Hamburger Beiträge
zur Angewandten Mathematik

The semi-smooth Newton method for variationally discretized control constrained elliptic optimal control problems; implementation, convergence and globalization

Michael Hinze and Morten Vierling

Nr. 2009-15
November 2009; replaced in January 2012

The semi-smooth Newton method for variationally discretized control constrained elliptic optimal control problems; implementation, convergence and globalization

Michael Hinze*& Morten Vierling†

January 9, 2012

Abstract: Combining the numerical concept of variational discretization introduced in [11, 12] and semi-smooth Newton methods for the numerical solution of pde constrained optimization with control constraints [9, 22] we place special emphasis on the implementation and globalization of the numerical algorithm. We prove fast local convergence of a globalized algorithm and illustrate our analytical and algorithmical findings by numerical experiments.

Mathematics Subject Classification (2010): 49J20, 49K20, 49M15

Keywords: Variational discretization, semi-smooth Newton method, primal-dual active set strategy, Elliptic optimal control problem, control constraints, error estimates.

1 Introduction and Mathematical Setting

Let $\Omega \subset \mathbb{R}^d\,(d \geq 1)$ denote an open, bounded sufficiently smooth (or convex polyhedral) domain with boundary Γ, and let $Y \subseteq H^1(\Omega)$ denote a closed subspace. Given some Hilbert space U and some closed, convex subset $U_{ad} \subset U$ of admissible controls together with and a linear, continuous control operator $B : U \to Y^*$, we are interested in the numerical treatment of the following control problem

$$(\mathbb{P}) \quad \begin{cases} \min_{(y,u)\in Y\times U_{ad}} J(y,u) := \frac{1}{2}\|y-z\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_U \\ \text{s.t.} \\ a(y,v) = \langle Bu, v\rangle_{Y^*,Y} \text{ for all } v \in Y. \end{cases} \tag{1.1}$$

Here, $a : Y \times Y \to \mathbb{R}$ denotes a continuous, coercive bilinear form associated with a boundary value problem on Ω for an uniformly elliptic operator A, i.e. we assume that a satisfies the

*Schwerpunkt Optimierung und Approximation, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany.

†Schwerpunkt Optimierung und Approximation, Universität Hamburg, Bundesstraße 55, 20146 Hamburg, Germany.

suppositions of the Lax-Milgram theorem. Furthermore, $z \in L^2(\Omega)$ denotes the given desired state.
Typical configurations of $\mathbb{P}$ are

Examples.

(i) (Finite dimensional control space) $Y := H_0^1(\Omega)$, $a(y,v) := \int\limits_\Omega \nabla y \nabla v dx$, $U := \mathbb{R}^m$, $B : \mathbb{R}^m \to H^{-1}(\Omega)$, $Bu := \sum\limits_{j=1}^m u_j F_j$, $F_j \in H^{-1}(\Omega)$, $U_{\text{ad}} := \{v \in \mathbb{R}^m; a_j \leq v_j \leq b_j\}$, $a, b \in \mathbb{R}^m$, $a < b$.

(ii) (Distributed control in $L^2(\Omega)$) $Y := H_0^1(\Omega)$, $a(y,v) := \int\limits_\Omega \nabla y \nabla v dx$, $U := L^2(\Omega)$, $B = \imath : L^2(\Omega) \to H^{-1}(\Omega)$, $\imath$ being the canonical injection, $U_{\text{ad}} := \{v \in L^2(\Omega); a \leq v \leq b\}$, $a, b \in L^\infty(\Omega)$, $a < b$.

(iii) (Robin boundary control in $L^2(\Gamma)$) $Y := H^1(\Omega)$, $a(y,v) := \int\limits_\Omega \nabla y \nabla v dx + \delta \int\limits_\Gamma yv \mathrm{d}\Gamma$ ($\delta > 0$), $U := L^2(\Gamma)$, $B : L^2(\Gamma) \to Y^*$, $\langle Bu, v\rangle_{Y^*,Y} = \int_\Gamma uv \, \mathrm{d}\Gamma$, $U_{\text{ad}} := \{v \in L^2(\Gamma); a \leq v \leq b\}$, $a, b \in L^\infty(\Gamma)$, $a < b$.

(iv) (Distributed control in $H^1(\Omega)$) $Y := H_0^1(\Omega)$, $a(y,v) := \int\limits_\Omega \nabla y \nabla v dx$, $U := H^1(\Omega)$, $B = \imath : H^1(\Omega) \to H^{-1}(\Omega)$, $\imath$ being the canonical injection, $U_{\text{ad}} := \{v \in H^1(\Omega); a \leq v \leq b\}$, $a, b \in L^\infty(\Omega)$, $a < b$.

Problem $\mathbb{P}$ admits a unique solution $(y,u) \in Y \times U_{\text{ad}}$, and can equivalently be rewritten as the optimization problem

$$\min_{u \in U_{\text{ad}}} \hat{J}(u)$$

for the reduced functional $\hat{J}(u) := J(SBu, u)$ over the set U_{ad}, where $S : Y^* \to Y$ denotes the solution operator associated with the bilinear form a. The first order necessary (and here also sufficient) optimality conditions take the form

$$\langle \hat{J}'(u), v - u\rangle_{U^*,U} \geq 0 \text{ for all } v \in U_{\text{ad}} \tag{1.2}$$

where $\hat{J}'(u) = \alpha(\cdot,u)_U + B^*S^*(SBu - z) \equiv \alpha(\cdot,u)_U + B^*p$, with $p := S^*(SBu - z)$ denoting the adjoint variable. The function p in our setting satisfies

$$a(w,p) = (y - z, w)_{L^2} \text{ for all } w \in Y. \tag{1.3}$$

For the numerical treatment of problem (1.1) it is convenient to rewrite (1.2) for $\sigma > 0$ arbitrary in form of the following non–smooth operator equation;

$$u = P_{U_{\text{ad}}}\left(u - \sigma \nabla \hat{J}(u)\right) \overset{\sigma=1/\alpha}{\equiv} P_{U_{\text{ad}}}\left(-\frac{1}{\alpha} R^{-1} B^* p\right),$$

with $R : U \to U^*$ denoting the Riesz isomorphism, and $\nabla \hat{J}(u) = R^{-1}\hat{J}'(u)$ denoting the gradient of $\hat{J}(u)$.

2 Finite Element Discretization

To discretize ($\mathbb{P}$) we concentrate on Finite Element approaches and make the following assumptions.

Assumption 2.1.
$\Omega \subset \mathbb{R}^d$ denotes a convex polyhedral domain with an admissible quasi-uniform sequence of partitions $\{T_{h_i}\}_{i=1}^{\infty}$, $h_i \to 0$ and $T_h := \{T_j^h\}_{j=1}^{n_h}$. We abbreviate $T_h := \{T_j\}_{j=1}^{n_h}$.

For $k \in \mathbb{N}$ we set

$$W_h := \{v \in C^0(\bar{\Omega}); v_{|T_j} \in \mathbb{P}_k(T_j) \text{ for all } 1 \le j \le n_h\}, \text{ and}$$
$$Y_h := Y \cap W_h =: \langle \phi_1, \dots, \phi_n \rangle \subseteq Y,$$

with some $n \in \mathbb{N}$. The resulting Ansatz for y_h then is of the form $y_h = \sum_{i=1}^{n} y_i \phi_i$. Following [11, 12] we approximate problem ($\mathbb{P}$) by

$$(\mathbb{P}_h) \quad \begin{cases} \min_{(y_h,u)\in Y_h \times U_{\text{ad}}} J(y_h,u) := \frac{1}{2}\|y_h - z\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_U \\ \text{s.t.} \\ a(y_h, v_h) = \langle Bu, v_h \rangle_{Y^*,Y} \text{ for all } v_h \in Y_h. \end{cases} \tag{2.1}$$

Problem ($\mathbb{P}_h$) admits a unique solution $(y_h, u) \in Y_h \times U_{\text{ad}}$ and, as above, can equivalently be rewritten as the optimization problem

$$\min_{u \in U_{\text{ad}}} \hat{J}_h(u)$$

for the discrete reduced functional $\hat{J}_h(u) := J(S_h B u, u)$ over the set U_{ad}, where $S_h : Y^* \to Y_h \subset Y$ denotes the discrete solution operator associated with the equality constraint in $\mathbb{P}_h$. The first order necessary (and here also sufficient) optimality conditions take the form

$$\langle \hat{J}_h'(u_h), v - u_h \rangle_{U^*,U} \ge 0 \text{ for all } v \in U_{\text{ad}}$$

where $\hat{J}_h'(v) = \alpha(\cdot, v)_U + B^* S_h^*(S_h B v - z) \equiv \alpha(\cdot, v)_U + B^* p_h$, with $p_h := S_h^*(S_h B v - z) \in Y_h$ denoting the adjoint variable. The function p_h in our setting satisfies

$$a(w_h, p_h) = (y_h - z, w_h)_{L^2} \text{ for all } w_h \in Y_h. \tag{2.2}$$

Analogously to (1.2), for $\sigma > 0$ arbitrary, we have

$$u_h = P_{U_{\text{ad}}}\left(u_h - \sigma \nabla \hat{J}_h(u_h)\right) \overset{\sigma=1/\alpha}{\equiv} P_{U_{\text{ad}}}\left(-\frac{1}{\alpha} R^{-1} B^* p_h\right). \tag{2.3}$$

Problem (2.1) is infinite–dimensional, if the control space U and the feasible subset U_{ad} are infinite–dimensional. The numerical challenge now consists in designing numerical solution algorithms for problem (2.1) which are implementable, and which reflect the (infinite–dimensional) structure of the *discrete* problem (2.1). In the following we investigate the semi–smooth Newton method for the setting of Example (ii). Our considerations are also applicable to the settings in Example (i) and (iii). In these three settings the control space is either

finite–dimensional, or a L^2–space, so that these settings have in common, that the orthogonal projection $P_{U_{\text{ad}}}$ for pointwise bounds is easy to evaluate, and that the Riesz isomorphism R is simple. This is different for Example iv.). The numerical evaluation of R^{-1} there requires the solution of an elliptic PDE, and the numerical evaluation of $P_{U_{\text{ad}}}\left(-\frac{1}{\alpha}R^{-1}B^*p_h\right)$ in (2.3) the numerical solution of an obstacle problem. However, we emphasize that the discretization of R^{-1} and of the obstacle problem need not to be coupled with the finite element discretization in problem ($\mathbb{P}$). This freedom in choice is caused by the concept of variational discretization. For a further discussion we refer to e.g. [11, 12, 14], [16].
The author's are not aware of any contributions in the literature to semi-smooth Newton methods applied to the setting of Example iv.). A major ingredient in this context would be appropriate differentiability properties of the solution operator associated with an obstacle problem. Here we refer to the work of Mignot and Puel [17] and the references cited there.

3 Semi-Smooth Newton Algorithm

In the following we restrict our considerations to the case of Example (ii), which requires the implementation of variational discretization in $\mathbb{R}^d$. The setting of Example (i) leads to the semi-smooth Newton method in finite dimensions, the setting of Example (iii) requires the implementation of variational discretization on the $d-1$–dimensional boundary Γ of Ω, see [13] for an application of the method.
Let us recall that now $U = L^2(\Omega)$, $Y = H_0^1(\Omega)$, $U_{ad} = \{v \in L^2(\Omega);\, a \le v \le b\}$ with $a, b \in L^\infty(\Omega)$, $b - a \ge \delta > 0$ which we require from now onwards. The control operator B is the injection $\imath : L^2(\Omega) \to Y^*$, hence the adjoint $B^* = \imath^*$ is the injection from Y into $L^2(\Omega)$. Below, the operators $\imath$, $\imath^*$ and R are omitted for notational convenience.
The exact application of the semi–smooth Newton algorithm proposed in the following requires bounds $a, b \in Y_h$ which we assume w.l.o.g..
Using the notation $P_{[a,b]}(x) = \max(\min(x, b), a)$ for $x \in \mathbb{R}$ we set

$$G(v) := v - P_{[a,b]}\left(-\frac{1}{\alpha}p(v)\right), \text{ and } G_h(v) := v - P_{[a,b]}\left(-\frac{1}{\alpha}p_h(v)\right),$$

where for given $v \in L^2(\Omega)$ the functions p, p_h are defined through (1.3) and (2.2), respectively and $P_{U_{\text{ad}}}(v)(\omega) = P_{[a,b]}(v(\omega))$ for $\omega \in \Omega$. It follows from the characterization of orthogonal projectors in real Hilbert spaces that the unique solutions u, u_h to (1.1) and (2.1) are characterized by the equations

$$G(u),\ G_h(u_h) = 0 \text{ in } L^2(\Omega). \tag{3.1}$$

These equations are amenable to semi–smooth Newton methods as proposed in [9] and [22], see Section 3.1. We begin with formulating

Algorithm 3.1. (Semi–smooth Newton algorithm for (3.1))

Start with $v \in L^2(\Omega)$ given. Do until convergence

Choose $M \in \partial G_h(v)$.

Solve $M\delta v = -G_h(v)$, $v := v + \delta v$.

If we choose Jacobians $M \in \partial G_h(v)$ with $\|M^{-1}\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))}$ uniformly bounded throughout the iteration, and at the solution u_h of $(\mathbb{P}_h)$ the function G_h is ∂G_h-semismooth of order μ, this algorithm is locally superconvergent of order $1+\mu$, see [22]. Although Algorithm 3.1 works on the infinite dimensional space $L^2(\Omega)$, it is possible to implement it numerically, as is shown subsequently.

3.1 Smoothness

Following [10], under mild assumptions it can be shown that the family $\{G_h\}_{h\geq 0}$ is mesh-indepently semismooth with generalized differential

$$\partial G_h(v) = I + \partial P_{[a,b]}\big(-\frac{1}{\alpha}p_h(v)\big)\frac{1}{\alpha}S_h^*S_h \ni I + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h\,.$$

The set $\partial P_{[a,b]}\big(-\frac{1}{\alpha}p_h(v)\big) \subset L^2(\Omega)$ consists of the (measurable) pointwise evaluations of generalized differentials of the projection $P_{[a(\omega),b(\omega)]} : \mathbb{R} \to [a(\omega), b(\omega)]$ and contains the indicator function $\mathbb{1}_{\mathcal{I}(p_h^v)}$ of the inactive set

$$\mathcal{I}\left(p_h^v\right) = \big\{\omega\in\Omega \,\big|\, \big(-\frac{1}{\alpha}p_h(v)\big)(\omega) \in \big(a(\omega), b(\omega)\big)\big\}\,, \quad \mathbb{1}_{\mathcal{I}(p_h^v)}(\omega) = \begin{cases} 1 & \omega\in\mathcal{I}\left(p_h^v\right) \\ 0 & \omega\in\Omega\setminus\mathcal{I}\left(p_h^v\right)\end{cases}.$$

By $\mathbb{1}_{p_h(v)}$ we denote the self-adjoint endomorphism in $L^2(\Omega)$ given by the pointwise multiplication with $\mathbb{1}_{\mathcal{I}(p_h^v)}$.
The results of [10] are generalized in [23] and [24]. In our situation [24, Thm. 6.23] implies the following lemma.

Lemma 3.2. *Let u and u_h denote the solutions of $(\mathbb{P})$ and $(\mathbb{P}_h)$, respectively. Let further $\gamma\in(0,1]$ and $s_0 > 0$ and suppose that*

$$\operatorname{meas}\left(\left\{\omega\in\Omega :\ 0 \leq |-\frac{1}{\alpha}p(u)(\omega) - a| < s \vee 0 \leq |-\frac{1}{\alpha}p(u)(\omega)-b| < s\right\}\right) \leq Cs^\gamma$$

holds for $s\in[0,s_0)$. Then there exist $h_0>0$, $\delta>0$ and $C>0$, such that for $0<h\leq h_0$

$$\begin{aligned}\|G_h(v) - G_h(u_h) - M_v(v-u_h)\|_{L^2(\Omega)} &\leq \dots \\ &\leq C\max\left(\frac{1}{\alpha}\|p_h(u_h)-p(u)\|_\infty, \|v-u_h\|_{L^2(\Omega)},\right)^{\frac{1}{2}}\|v-u_h\|_{L^2(\Omega)}\,,\end{aligned}$$

holds for all $v\in L^2(\Omega)$ with $\|v-u_h\|_{L^2(\Omega)}<\delta$ and for all $M_v\in\partial G_h(v)$.

This means, that Algorithm 3.1 behaves like a semismooth Newton method of order $\frac{1}{2}$, as long as the discretization error $\|p_h(u_h)-p(u)\|_\infty$ does not dominate $\|v-u_h\|_{L^2(\Omega)}$. Upon reaching the point at which both errors are balanced one could stop the iteration, since the accuracy of $(\mathbb{P}_h)$ is reached.
Under reasonable assumptions on the desired state z and on the solution operator S_h one can show true mesh independent Fréchet differentiability.

Assumption 3.3. S_h satisfies the following stability and convergence properties

$$\|S_h v\|_{1,\infty} \leq C\|Sv\|_{1,\infty}\,, \quad \|(S_h - S)v\|_{1,\infty} \leq Ch\|Sv\|_{2,\infty}\,, \quad \|Sv\|_{W^{2,p}(\Omega)} \leq C\|v\|_{L^p(\Omega)}\,,$$

where v is chosen such that these quantities exist and $C > 0$ does not depend on h and v.

Assumption 3.3 is fulfilled in the situation of Example (ii), see [19] for the $W^{1,\infty}(\Omega)$-estimates.

Lemma 3.4. *Let u solve ($\mathbb{P}$) over $\Omega \subset \mathbb{R}^2$ open and bounded and with $a, b \in \mathbb{R}$. Further let Assumption 3.3 hold and let $p(u) \in C^2(\bar{\Omega})$ for the adjoint state $p(u)$ associated to u and $\|\frac{1}{\alpha}\nabla p(u)\| \geq 2\beta > 0$ on*

$$\mathcal{C}_u = \left\{\omega \in \Omega :\; -\frac{1}{\alpha}p(u)(\omega) = a \vee -\frac{1}{\alpha}p(u)(\omega) = b\right\}.$$

Then there exists $h_0, \delta > 0$, such that for $0 \leq h \leq h_0$ the operator $G_h : L^2(\Omega) \to L^2(\Omega)$ is uniformly strictly Fréchet differentiable with $\frac{1}{2}$-Hölder continuous derivative on the $L^2(\Omega)$-ball $B_\delta(u_h)$, i.e. there exists $C > 0$, independent of h such that for $u_1, u_2 \in B_\delta(u_h)$

$$\|G_h(u_1) - G_h(u_2) - \partial G_h(u_2)(u_1 - u_2)\|_{L^2(\Omega)} \leq C\|u_1 - u_2\|^{\frac{3}{2}}_{L^2(\Omega)}\,,$$

and

$$\|\partial G_h(u_1) - \partial G_h(u_2)\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))} \leq C\|S_h(u_1 - u_2)\|^{\frac{1}{2}}_{L^2(\Omega)} \tag{3.2}$$

holds.

The poof is contained in Appendix A. Note that it generalizes to higher order (globally continuous) finite element discretizations as well as higher dimensions. Note that the condition $p(u) \in C^2(\bar{\Omega})$ can bee satisfied by requiring $z \in C(\bar{\Omega})$ and a sufficiently smooth boundary $\partial\Omega$.
Without the additional assumptions of Lemma 3.2 and 3.4 it is possible to show that for any fixed $h > 0$ the operator G_h corresponding to the finite element solution operator S_h is semismooth. This is caused by the piecewise polynomial Ansatz in S_h. Strict complementarity need not hold in this case.

3.2 Implementation of the Newton-Algorithm

Choosing $M = \mathbb{1}_{p_h(v)} \in \partial G_h(v)$ in Algorithm 3.1, the Newton-step reads

$$\Big(I + \frac{1}{\alpha}\mathbb{1}_{p_h(v)} S_h^* S_h\Big)\delta v = -v + P_{[a,b]}\left(-\frac{1}{\alpha}p_h(v)\right). \tag{3.3}$$

The equation admits of a unique solution δv and the Jacobian has a bounded inverse, see Section 3.3. To obtain an impression of the structure of the next iterate $v^+ = v + \delta v$ we rewrite (3.3) as

$$v^+ = P_{[a,b]}\left(-\frac{1}{\alpha}p_h(v)\right) - \frac{1}{\alpha}\mathbb{1}_{p_h(v)} S_h^* S_h \delta v\,.$$

Since the range of S_h^* is Y_h, the first addend is continuous and piecewise polynomial (of degree k) on a refinement K_h of T_h. The partition K_h is obtained from T_h by inserting nodes and

edges along the boundary between the inactive set $\mathcal{I}(p_h^v)$ and the according active set. The inserted edges are zero-level sets of polynomials of order $\le k$ since we assume $a, b \in Y_h$. We note that the refinement K_h may develop complicated structures in the case $k \ge 2$. The case $k = 1$ is studied in detail in [12] and [14, Chap.3]. The case $k = 2$ is considered in [20] and also in [8] where piecewise quadratic polynomials in the context of a posteriori error estimates in the presence of control and state constraints.
The second addend, involving the cut-off function $\mathbb{1}_{p_h(v)}$, is also piecewise polynomial of degree k on K_h but may jump along the edges not contained in T_h.
Finally v^+ lies in the following finite dimensional subspace of $L^2(\Omega)$

$$Y_h^+ = \left\{ \mathbb{1}_{\mathcal{I}(p_h^v)}\varphi_1 + \mathbb{1}_{\hat{\mathcal{A}}(p_h^v)}\varphi_2 + \mathbb{1}_{\check{\mathcal{A}}(p_h^v)}\varphi_3 \,\middle|\, \varphi_1, \varphi_2, \varphi_3 \in Y_h \right\},$$

where

$$\check{\mathcal{A}}(p_h^v) = \left\{\omega \in \Omega \,\middle|\, \big(-\frac{1}{\alpha}p_h(v)\big)(\omega) \le a\right\} \text{ and } \hat{\mathcal{A}}(p_h^v) = \left\{\omega \in \Omega \,\middle|\, \big(-\frac{1}{\alpha}p_h(v)\big)(\omega) \ge b\right\}.$$

The iterates generated by the Newton-algorithm can be represented exactly with about constant effort, since the number of inserted nodes varies only mildly from step to step, once the algorithm begins to converge. Furthermore the number of inserted nodes and edges is bounded in terms of the polynomial degree k and the dimension of Y_h, see [11, 12, 14].
Since the Newton-increment δv may have jumps along the borders of both the new and the old active and inactive sets, it is advantageous to compute v^+ directly, because v^+ lies in Y_h^+. To achieve an equation for v^+ we add $G_h'(v)v$ on both sides of (3.3) to obtain

$$\Big(I + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h\Big)v^+ = P_{[a,b]}\left(-\frac{1}{\alpha}p_h(v)\right) + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h v\,, \tag{3.4}$$

and reformulate Algorithm 3.1 as

Algorithm 3.5 (Newton Algorithm)**.**

$v \in U$ given. Do until convergence

Solve (3.4) for v^+, $v := v^+$.

3.3 Computing the Newton-Step 3.4

Since v^+ defined by (3.4) is known on the active set $\mathcal{A}(p_h^v) := \check{\mathcal{A}}(p_h^v) \cup \hat{\mathcal{A}}(p_h^v)$ it remains to compute v^+ on the inactive set. So we rewrite (3.4) in terms of the unknown $\mathbb{1}_{p_h(v)}v^+$ by splitting v^+ as

$$v^+ = (1 - \mathbb{1}_{\mathcal{I}(p_h^v)})v^+ + \mathbb{1}_{\mathcal{I}(p_h^v)}v^+$$

and obtain

$$\Big(I+\frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h\Big)\mathbb{1}_{p_h(v)}v^+ = P_{[a,b]}\left(-\frac{1}{\alpha}p_h(v)\right)+\frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h v-\Big(I+\frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h\Big)(1-\mathbb{1}_{p_h(v)})v^+\,.$$

Since $(1 - \mathbb{1}_{p_h(v)})v^+$ is already known, we can restrict the latter equation to the inactive set $\mathcal{I}(p_h^v)$

$$\Big(\mathbb{1}_{p_h(v)} + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h\mathbb{1}_{p_h(v)}\Big)v^+ = \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*z - \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^*S_h(1 - \mathbb{1}_{p_h(v)})v^+\,. \tag{3.5}$$

On the left-hand side of (3.5) we have now a continuous, self-adjoint Operator on $L^2(\mathcal{I}(p_h^v))$, which is positive definite, because it is the restriction of the positive definite Operator $\left(I + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^* S_h \mathbb{1}_{p_h(v)}\right)$ to $L^2(\mathcal{I}(p_h^v))$.
Hence we are in the position to apply a CG-algorithm to solve (3.5). Moreover assuming that the first iterate is an element of

$$Y_h^+\big|_{\mathcal{I}(p_h^v)} = \left\{\mathbb{1}_{p_h(v)}\varphi \,\middle|\, \varphi \in Y_h\right\} ,$$

the algorithm does not leave $Y_h^+\big|_{\mathcal{I}(p_h^v)}$ because of the $\left(I + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^* S_h \mathbb{1}_{p_h(v)}\right)$-invariance of the subspace $Y_h^+\big|_{\mathcal{I}(p_h^v)}$.
These considerations lead to the following

Algorithm 3.6 (Solving (3.4))**.**

Compute the active and inactive sets $\mathcal{A}(p_h^v)$ and $\mathcal{I}(p_h^v)$.

$\forall q \in \mathcal{A}(p_h^v)$ set

$$v^+(q) = P_{[a,b]}\left(-\frac{1}{\alpha}p_h(v)(q)\right) .$$

Solve

$$\left(I + \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^* S_h\right)\mathbb{1}_{p_h(v)}v^+ = \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^* z - \frac{1}{\alpha}\mathbb{1}_{p_h(v)}S_h^* S_h(1 - \mathbb{1}_{p_h(v)})v^+$$

for $\mathbb{1}_{p_h(v)}v^+$ by the CG-algorithm.

$v^+ = (1 - \mathbb{1}_{p_h(v)})v^+ + \mathbb{1}_{p_h(v)}v^+$.

We note that the use of this procedure in Algorithm 3.5 coincides with the active set strategy for the solution of $(\mathbb{P}_h)$ proposed in [9].

4 Globalization

In order to formulate a globally convergent algorithm, we want to apply inexact Armijo line-search. For this purpose we construct a sufficiently smooth merit function in the present section.
To begin with, following [6, 7], we introduce a multiplier $w \in L^2(\Omega)$ associated with the equality constraints and reformulate Algorithm 3.5 by means of the Lagrange dual function $\phi : L^2(\Omega) \to \mathbb{R}$,

$$\phi(w) = -\inf_{u,y \in L^2(\Omega)} \left(\underbrace{\frac{1}{2}\|y - z_h\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \chi_{U_{ad}}(u) - \langle w, y - S_h u\rangle_{L^2(\Omega)}}_{\mathcal{L}(u,y,w)} \right), \quad (4.1)$$

where z_h is the $L^2(\Omega)$-orthogonal projection of z onto Y_h. By

$$\chi_{U_{ad}} = \begin{cases} 0 & \text{on } U_{ad} \\ \infty & \text{on } L^2(\Omega) \setminus U_{ad} \end{cases}$$

we denote the characteristic function of the set U_{ad} in the sense of convex analysis.
It turns out that ϕ is differentiable with Lipschitz continuous derivative and strongly convex.

Lemma 4.1 (Lagrange dual function)**.** *The function $\phi : L^2(\Omega) \to \mathbb{R}$ from* (4.1) *is strongly convex and Fréchet differentiable with Lipschitz continuous $L^2(\Omega)$-gradient*

$$\nabla\phi(w) = y(w) - S_h u(w)\,,$$

where $y(w) = w + z_h$ and $u(w) = P_{[a,b]}\left(-\frac{1}{\alpha}S_h^* w\right)$ are the unique minimizers of the Lagrange function $\mathcal{L}(u, y, w)$ from (4.1) *for any given $w \in L^2(\Omega)$.*

A proof of the lemma is included in the appendix, proceeding along the lines of the proof given in [7] for the finite dimensional setting. Note that all results from this section hold for $h \geq 0$ in the sense of $S_0 = S : L^2(\Omega) \to L^2(\Omega)$.
Observe that, while the gradient $\nabla\phi$ is semi-smooth, the differentiability assertion from Lemma 3.4 does not apply since the smoothing properties of S_h^* alone are not strong enough. Nevertheless we are now in the position to apply a semi-smooth Newton strategy to the dual problem

$$\min_{w \in L^2(\Omega)} \phi(w)\,. \tag{$\mathbb{P}'_h$}$$

Due to strong convexity, problem ($\mathbb{P}'_h$) admits a unique solution w^* sasisfying $\nabla\phi(w^*) = 0$. Hence $u(w^*)$ solves (2.3), compare Lemma 4.1; there is no duality gap.
A semi-smooth Newton step for ($\mathbb{P}'_h$) reads

$$\left(I + \frac{1}{\alpha}S_h \mathbb{1}_{S_h^* w} S_h^*\right)\delta w = -(w + z_h) + S_h P_{[a,b]}\left(-\frac{1}{\alpha}S_h^* w\right) = -\nabla\phi(w)\,, \tag{4.2}$$

which is obtained from Equation (3.3) by applying S_h to both sides, if we think of w as a function of the iterate v, i.e.

$$w = w(v) = S_h v - z_h\,, \tag{v-w}$$

and take into account that $S_h^* z = S_h^* z_h$. Thus both Newton iterations are equivalent because the next iterate v^+ in (3.4) only depends on the adjoint state $p_h(v) = S_h^* w(v)$.
Step (4.2) is implementable since beginning with the first iteration there holds $w^+ \in Y_h$. Observe that the generalized Hessian on the left hand side of (4.2) is a self-adjoint and positive definite endomorphism of $L^2(\Omega)$. Also ϕ is easy to evaluate, since by Lemma 4.1 we have

$$\phi(w) = \frac{1}{2}\|w\|^2_{L^2(\Omega)} - \frac{\alpha}{2}\|u(w)\|^2_{L^2(\Omega)} + \langle w, z_h - S_h u(w)\rangle_{L^2(\Omega)}\,.$$

Thus, with ϕ as merit function we obtain a globalization of Algorithm 3.5.

Algorithm 4.2 (Damped Newton-Algorithm)**.** $w := w_0 \in L^2(\Omega)$, $\beta \in (0, 1)$ given.
Do until convergence

i) Solve Equation (4.2) for δw. Set $\lambda := 1$.

ii) If $\phi(w + \lambda\delta w) > \phi(w) + \frac{1}{3}\lambda \underbrace{\langle\nabla\phi(w), \delta w\rangle_{L^2(\Omega)}}_{\leq 0 \text{ by } (4.2)}$, set $\lambda := \beta\,\lambda$ and return to ii).

iii) Set $w := w + \lambda\delta w$. Return to i).

As mentioned above, in the sense of (v-w) we can interpret Algorithm 4.2 as a Newton algorithm with respect to v, just like Algorithm 3.5, but working exclusively on $w(v)$.
Since ϕ is sufficiently smooth the number of damping steps in the line search algorithm is bounded, and Algorithm 4.2 converges for any initial value w_0.

Lemma 4.3 (Global convergence)**.** *Let $L = 1 + \frac{\|S_h\|^2}{\alpha}$ denote the Lipschitz constant of $\nabla\phi$ and $\beta \in (0,1)$ as in Algorithm 4.2. Let u_h and w^* denote the solutions of $(\mathbb{P}_h)$ and $(\mathbb{P}'_h)$, respectively. Then we have*

i) A step with damping parameter $\lambda \leq \beta^{K(L,\beta)}$ is always accepted, where

$$K(L,\beta) = \frac{\log(2) - \log(3L)}{\log\beta} .$$

ii) Hence Algorithm 4.2 converges for arbitrary initial data $w_0 \in L^2(\Omega)$, in the sense that $w \to w^*$, $u(w) \to u_h$ and $v \to u_h$ in $L^2(\Omega)$.

iii) The stopping criterion $\|\nabla\phi(w)\|_{L^2(\Omega)} \leq \mathrm{Tol}$ is reasonable since

$$\|\nabla\phi(w)\|_{L^2(\Omega)} \geq \|w - w^*\|_{L^2(\Omega)} .$$

Proof. i) By the mean value theorem one has $\Theta \in (0,1)$ such that

$$\begin{aligned}\phi(w + \lambda\delta w) &= \phi(w) + \lambda\langle\nabla\phi(w + \Theta\lambda\,\delta w), \delta w\rangle_{L^2(\Omega)} \\ &\leq \phi(w) + \lambda\left(\langle\nabla\phi(w), \delta w\rangle_{L^2(\Omega)} + \lambda L\|\delta w\|^2_{L^2(\Omega)}\right) .\end{aligned}$$

Using (4.2), standard arguments deliver that sufficient descent is achieved for $\beta^k L \leq \frac{2}{3}$.
ii)+iii) Denoting by λ_k the damping parameter generated in step ii) of Algorithm 4.2 with associated iterates w_k the descend condition reads

$$-\frac{1}{3}\lambda\langle\nabla\phi(w_k), \delta w_k\rangle_{L^2(\Omega)} \leq \phi(w_k) - \phi(w_{k+1})$$

Because ϕ is bounded from below by $\phi(w^*)$, summation over k yields

$$\phi(w_0) - \phi(w^*) \geq \phi(w_0) - \liminf_{k\to\infty}\phi(w_k) \geq -\sum_{k=1}^{\infty}\frac{\lambda_k}{3}\langle\nabla\phi(w_k), \delta w_k\rangle_{L^2(\Omega)} \geq \sum_{k=1}^{\infty}\frac{2\beta}{3L}\|\delta w_k\|^2_{L^2(\Omega)} ,$$

where we again use (4.2) in the last estimate. Hence $\delta w_k \to 0$ and thus $\nabla\phi(w_k) \to 0$.
Inserting $\nabla\phi(w^*) = 0$ into the strong convexity relation (A.6) we arrive at the estimate $\|\nabla\phi(w)\|_{L^2(\Omega)} \geq \|w - w^*\|_{L^2(\Omega)}$ and thus finally at $w \to w^*$. $u(w) \to u_h$ follows by continuity and the fact that $u_h = u(w^*)$. ∎

Under the conditions of Lemma 3.4 the damping steps generated by Algorithm 4.2 do not affect fast local convergence of the Newton iteration.

Lemma 4.4 (Transition to fast local convergence)**.** *Let u_h denote the solution of $(\mathbb{P}_h)$. In the situation of Lemma 3.4, given $\beta \in (0,1)$ there exists $\tilde{\delta} > 0$ such that the Algorithm 4.2, upon reaching the ball $B_{\tilde{\delta}}(u_h) \subset L^2(\Omega)$, proceeds with full Newton steps, i.e. with $\lambda = 1$.*

Proof. The function $v \mapsto \phi(w(v))$ is $C^{2,\frac{1}{2}}$-smooth in the ball $B_\delta(u_h)$ from Lemma 3.4. Hence, using Taylor expansion and (4.2) together with the improved Hölder continuity from (3.2) we get

$$\begin{aligned}\phi(S_h(v+\delta v)) &= \phi(S_h(v)) + \langle \nabla\phi(S_h v), S_h\delta v\rangle_{L^2(\Omega)} \dots \\ &\quad + \frac{1}{2}\langle S_h\delta v, \left(I + \frac{1}{\alpha}S_h \mathbb{1}_{p_h(v+\Theta\delta v)} S_h^*\right) S_h\delta v\rangle_{L^2(\Omega)} \\ &\le \phi(S_h(v)) - \frac{1}{2}\langle S_h\delta v, \left(I + \frac{1}{\alpha}S_h \mathbb{1}_{p_h(v)} S_h^*\right) S_h\delta v\rangle_{L^2(\Omega)} + C\|S_h\delta v\|^{\frac{5}{2}}_{L^2(\Omega)} .\end{aligned} \tag{4.3}$$

Upon reaching the neighborhood of u_h in which the undamped Newton iteration converges $\frac{1}{2}$-superlinearly there holds $\|\delta v\| \le 2\|v - u_h\|$. Lemma 4.3 ensures that this happens after finitely many steps, and for $\tilde{\delta} > 0$ sufficiently small $\|v - u_h\|_{L^2(\Omega)} \le \tilde{\delta}$ together with (4.3) implies

$$\phi(S_h(v+\delta v)) \le \phi(S_h(v)) - \frac{1}{3}\langle S_h\delta v, \left(I + \frac{1}{\alpha}S_h \mathbb{1}_{p_h(v)} S_h^*\right) S_h\delta v\rangle_{L^2(\Omega)} .$$

∎

If $\lambda = 1$ is always accepted after sufficiently many iterations, locally super-linear convergence occurs with respect to both v and $w = w(v)$. The mesh-independence result from Lemma 3.4 however only applies to v.

Note that one could compute the next full step iterate v^+ following a full step from w by solving (3.4) at any time during the iteration, whereas $v + \lambda\delta v$, $\lambda \neq 1$ is not so readily accessible.

Another possible globalization for our semi-smooth Newton method was described in [24, Alg. 7.27]. The trust-region algorithm proposed there uses a merit function based either on a Fisher-Burmeister function or on the objective J itself. For a number of other trust-region approaches that are not directly linked to our method we refer to the references in [24].

4.1 Global Convergence of the undamped Newton Algorithm

It is not difficult to see, that the fixed-point equation for problem $(\mathbb{P}_h)$

$$u_h = P_{[a,b]}\left(-\frac{1}{\alpha}S_h^*(S_h u_h - z)\right)$$

can be solved by simple fixed-point iteration that converges globally for $\alpha > \|S_h\|^2_{\mathcal{L}(L^2(\Omega),L^2(\Omega))}$, see [11, 12]. A similar global convergence result holds for the undamped Newton algorithm 3.1

Lemma 4.5. *The Newton algorithm 3.1 converges globally if* $\alpha > \frac{4}{3}\|S\|^2_{\mathcal{L}(L^2(\Omega),L^2(\Omega))}$.

Proof. See [25]. ∎

Similar results are published in [1].

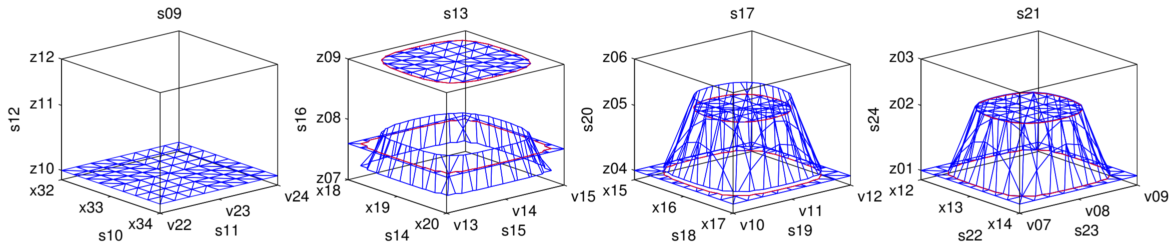

Figure 1: The first four Newton-iterates for Example 5.1 (Dirichlet) with parameter $\alpha = 0.001$

5 Numerical examples

Finally, we illustrate our theoretical findings by numerical examples. The first two examples are solved by Algorithm 3.5, i.e. Algorithm 4.2 without damping, making use of the global convergence property from Lemma 4.5. The third one involves a small parameter $\alpha = 10^{-8}$ and is treated using the globalization strategy 4.2 with inexact Armijo line search.
As stopping criterion we require $\|P_{[a,b]}(v) - \bar{u}_h\|_{L^2(\Omega)} < 10^{-11}$ in Algorithm 3.5 and $\alpha\|S_h\|^{-2}\|u(w) - u_h\|_{L^2(\Omega)} \leq \|\nabla\phi(w)\|_{L^2(\Omega)} \leq 10^{-14}$ in Algorithm 4.2, compare Lemma 4.3. The first criterion uses the a posteriori bound for *admissible* $v \in U_{ad}$

$$\|v - \bar{u}_h\|_{L^2(\Omega)} \leq \frac{1}{\alpha}\|\zeta\|_{L^2(\Omega)}, \qquad \zeta(\omega) = \begin{cases} [\alpha v + p_h(v)]_- & \text{if } v(\omega) = a \\ [\alpha v + p_h(v)]_+ & \text{if } v(\omega) = b \\ \alpha v + p_h(v) & \text{if } a < v(\omega) < b \end{cases},$$

presented in [15] and [21].

Example 5.1 (Dirichlet)**.** We consider problem ($\mathbb{P}$) from (1.1) in the situation of Example (ii), i.e. with a Poisson state equation with homogeneous boundary conditions and distributed control $u \in L^2(\Omega)$ on the unit square $\Omega = (0,1)^2$. Let $a \equiv 0.3$, $b \equiv 1$ and

$$z = 4\pi^2\alpha\sin(\pi x)\sin(\pi y) + (S \circ \imath)r\,, \quad \text{where } r = \min\big(1, \max\big(0.3, 2\sin(\pi x)\sin(\pi y)\big)\big)\,.$$

The choice of parameters implies a unique solution $\bar{u} = r$ to the continuous problem ($\mathbb{P}$).

Throughout this section, solutions to the state equation are approximated by continuous, piecewise linear finite elements on a quasi-uniform triangulation T_h with maximal edge length $h > 0$. The meshes are generated through regular refinement starting from the coarsest mesh.

Problem ($\mathbb{P}_h$) admits a unique solution $\bar{u}_h$, compare Section 2. As shown in [11, 12], and [14, Chapter 3], we have

$$\|\bar{u}_h - \bar{u}\|_{L^2(\Omega)} = O(h^2)$$

as $h \to 0$. There also holds quadratic convergence up to a logarithmic factor in $L^\infty(\Omega)$

$$\|\bar{u}_h - \bar{u}\|_{L^\infty(\Omega)} = O(|\log(h)|^{\frac{1}{2}}h^2)$$

for domains $\Omega \subset \mathbb{R}^2$, see also [11, 12], and [14, Chapter 3]. Both convergence rates are observed numerically and are presented in Table 1, which shows the L^2- and the L^∞-errors together with the corresponding experimental orders of convergence

$$EOC_i = \frac{\ln ERR(h_{i-1}) - \ln ERR(h_i)}{\ln(h_{i-1}) - \ln(h_i)}$$

for Example 5.1. Lemma 4.5 ensures global convergence of the undamped Algorithm 3.5 only for $\alpha > 1/(3\pi^4) \simeq 0.0034$, but it is still observed for $\alpha = 0.001$.
The algorithm is initialized with $v_0 \equiv 0.3$. The resulting number of Newton steps as well as the value of ζ/α for the computed solution are also given in Table 1.
Figure 1 shows the Newton iterates, active and inactive sets are very well distinguishable, the jumps along their frontier can be observed.

Example 5.2 (Neumann)**.** We next consider an elliptic problem with Neumann boundary conditions

$$\begin{aligned} -\Delta y + y &= u \qquad \text{in } \Omega\,, \\ \partial_n y &= 0 \qquad \text{on } \partial\Omega\,, \end{aligned}$$

on $\Omega = (0,1)^2$, with a similar discrete setting as in the previous example. It then is clear, how $(\mathbb{P})$ and $(\mathbb{P}_h)$ have to be understood. We set $\alpha = 1$ and choose $a \equiv -1$, $b \equiv 1$ together with

$$z = -2(2\pi^2+1)\alpha\cos(\pi x)\cos(\pi y) + (S\circ\imath)r\ , \quad \text{with } r = \min\big(1, \max\big(-1, 2\cos(\pi x)\cos(\pi y)\big)\big)\,.$$

The optimal control to the continuous problem then is $\bar{u} = r$.

For $\alpha = 1$ the undamped iteration still converges globally, although the solution operator has norm $\|S\| = 1$ as an endomorphism in $L^2(\Omega)$. The predicted convergence properties and the stopping criterion are the same as above; Algorithm 4.2 is initialized by $v_0 \equiv -1$. The first four steps of the iteration are displayed in Figure 2 and the behaviour of the approximation error between the exact and the semidiscrete solution, as well as the number of iterations and the final value of ζ/α, is shown in Table 2.
To investigate Algorithm 4.2 with damping we again consider Example 5.1, this time with $\alpha = 10^{-8}$. We choose $\beta = \frac{1}{2}$ for the Armijo line search and initialize the algorithm with $w_0 \equiv 0$.

mesh param. h	ERR	ERR_∞	EOC	EOC_∞	Iterations	Quality
$\sqrt{2}/16$	2.5865e-03	1.2370e-02	1.95	1.79	4	2.16e-15
$\sqrt{2}/32$	6.5043e-04	3.2484e-03	1.99	1.93	4	2.08e-15
$\sqrt{2}/64$	1.6090e-04	8.1167e-04	2.02	2.00	4	2.03e-15
$\sqrt{2}/128$	4.0844e-05	2.1056e-04	1.98	1.95	4	1.99e-15
$\sqrt{2}/256$	1.0025e-05	5.3806e-05	2.03	1.97	4	1.69e-15
$\sqrt{2}/512$	2.5318e-06	1.3486e-05	1.99	2.00	4	1.95e-15

Table 1: L^2- and L^∞-error development for Example 5.1 (Dirichlet)

mesh param. h	ERR	ERR_∞	EOC	EOC_∞	Iterations	Quality
$\sqrt{2}/16$	3.9866e-03	1.1218e-02	1.94	1.74	3	1.81e-12
$\sqrt{2}/32$	1.0025e-03	3.2332e-03	1.99	1.79	3	2.31e-12
$\sqrt{2}/64$	2.5188e-04	8.4398e-04	1.99	1.94	3	9.74e-13
$\sqrt{2}/128$	6.2936e-05	2.1856e-04	2.00	1.95	3	9.37e-13
$\sqrt{2}/256$	1.5740e-05	5.5223e-05	2.00	1.99	3	8.91e-13
$\sqrt{2}/512$	3.9346e-6	1.3928e-05	2.00	2.00	3	8.86e-13

Table 2: Development of the error in Example 5.2 (Neumann)

mesh param. h	ERR	ERR_∞	EOC	EOC_∞	Iterations
$\sqrt{2}/2$	1.1230e-01	1.6724e-01	-	-	7
$\sqrt{2}/4$	3.8502e-02	1.1784e-01	1.54	0.51	35
$\sqrt{2}/8$	1.0812e-02	3.4228e-02	1.83	1.78	19
$\sqrt{2}/16$	2.1770e-03	1.8471e-02	2.31	0.89	28
$\sqrt{2}/32$	5.6915e-04	3.1867e-03	1.94	2.54	28
$\sqrt{2}/64$	1.0307e-04	1.0031e-03	2.47	1.67	32
$\sqrt{2}/128$	2.5753e-05	2.5104e-04	2.00	2.00	32
$\sqrt{2}/256$	5.2530e-06	5.1462e-05	2.29	2.29	30
$\sqrt{2}/512$	1.2863e-06	1.6448e-05	2.03	1.66	31

Table 3: Development of the error in Example 5.1 (Dirichlet) for $\alpha = 10^{-8}$.

Table 3 shows errors and the number of iterations for different mesh parameters h. The number of iterations appears to be independent of h. For every mesh size the maximum number of damping steps applied is 9 amounting to a minimal damping factor $\lambda_{\min} \sim 2 \cdot 10^{-3}$, which is much larger than the magnitude $\sim 3 \cdot 10^{-6}$ predicted by Lemma 4.3. In all cases the algorithm terminates without damping during the last steps.

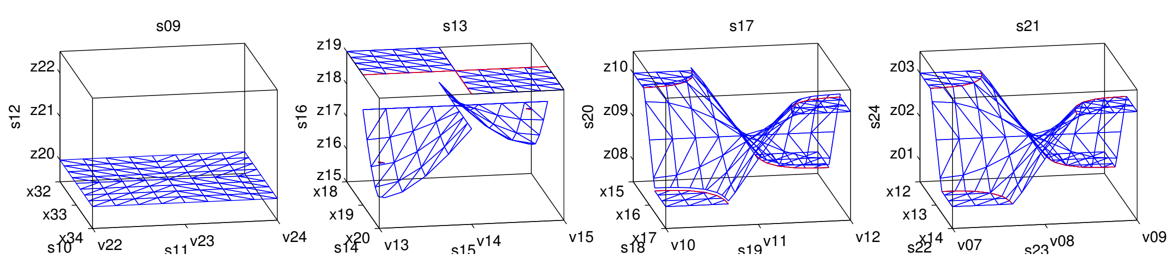

Figure 2: The first steps of the Newton-algorithm for Example 5.2 (Neumann) with $\alpha = 1$.

Acknowledgements

The first author gratefully acknowledges the support of the DFG Priority Program 1253 entitled Optimization With Partial Differential Equations. We wish to thank the anonymous referee for his valuable comments which helped us to improve the material of the paper, in particular in Section 3.1. We also thank Andreas Günther for some fruitful discussions.

A Proof of Lemma 3.4

Lemma A.1. *Let u_h denote the solution of $(\mathbb{P}_h)$. Under the assumptions of Lemma 3.4 there exist $\delta, h_0, C, s_0 > 0$, such that for $0 \le h < h_0$ the set*

$$\mathcal{C}_h^{\tilde{u}}(s) = \left\{ \omega \in \Omega : \ 0 \le |-\frac{1}{\alpha} p_h(\tilde{u})(\omega) - a| < s \vee 0 \le |-\frac{1}{\alpha} p_h(\tilde{u})(\omega) - b| < s \right\}$$

satisfies $\operatorname{meas}(\mathcal{C}_h^{\tilde{u}}(s)) \le Cs$ *for $0 \le s < s_0$ and for all $\tilde{u} \in B_\delta(u_h)$.*

Proof. Our approach is inspired by [2] and [3] and also relies on $W^{1,\infty}$-convergence results from [19] for the finite element discretization of elliptic operators. W.l.o.g. we assume that $b = \infty$, that $\mathcal{C} := \mathcal{C}_u$ is connected, and that $\mathcal{C} \cap \partial\Omega = \emptyset$. It follows from the assumptions on $p(u)$ that the level set $\mathcal{C}$ is a compact non self-intersecting C^2-curve. Also there exists a signed distance function d of class C^2 defined in a neighborhood of $\mathcal{C}$, such that $|d(\omega)| = \operatorname{dist}(\omega, \mathcal{C})$ and $\nabla d = -\frac{\nabla p(u)}{\|\nabla p(u)\|}$, see for example [5, Lemma 14.16]. Using d we can define the unique projection of ω onto $\mathcal{C}$

$$\mathrm{P}_{\mathcal{C}}(\omega) = \omega - d(\omega)\nabla d(\omega) = \arg\inf_{\tilde{\omega} \in \mathcal{C}} \|\tilde{\omega} - \omega\| \,.$$

One can show that $\nabla d(\omega) = \nabla d(\mathrm{P}_{\mathcal{C}}(\omega))$ which is also a unit normal vector to $\mathcal{C}$ at $\mathrm{P}_{\mathcal{C}}(\omega)$, compare for example [18, Prop. 5.1]. One easily sees that inside $\mathcal{U}_\eta$ the projection is constant in the normal direction of $\mathcal{C}$, i.e. for $\omega \in \mathcal{C}$, $t \in (-\eta, \eta)$ one has

$$\omega = \mathrm{P}_{\mathcal{C}}\big(\omega + t\nabla d(\omega)\big) \,. \tag{A.1}$$

Finally, by the assumptions of Lemma 3.4 there exists $\eta > 0$ such that $-\frac{1}{\alpha}\langle \nabla p(u), \nabla d \rangle \ge \beta$ holds on the tube

$$\mathcal{U}_\eta = \{\omega \in \Omega : \ |d(\omega)| \le \eta\}$$

around $\mathcal{C}$, so that $\mathcal{U}_\eta$ is contained in the domain of the projection $\mathrm{P}_{\mathcal{C}}$. Since $\bar{\Omega}$ is compact, there exists $\mu_\eta > 0$ with

$$\inf_{\omega \in \Omega \setminus \mathcal{U}_\eta} |-\frac{1}{\alpha} p(u)(\omega) - a| \ge \mu_\eta > 0 \,. \tag{A.2}$$

Now choose $\delta, h_0 > 0$ such that for all $h \in [0, h_0)$ and $\tilde{u} \in B_\delta(u_h)$

$$\begin{aligned} \|p_h(\tilde{u}) - p(u)\|_{W^{1,\infty}} &\le \|(S_h - S)(Su - z)\|_{W^{1,\infty}} + \|S_h(S_h\tilde{u} - Su)\|_{W^{1,\infty}} \\ &\le Ch\|p(u)\|_{W^{2,\infty}(\Omega)} + C\|(S_h\tilde{u} - Su)\|_{L^p(\Omega)} \\ &\le Ch\|p(u)\|_{W^{2,\infty}(\Omega)} + C\left(\|(S_h - S)\tilde{u}\|_{L^p(\Omega)} + \|S(\tilde{u} - u_h)\|_{L^p(\Omega)} + \|S(u_h - u)\|_{L^p(\Omega)}\right) \\ &\le \alpha \min(\frac{\beta}{2}, \frac{\mu_\eta}{2}) \end{aligned} \tag{A.3}$$

holds, where $p > 2$ is arbitrary. The second estimate follows from Assumption 3.3, in particular one has for $v \in L^p(\Omega)$

$$\|S_h v\|_{1,\infty} \leq c\|Sv\|_{1,\infty} \leq \tilde{c}\|Sv\|_{W^{2,p}(\Omega)} \leq C\|v\|_{L^p(\Omega)}\,.$$

Consider $\tilde{u} \in B_\delta(u_h)$. Together, (A.2) and (A.3) imply that $\mathcal{U}_\eta$ contains the level set

$$\mathcal{C}_{h,\tilde{u}} = \left\{\omega \in \Omega:\ -\frac{1}{\alpha}p_h(\tilde{u})(\omega) = a\right\}\,.$$

The rest of the proof is divided into three steps.

a) First we prove that for $h \in [0, h_0)$ the restriction $\mathrm{P}_{\mathcal{C}}\big|_{\mathcal{C}_{h,\tilde{u}}}$ is one-to-one and onto. Taking into account (A.1) this is equivalent to the following statement: for $\omega \in \mathcal{C}$ the function $t \mapsto -\frac{1}{\alpha}p_h(\tilde{u})(\omega + t\nabla d(\omega)) - a$ has a unique zero in $(-\eta, \eta)$.

Observe that $p_h(\tilde{u})$ is piecewise smooth and globally continuous and thus for $s \in [-\eta, t]$

$$-\frac{1}{\alpha}p_h(\tilde{u})(\omega + t\nabla d(\omega)) + \frac{1}{\alpha}p_h(\tilde{u})(\omega + s\nabla d(\omega)) = -\frac{1}{\alpha}\int_s^t \langle\nabla p_h(\tilde{u}), \nabla d\rangle(\omega + t\nabla d(\omega))\mathrm{d}t$$

and using the L^∞-convergence of the gradients stated in (A.3) we can continue to estimate

$$\begin{aligned}&\geq \frac{1}{\alpha}\int_s^t -\langle\nabla p(u), \nabla d\rangle(\omega + t\nabla d(\omega))\mathrm{d}t - \frac{\beta}{2}(t-s)\\ &\geq \frac{\beta}{2}(t-s)\,.\end{aligned} \tag{A.4}$$

Hence $t \mapsto p_h(\tilde{u})(\omega + t\nabla d(\omega))$ is strictly increasing.

The change of sign, and hence the existence of a unique zero, now follows from the L^∞-bound in (A.3) together with

$$-\frac{1}{\alpha}p(u)(\omega - \eta\nabla d(\omega)) - a \geq \mu_\eta \quad \text{and} \quad -\frac{1}{\alpha}p(u)(\omega + \eta\nabla d(\omega)) - a \leq -\mu_\eta\,,$$

which in turn is a consequence of (A.2) and (A.4).

b) Next we derive a bound to the measure of the critical set $\mathcal{C}_h^{\tilde{u}}(s)$. To this end let us define the projection $\mathrm{P}_{\mathcal{C}_{h,\tilde{u}}}: \mathcal{U}_\eta \to \mathcal{C}_{h,\tilde{u}}$ onto $\mathcal{C}_{h,\tilde{u}}$ along the normal field ∇d of $\mathcal{C}$

$$\mathrm{P}_{\mathcal{C}_{h,\tilde{u}}}(\omega) = \mathrm{P}_{\mathcal{C}}(\omega) + t_{\omega,h}\nabla d(\omega) = \omega + (t_{\omega,h} - d(\omega))\nabla d(\omega)\,,$$

where for $\omega \in \mathcal{U}_\eta$ and $h \in [0, h_0)$ the parameter $t_{\omega,h} \in (-\eta, \eta)$ is the unique zero from *a)*. From (A.4) we deduce that

$$\begin{aligned}|-\frac{1}{\alpha}p_h(\tilde{u})(\omega) - a| &= \frac{1}{\alpha}|-p_h(\tilde{u})\big(\mathrm{P}_{\mathcal{C}}(\omega) + d(\omega)\nabla d(\omega)\big) + p_h(\tilde{u})\big(\mathrm{P}_{\mathcal{C}}(\omega) + t_{\omega,h}\nabla d(\omega)\big)|\\ &\geq \frac{\beta}{2\alpha}|d(\omega) - t_{\omega,h}| = \frac{\beta}{2\alpha}\|\omega - \mathrm{P}_{\mathcal{C}_{h,\tilde{u}}}(\omega)\| \geq \frac{\beta}{2\alpha}\mathrm{dist}(\omega, \mathcal{C}_{h,\tilde{u}})\,.\end{aligned}$$

Together with (A.3) this shows, that for $0 \leq s < s_0 =: \min(\frac{\beta\eta}{4\alpha}, \frac{\mu_\eta}{2})$ the critical set $\mathcal{C}_h^{\tilde{u}}(s)$ is contained in the intersection of $\mathcal{U}_\eta$ and a tube of radius $\frac{2\alpha}{\beta}s$ around $\mathcal{C}_{h,\tilde{u}}$. The measure of this tube is $2\,\mathrm{length}(\mathcal{C}_{h,\tilde{u}})\frac{2\alpha}{\beta}s$.

c) It remains to bound the length of the pice-wise smooth curve $\mathcal{C}_{h,\tilde{u}}$. In order to investigate the volume of the manifold $\mathcal{C}_{h,\tilde{u}}$ we use the restricted projection $\mathrm{P}_{\mathcal{C}}\big|_{\mathcal{C}_{h,\tilde{u}}}$ which according to *a)* is one-to-one and onto.
Due to the polygonal structure of $\mathcal{C}_{h,\tilde{u}}$ the set of non-smooth points of $\mathcal{C}_{h,\tilde{u}}$ has measure zero. First we exploit the C^2-smoothness of d to compute the Derivative of $\mathrm{P}_{\mathcal{C}}$,

$$D\mathrm{P}_{\mathcal{C}} = I - \nabla d(\nabla d)^T - d\nabla^2 d\,.$$

The Jacobian of $\mathrm{P}_{\mathcal{C}}\big|_{\mathcal{C}_{h,\tilde{u}}} : \mathcal{C}_{h,\tilde{u}} \to \mathcal{C}$ can then be computed from $D\mathrm{P}_{\mathcal{C}}$ via projection onto the tangential spaces. To this end let λ, λ_h denote the unique unit tangential vectors, that belong to the orientation induced on $\mathcal{C}$ and $\mathcal{C}_{h,\tilde{u}}$ by d and $\nabla p_h(\tilde{u})$, respectively. After possibly decreasing h_0 and δ we get for all $\omega \in \mathcal{C}_{h,\tilde{u}}$

$$D_{\mathcal{C}_{h,\tilde{u}}}\mathrm{P}_{\mathcal{C}}(\omega) = \langle \lambda^{\mathrm{P}_{\mathcal{C}}(\omega)}, D\mathrm{P}_{\mathcal{C}}\left(\lambda_h^{\omega}\right)\rangle = \langle \lambda^{\mathrm{P}_{\mathcal{C}}(\omega)}, \lambda_h^{\omega}\rangle + O(|d(\omega)|) \geq \frac{1}{2}\,,$$

because $\langle \lambda^{\mathrm{P}_{\mathcal{C}}(\omega)}, \nabla d(\omega)\rangle = \langle \lambda^{\mathrm{P}_{\mathcal{C}}(\omega)}, \nabla d(\mathrm{P}_{\mathcal{C}}(\omega))\rangle = 0$ and $\langle \lambda, \lambda_h\rangle = \langle \frac{\nabla p(u)}{\|\nabla p(u)\|}, \frac{\nabla p_h(\tilde{u})}{\|\nabla p_h(\tilde{u})\|}\rangle \overset{h,\delta\to 0}{\longrightarrow} 1$ uniformly. Thus $\mathrm{P}_{\mathcal{C}}\big|_{\mathcal{C}_{h,\tilde{u}}}$ is a local diffeomorphism, a.e. on $\mathcal{C}_{h,\tilde{u}}$. Hence

$$\mathrm{length}(\mathcal{C}_{h,\tilde{u}}) = \int_{\mathcal{C}_{h,\tilde{u}}} 1\mathrm{d}A_h = \int_{\mathcal{C}} |\det\left(D_{\mathcal{C}_{h,\tilde{u}}}\mathrm{P}_{\mathcal{C}}\right)|^{-1}\mathrm{d}A \leq \int_{\mathcal{C}} 2\mathrm{d}A = 2\,\mathrm{length}(\mathcal{C})\,.$$

∎

Proof of Lemma 3.4. Choose $\delta, h_0, s_0, C > 0$ as in Lemma A.1. If necessary, further decrease δ and h_0 until $\|p_h(u_1) - p_h(u_2)\|_\infty < \min(b-a, \alpha s_0)$ for all $u_1, u_2 \in B_\delta(u_h)$ and $0 \leq h < h_0$ holds, so that in particular $-\frac{1}{\alpha}p_h(u_1)(\omega)$ and $-\frac{1}{\alpha}p_h(u_2)(\omega)$ can only be active at the same bound at a given $\omega \in \Omega$. In this situation we have

$$\begin{aligned}
&\left\| G_h(u_1) - G_h(u_2) - \left(I + \frac{\mathbb{1}_{p_h(u_2)}}{\alpha}S_h^*S_h\right)(u_1 - u_2)\right\|_{L^2(\Omega)}^2 = \dots \\
&\quad = \int_\Omega \left(P_{[a,b]}\left(-\frac{1}{\alpha}p_h(u_1)\right) - P_{[a,b]}\left(-\frac{1}{\alpha}p_h(u_2)\right) - \frac{\mathbb{1}_{p_h(u_2)}}{\alpha}S_h^*S_h(u_1-u_2)\right)^2 \mathrm{d}\omega \\
&\quad = \int_{\mathcal{A}\left(p_h^{u_1}\right)\cap\mathcal{I}\left(p_h^{u_2}\right)} \Bigg(\overbrace{P_{[a,b]}\left(-\frac{1}{\alpha}p_h(u_1)\right)}^{=a \text{ or } =b} - \overbrace{\left(-\frac{1}{\alpha}p_h(u_2)\right) - \frac{1}{\alpha}S_h^*S_h(u_1-u_2)}^{=-\frac{1}{\alpha}p_h(u_1)}\Bigg)^2 \mathrm{d}\omega \dots \\
&\qquad + \int_{\mathcal{I}\left(p_h^{u_1}\right)\cap\mathcal{A}\left(p_h^{u_2}\right)} \Bigg(-\frac{1}{\alpha}p_h(u_1) - \underbrace{P_{[a,b]}\left(-\frac{1}{\alpha}p_h(u_2)\right)}_{=a \text{ or } =b}\Bigg)^2 \mathrm{d}\omega\,,
\end{aligned}$$

where $\mathcal{A}$ and $\mathcal{I}$ again denote the corresponding active and inactive sets. It follows that

$$\begin{aligned}
&\left\| G_h(u_1) - G_h(u_2) - \left(I + \frac{\mathbb{1}_{p_h(u_2)}}{\alpha}S_h^*S_h\right)(u_1 - u_2)\right\|_{L^2(\Omega)} \leq \dots \\
&\qquad \leq \mathrm{meas}((\mathcal{A}\left(p_h^{u_1}\right)\cap\mathcal{I}\left(p_h^{u_2}\right)) \cup (\mathcal{I}\left(p_h^{u_1}\right)\cap\mathcal{A}\left(p_h^{u_2}\right)))^{\frac{1}{2}}\frac{1}{\alpha}\|p_h(u_1) - p_h(u_2)\|_\infty\,.
\end{aligned}$$

Now we have

$$\left(\mathcal{A}\left(p_h^{u_1}\right) \cap \mathcal{I}\left(p_h^{u_2}\right)\right) \cup \left(\mathcal{I}\left(p_h^{u_1}\right) \cap \mathcal{A}\left(p_h^{u_2}\right)\right) \subset \mathcal{C}_h^{u_2}\left(\frac{1}{\alpha}\|p_h(u_1) - p_h(u_2)\|_\infty\right) ,$$

and by the choice of h_0 we can use Lemma A.1 to get

$$\begin{aligned}\left\|G_h(u_1) - G_h(u_2) - \left(I + \frac{\mathbb{1}_{p_h(u_2)}}{\alpha} S_h^* S_h\right)(u_1 - u_2)\right\|_{L^2(\Omega)} &\le C\frac{1}{\alpha}\|p_h(u_1) - p_h(u_2)\|_\infty^{\frac{3}{2}} \\ &\le C\frac{\|S_h^* S_h\|_{\mathcal{L}(L^2(\Omega),L^\infty(\Omega))}^{\frac{3}{2}}}{\alpha}\|u_1 - u_2\|_{L^2(\Omega)}^{\frac{3}{2}} .\end{aligned}$$

Similarly one shows the continuity property of the single valued ∂G on $B_\delta(u_h)$, which again implies Hölder continuity of ∂G in that neighborhood of u_h.

$$\begin{aligned}\|\partial G_h(u_1) - \partial G_h(u_2)\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))} &= \left\|\left(I + \frac{\mathbb{1}_{p_h(u_1)}}{\alpha} S_h^* S_h\right) - \left(I + \frac{\mathbb{1}_{p_h(u_2)}}{\alpha} S_h^* S_h\right)\right\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))} \\ &\le \frac{\|S_h\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))}^2}{\alpha}\|\mathbb{1}_{p_h(u_1)} - \mathbb{1}_{p_h(u_2)}\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))} \\ &\le \frac{\|S_h\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))}^2}{\alpha}\mathrm{meas}\left(\left(\mathcal{A}\left(p_h^{u_1}\right) \cap \mathcal{I}\left(p_h^{u_2}\right)\right) \cup \left(\mathcal{I}\left(p_h^{u_1}\right) \cap \mathcal{A}\left(p_h^{u_2}\right)\right)\right)^{\frac{1}{2}} \\ &\le C\frac{\|S_h\|_{\mathcal{L}(L^2(\Omega),L^2(\Omega))}^2}{\alpha}\|S_h\|_{\mathcal{L}(L^2(\Omega),L^\infty(\Omega))}^{\frac{1}{2}}\|S_h(u_1 - u_2)\|_{L^2(\Omega)}^{\frac{1}{2}} .\end{aligned}$$

∎

Proof of Lemma 4.1. The proof is inspired by the one given in [7]. First we note that the minimization problem in (4.1) can be rewritten as

$$\min_{y \in L^2(\Omega)}\left(\frac{1}{2}\|y - z_h\|_{L^2(\Omega)}^2 - \langle w, y\rangle_{L^2(\Omega)}\right) + \min_{u \in L^2(\Omega)}\left(\frac{\alpha}{2}\|u\|_{L^2(\Omega)}^2 + \chi_{U_{ad}}(u) + \langle w, S_h u\rangle_{L^2(\Omega)}\right) ,$$

with unique minimizers $u(w)$ and $y(w)$, respectively. For $y(w)$ this is obvious. For the problem in u we get necessary conditions like in (1.2) which guarantee a unique solution $u(w)$. From the viewpoint of convex analysis these conditions read

$$v \in \alpha u(w) + \partial\chi_{U_{ad}}(u(w)) , \tag{A.5}$$

where $v = -S_h^* w$ and the subdifferential $\partial\chi_{U_{ad}}$ is the Fréchet normal cone of the convex set U_{ad}, i.e. $\partial\chi_{U_{ad}}(u) = \chi_{\hat{\mathcal{A}}(p_h^u)} - \chi_{\check{\mathcal{A}}(p_h^u)}$, whence one has the formula $u(w) = P_{[a,b]}\left(-\frac{1}{\alpha}S_h^* w\right)$. With the smoothness of ϕ we have

$$\phi(w) = -\frac{1}{2}\|w\|_{L^2(\Omega)}^2 + \psi^*(-S_h^* w) + \langle w, w + z_h\rangle_{L^2(\Omega)} ,$$

ψ^* being the polar function of the convex functional $\psi(u) = \frac{\alpha}{2}\|u\|_{L^2(\Omega)}^2 + \chi_{U_{ad}}(u)$ defined as

$$\psi^*(v) = \sup_{u \in L^2(\Omega)}\left(\langle v, u\rangle_{L^2(\Omega)} - \psi(u)\right) .$$

Since ψ is convex and lower semicontinuous we have the property $v \in \partial\psi(u) \Leftrightarrow u \in \partial\psi^*(v)$, see [4, Cor. 5.2]. Furthermore, because for any $v \in L^2(\Omega)$ equation (A.5) can be uniquely solved for $u = P_{[a,b]}(\frac{1}{\alpha}v)$ we conclude that $\partial\psi^*(v)$ is single valued, thus ψ^* is Gâteaux differentiable, compare [4, Prop. 5.3]. Since $\partial\psi^*(v) = P_{[a,b]}\left(\frac{1}{\alpha}v\right)$ the function ψ^* is even Lipschitz continuously Fréchet differentiable. Hence

$$\phi'(w)v = \langle w + z_h, v\rangle_{L^2(\Omega)} + \left\langle P_{[a,b]}\left(-\frac{1}{\alpha}S_h^* w\right), (-S_h^*)v\right\rangle_{L^2(\Omega)},$$

and the formula for the gradient follows.
The strong convexity now is a consequence of the monotonicity of the derivative

$$\begin{aligned}\langle\phi'(w_1) - \phi'(w_2), w_1 - w_2\rangle_{L^2(\Omega)} &= \|w_1 - w_2\|^2_{L^2(\Omega)} \cdots \\ &+ \left\langle \left(P_{[a,b]}\left(-\frac{1}{\alpha}S_h^* w_2\right) - P_{[a,b]}\left(-\frac{1}{\alpha}S_h^* w_1\right)\right), S_h^*(w_1 - w_2)\right\rangle_{L^2(\Omega)} \geq \|w_1 - w_2\|^2_{L^2(\Omega)}\end{aligned} \tag{A.6}$$

which again follows from the definition of the orthogonal projection $P_{U_{ad}}$. ∎

References

[1] M. Bergounioux, K. Ito, and K. Kunisch. Primal-dual strategy for constrained optimal control problems. *SIAM J. Control Optimization*, **37**(4), 1176–1194 (1999).

[2] G. Dziuk: *Finite elements for the Beltrami operator on arbitrary surfaces* Partial differential equations and calculus of variations, Lect. Notes Math. 1357, 142-155 (1988).

[3] G. Dziuk, C.M. Elliott: *Finite elements on evolving surfaces* IMA J. Numer. Anal. **27** (2), 262–292 (2007).

[4] I. Ekeland and R. Témam. *Convex analysis and variational problems. Unabridged, corrected republication of the 1976 English original.* Classics in Applied Mathematics. 28. Philadelphia, SIAM (1999).

[5] D. Gilbarg, N. S. Trudinger: *Elliptic partial differential equations of second order*, Springer, Berlin (1998).

[6] C. Gräser. Globalization of nonsmooth Newton methods for optimal control problems. Kunisch, Karl (ed.) et al., Numerical mathematics and advanced applications. Proceedings of ENUMATH 2007, the 7th European conference on numerical mathematics and advanced applications, Graz, Austria, September 10–14, 2007. Berlin: Springer. 605-612 (2008).

[7] C. Gräser and R. Kornhuber. Nonsmooth Newton methods for set-valued saddle point problems. *SIAM J. Numer. Anal.*, **47**(2), 1251–1273 (2009).

[8] A. Günther, M.H. Tber: *A goal-oriented adaptive Moreau-Yosida algorithm for control- and state-constrained elliptic control problems*, Preprint No. SPP1253-089, DFG Schwerpunktprogramm 1253 (2009).

[9] M. Hintermüller, K. Ito and K. Kunisch: *The primal-dual active set method as a semi-smooth Newton method*. SIAM J. Control and Optim. **13** (3), 865–888 (2003).

[10] M. Hintermüller and M. Ulbrich: *A mesh-independence result for semismooth Newton methods* Mathematical Programming **101**, 151–184 (2004).

[11] M. Hinze: *A generalized discretization concept for optimal control problems with control constraints*, Preprint MATH-NM-02-2003, Institut für Numerische Mathematik, Technische Universität Dresden (2003).

[12] M. Hinze: *A variational discretization concept in control constrained optimization: the linear-quadratic case*, J. Computational Optimization and Applications **30** (1), 45–61 (2005).

[13] M. Hinze and U. Matthes: *A note on variational discretization of elliptic Neumann boundary control*, Control & Cybernetics **38**, 577-591 (2009).

[14] M. Hinze, R. Pinnau, M. Ulbrich, and S. Ulbrich: *Optimization with PDE constraints.* Mathematical Modelling: Theory and Applications **23**. Dordrecht: Springer (2009).

[15] K. Krumbiegel and A. Rösch: *A new stopping criterion for iterative solvers for control constrained optimal control problems*, Archives of Control Sciences **18** (1), 17–42 (2008).

[16] C. Meyer and A. Rösch: *Superconvergence properties of optimal control problems.* SIAM J. Control Optim. **43** (3), 970–985 (2004).

[17] F. Mignot and F.P. Puel: *Optimal control in some variational inequalities.* SIAM J. Control Optim. **22**, 466–477 (1984).

[18] M. Moubachir and J.-P. Zolésio. *Moving shape analysis and control. Applications to fluid structure interactions.* Boca Raton, FL: Chapman & Hall/CRC, 2006.

[19] R. Rannacher and R. Scott: *Some optimal error estimates for piecewise linear finite element approximations.* Math. Comput. **38**, 437–445 (1982).

[20] A. Schiela: *The Control Reduced Interior Point Method. A Function Space Oriented Algorithmic Approach.* Verlag Dr. Hut, München 2006.

[21] F. Tröltzsch and S. Volkwein: *POD a-posteriori error estimates for linear-quadratic optimal control problems*, Computational Optimization and Applications, Online First (2008).

[22] M. Ulbrich: *Semismooth Newton methods for operator equations in function spaces*, SIAM J. Optim. **13**, 805–841 (2003).

[23] M. Ulbrich:*A new mesh-Independence result for semismooth Newton methods*, Oberwolfach Report **4/2009**, 78–81 (2009).

[24] M. Ulbrich. *Semismooth Newton methods for variational inequalities and constrained optimization problems in function spaces.* MOS-SIAM Series on Optimization. Philadelphia (2011).

[25] M. Vierling: *Ein semiglattes Newtonverfahren für semidiskretisierte steuerungsbeschränkte Optimalsteuerungsprobleme*, Diplomarbeit, Universität Hamburg (2007).