Ch = \chi_{\tilde{\mathcal{A}}}\, g_1 \end{cases} \tag{78}$$

with $(g_1, g_2) = (\frac{1}{\alpha} f_1, \frac{1}{\alpha} f_2)$.

We again define $\mathcal{E} : Y \times U \to Y' \times \bigotimes_{i=1}^m L^2(\tilde{\mathcal{A}}_i, \mathbb{R})$ by

$$\mathcal{E}(v, h) = \begin{pmatrix} e_1'(y)v + e_2 h \\ \chi_{\tilde{\mathcal{A}}}(C\,h - g_1) \end{pmatrix},$$

with $\ker(\mathcal{E}') = \{(v, h) \in Y \times U : \chi_{\tilde{\mathcal{A}}} Ch = 0, \quad e_1'(y)v + e_2 h = 0\}$.

The Hessian of $\mathcal{J}_a$ is given by

$$\mathcal{J}_a''(\delta u, \delta h)^2 = (\delta v, Q\,\delta v)_Y + \alpha|\chi_{\tilde{\mathcal{I}}} C\,\delta h|^2_{L^2(\tilde{\mathcal{I}}, \mathbb{R}^l)} + \alpha|P\,\delta h|_U^2 + \langle p, e_1''(y)[\delta v]^2\rangle_{Y,Y'}.$$

For $(\delta v, \delta h) \in \ker(\mathcal{E}')$ we therefore obtain that

$$\mathcal{J}_a''(\delta v, \delta h)^2 \geq \alpha \int_{\hat{\Omega}} |C\,\delta h|^2 + \alpha \int_{\hat{\Omega}} |P\,\delta h|^2 + (\delta v, Q\delta v) + \langle p, e_1''(y)[\delta v]^2 \rangle_{Y,Y'},$$

which by (SSC') and the Lipschitz continuity of e_1'' implies the existence of a constant $\bar{K} > 0$, independent of u, such that

$$\mathcal{J}_a''(\delta v, \delta h)^2 \geq \alpha \bar{K} |\delta h|_U^2 \text{ for all } (\delta v, \delta h) \in \ker(\mathcal{E}') \tag{79}$$

in a neighborhood of u^*. Additionally, due to (H2) we obtain that

$$\mathcal{J}_a''(\delta v, \delta h)^2 \geq \frac{\alpha \bar{K}}{2K} |(\delta v, \delta h)|_{Y \times U}^2 \text{ for all } (\delta v, \delta h) \in \ker(\mathcal{E}'), \tag{80}$$

The auxiliary problem is therefore a linear quadratic optimization problem with convex objective function and, consequently, there exists a unique solution to (77).

Moreover, since $\mathcal{E}'(y)$ is surjective, there exist multipliers (q, φ) such that the Lagrangian

$$\mathcal{L}(v, h, q, \varphi) = \mathcal{J}_a(v, h) + \langle q, e_y(y, u)v + e_u h \rangle_{Y,Y'} + (\mu, \chi_{\tilde{\mathcal{A}}}(Ch - g_1)). \tag{81}$$

is stationary at (v, h, q, φ), i.e.,

$$\begin{cases} e_1'(y)v + e_u h = 0, \\ (e_1'(y))^\star q = -Qv - ((e_1'(y))^\star)'(p, v) \\ \chi_{\tilde{\mathcal{A}}}(\,C\,h^* - g_1) = 0, \\ \alpha \chi_{\tilde{\mathcal{I}}}\, Ch^* + \chi_{\tilde{\mathcal{I}}}\, D^{-1} C e_u^\star\, q = \alpha \chi_{\tilde{\mathcal{I}}}\, g_1 \\ \chi_{\tilde{\mathcal{A}}} \varphi + \chi_{\tilde{\mathcal{A}}} D^{-1} C e_u^\star q = 0 \\ \alpha\, Ph + P e_u^\star q = \alpha\, g_2. \end{cases} \tag{82}$$

In particular, this implies the solvability of (77).

The bounded invertibility analysis is again based on (31) and (32). In the latter, $\mathcal{J}_a''$ now depends on (y, u) through the additional term $\langle ((e_1'(y))^\star)'(p(u), v_r),$ $v_r \rangle_{Y',Y}$ which then appears on the right hand side of (35). By Assumption 5.1 and (H2) there exists a neighborhood $\hat{U}(u^*)$ of u^* and $\hat{K}$ such that

$$|\langle ((e_1'(y))^\star)'(p(u), v_r), v_r \rangle_{Y^*,Y}| \leq \hat{K} |v_r|_Y^2$$

for all $u \in \hat{U}(u^*)$, where $y = y(u)$ and $p = p(y(u))$. ■

References

Alt, W. (1983) Lipschitzian perturbations of infinite optimization problems. In: A. V. Fiacco, ed., *Mathematical Programming with Data Perturbations III*. Marcel Dekker, New York, 7–12.

Bergounioux, M., Ito, K. and Kunisch, K. (1999) Primal-dual strategy for constrained optimal control problems. *SIAM Journal on Control and Optimization* **37** (4), 1176–1194. http://link.aip.org/link/?SJC/37/1176/1.

Bonnans, F. (1998) Second–order analysis for control constrained optimal control problems of semilinear elliptic systems. *Applied Mathematics and Optimization* **38**, 303–325.

Casas, E. and Tröltzsch, F. (2002) Second-order necessary and sufficient optimality conditions for optimization problems and applications to control theory. *SIAM Journal on Optimization* **13** (2), 406–431. http://link.aip.org/link/?SJE/13/406/1.

Casas, E., Mateos, M. and Raymond, J.P. (2007) Error estimates for the numerical approximation of a distributed control problem for the steady-state Navier–Stokes equations. *SIAM Journal on Control and Optimization* **46** (3), 952–982. http://link.aip.org/link/?SJC/46/952/1.

Casas, E., De Los Reyes, J.C. and Tröltzsch, F. (2008) Sufficient second–order optimality conditions for semilinear control problems with pointwise state constraints. *SIAM Journal on Optimization* **19** (2), 616–643.

De Los Reyes, J.C. (2006) Primal-dual active set method for control constrained optimal control of the Stokes equations. *Optimization Methods and Software* **21** (2), 267–293.

De Los Reyes, J.C. and Kunisch, K. (2005) A semi-smooth Newton method for control constrained boundary optimal control of the Navier-Stokes equations. *Nonlinear Analysis. Theory, Methods & Applications* **62** (7), 1289–1316.

Dunn, J.C. (1995) Second-order optimality conditions in sets of l^∞ functions with range in a polyhedron. *SIAM Journal on Control and Optimization* **33** (5), 1603–1635. http://link.aip.org/link/?SJC/33/1603/1.

Griesse,R. and Volkwein, S. (2005) A primal-dual active set strategy for optimal boundary control of a nonlinear reaction-diffusion system. *SIAM Journal on Control and Optimization* **44** (2), 467–494. http://link.aip.org/link/?SJC/44/467/1.

Gunzburger, M. (2000) Navier-Stokes equations for incompressible flows: finite-element methods. In: *Handbook of Computational Fluid Mechanics*, Academic Press, San Diego, 99–158.

Hintermüller, M. and Hinze, M. (2006) An SQP semi-smooth Newton-type algorithm applied to the instationary Navier-Stokes system subject to control constraints. *SIAM Journal on Optimization* **16** (4), 1177–1200.

Hintermüller, M., Ito, K. and Kunisch, K. (2002) The primal-dual active set strategy as a semismooth Newton method. *SIAM Journal on Optimization* **13** (3), 865–888. http://link.aip.org/link/?SJE/13/865/1.

Ito, K. and Kunisch, K. (1992) Sensitivity analysis of solutions to optimization problems in Hilbert spaces with applications to optimal control and estimation. *J. Differential Equations* **97**.

Ito, K. and Kunisch, K. (2004) The primal-dual active set method for nonlinear optimal control problems with bilateral constraints. *SIAM Journal on Control and Optimization* **43** (1), 357–376. http://link.aip.org/link/?SJC/43/357/1.

Kunisch, K. and Rösch, A. (2002) Primal-dual active set strategy for a general class of constrained optimal control problems. *SIAM Journal on Optimization* **13** (2), 321–334. http://link.aip.org/link/?SJE/13/321/1.

Meyer, C., Philip, P. and Tröltzsch, F. (2006) Optimal control of a semilinear pde with nonlocal radiation interface conditions. *SIAM Journal on Control and Optimization* **45** (2), 699–721. http://link.aip.org/link/?SJC/45/699/1.

Tröltzsch, F. (2005) *Optimale Steuerung partieller Differentialgleichungen.* Vieweg.

Wachsmuth, D. (2006) Sufficient second-order optimality conditions for convex control constraints. *Journal of Mathematical Analysis and Applications* **319** (1), 228–247.

Weiser, M. (2005) Lipschitzian perturbations of infin Interior point methods in function space. *SIAM Journal on Control and Optimization* **44** (5), 1766–1786. http://link.aip.org/link/?SJC/44/1766/1.