Hamburger Beiträge
zur Angewandten Mathematik

A finite element approximation to elliptic control problems in the presence of control and state constraints

Klaus Deckelnick and Michael Hinze

Nr. 2007-01
Januar 2007

A finite element approximation to elliptic control problems in the presence of control and state constraints

Klaus Deckelnick*& Michael Hinze†

Abstract: We consider an elliptic optimal control problem with control and pointwise state constraints. The cost functional is approximated by a sequence of functionals which are obtained by discretizing the state equation with the help of linear finite elements and enforcing the state constraints in the nodes of the triangulation. The control variable is not discretized. Error bounds for control and state are obtained both in two and three space dimensions. Finally, we discuss some implementation issues of a generalized Newton method applied to the numerical solution of the problem class under consideration.

Mathematics Subject Classification (2000): 49J20, 49K20, 35B37

Keywords: Elliptic optimal control problem, state constraints, control constraints, error estimates

1 Introduction

Let $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ be a bounded domain with a smooth boundary $\partial\Omega$ and consider the differential operator

$$Ay := -\sum_{i,j=1}^{d} \partial_{x_j}\big(a_{ij} y_{x_i}\big) + \sum_{i=1}^{d} b_i y_{x_i} + cy,$$

along with its formal adjoint operator

$$A^* y = -\sum_{i=1}^{d} \partial_{x_i}\big(\sum_{j=1}^{d} a_{ij} y_{x_j} + b_i y\big) + cy$$

where for simplicity the coefficients a_{ij}, b_i and c are assumed to be smooth functions on $\bar{\Omega}$. We associate with A the bilinear form

$$a(y,z) := \int_\Omega \big(\sum_{i,j=1}^{d} a_{ij}(x) y_{x_i} z_{x_j} + \sum_{i=1}^{d} b_i(x) y_{x_i} z + c(x) yz\big) dx, \qquad y, z \in H^1(\Omega)$$

and subsequently assume that there exists $c_0 > 0$ such that

$$\sum_{i,j=1}^{d} a_{ij}(x)\xi_i\xi_j \geq c_0|\xi|^2 \quad \text{ for all } \xi \in \mathbb{R}^d \text{ and all } x \in \Omega.$$

*Institut für Analysis und Numerik, Otto–von–Guericke–Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany

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

Furthermore we suppose that the form a is coercive on $H^1(\Omega)$, i.e. there exists $c_1 > 0$ such that

$$a(v,v) \geq c_1 \|v\|_{H^1}^2 \qquad \text{for all } v \in H^1(\Omega). \tag{1.1}$$

From the above assumptions it follows that for a given $f \in (H^1(\Omega))'$ the elliptic boundary value problem

$$\begin{aligned} Ay &= f \quad \text{in } \Omega \\ \textstyle\sum_{i,j=1}^{d} a_{ij} u_{x_i} \nu_j &= 0 \quad \text{on } \partial\Omega \end{aligned} \tag{1.2}$$

has a unique weak solution $y \in H^1(\Omega)$ which we denote by $y = \mathcal{G}(f)$. Here, ν is the unit outward normal to $\partial\Omega$. Furthermore, if $f \in L^2(\Omega)$, then the solution y belongs to $H^2(\Omega)$ and satisfies

$$\|y\|_{H^2} \leq C\|f\|,$$

where have used $\|\cdot\|$ to denote the L^2–norm.

Next, let $(U, (\cdot,\cdot)_U)$ be a Hilbert space and $B : U \to L^2(\Omega)$ a linear, continuous operator. We are interested in the following control problem

$$\begin{aligned} &\min_{u \in U_{ad}} J(u) = \frac{1}{2}\int_\Omega |y - y_0|^2 + \frac{\alpha}{2}\|u - u_0\|_U^2 \\ &\text{subject to } y = \mathcal{G}(Bu) \text{ and } y(x) \leq b(x) \text{ in } \Omega. \end{aligned} \tag{1.3}$$

Here, $U_{ad} \subseteq U$ denotes the set of admissible controls which is assumed to be closed and convex. Furthermore, we suppose that $\alpha > 0$ and that $y_0 \in H^1(\Omega)$, $u_0 \in U$ and $b \in W^{2,\infty}(\Omega)$ are given.

In the special case $U \equiv L^2(\Omega)$ without control constraints, i.e. $U_{ad} \equiv L^2(\Omega)$ the finite element analysis of problem (1.3) is carried out in [6]. In the present work we extend the analysis to the case of control and pointwise state constraints. Here we use techniques which are applicable to a wider class of control problems. In particular the results of [6] are contained as a special case.

From here onwards we impose the following assumption which is frequently referred to as *Slater condition* or interior point condition.

Assumption 1.1.

$$\exists \tilde{u} \in U_{ad} \qquad \mathcal{G}(B\tilde{u}) < b \text{ in } \bar{\Omega}.$$

Since the state constraints form a convex set and the set of admissible controls is closed and convex it is not difficult to establish the existence of a unique solution $u \in U_{ad}$ to this problem. In order to characterize this solution we introduce the space $\mathcal{M}(\bar{\Omega})$ of Radon measures which is defined as the dual space of $C^0(\bar{\Omega})$ and endowed with the norm

$$\|\mu\|_{\mathcal{M}(\bar{\Omega})} = \sup_{f \in C^0(\bar{\Omega}), |f| \leq 1} \int_{\bar{\Omega}} f \, d\mu.$$

Using [3, Theorem 5.2] we then infer (compare also [2, Theorem 2])

Theorem 1.2. *Let $u \in U_{ad}$ denote the unique solution to* (1.3)*. Then there exist $\mu \in \mathcal{M}(\bar{\Omega})$ and $p \in L^2(\Omega)$ such that with $y = \mathcal{G}(Bu)$ there holds*

$$\int_\Omega pAv = \int_\Omega (y - y_0)v + \int_{\bar{\Omega}} v \, d\mu \qquad \forall v \in H^2(\Omega) \text{ with } \sum_{i,j=1}^{d} a_{ij} v_{x_i} \nu_j = 0 \text{ on } \partial\Omega \tag{1.4}$$

$$(B^* p + \alpha(u - u_0), v - u_0)_U \geq 0 \qquad \forall v \in U_{ad} \tag{1.5}$$

$$\mu \geq 0,\ y(x) \leq b(x) \text{ in } \Omega \text{ and } \int_{\bar{\Omega}} (b - y) \, d\mu = 0. \tag{1.6}$$

Our aim is to develop and analyze a finite element approximation of problem (1.3). We start by approximating the cost functional J by a sequence of functionals J_h where h is a mesh parameter related to a sequence of triangulations. The definition of J_h involves only the approximation of the state equation by linear finite elements and enforces constraints on the state in the nodes of the triangulation, whereas the controls are still sought in U_{ad}. We shall prove that the minima of J_h converge in L^2 to the minimum of J as $h \to 0$ and that the states convergence strongly in H^1 with corresponding error bounds. We thereby extend the semi–discrete approach to purely control constrained control problems presented by the second author in [9] to problems with control and state constraints.
To the authors knowledge only few attempts have been made to develop a finite element analysis for elliptic control problems in the presence of control and state constraints. In [4] Casas proves convergence of finite element approximations to optimal control problems for semi-linear elliptic equations with finitely many state constraints. Casas and Mateos extend these results in [5] to a less regular setting for the states and prove convergence of finite element approximations to semi-linear distributed and boundary control problems. In [11] Meyer considers a fully discrete strategy to approximate an elliptic control problem with pointwise state and control constraints. He obtains the approximation order $\mathcal{O}(h^{2-d/2-\epsilon})$ for the state in H^1 and for the control in L^2, where d denotes the spatial dimension and $\epsilon > 0$ can be chosen arbitrarily. His results confirm those obtained by the authors in [6] for the purely state constrained case.
Let us comment on further approaches that tackle optimization problems for pdes with control and state constraints. A *Lavrentiev-type regularization* of problem (1.3) is investigated in [13]. In this approach the state constraint $y \le b$ in (1.3) is replaced by the mixed constraint $\epsilon u + y \le b$, with $\epsilon > 0$ denoting a regularization parameter. It turns out that the associated Lagrange multiplier μ_ϵ belongs to $L^2(\Omega)$. Numerical analysis for this approach with emphasis on the coupling of gridsize and regularization parameter ϵ is presented by the second author and Meyer in [10]. The resulting optimization problems are solved either by interior-point methods or primal-dual active set strategies, compare [12]. The development of numerical approaches to tackle (1.3) is ongoing. An excellent overview can be found in [7, 8], where also further references are given, for the latter see also [16].

The paper is organized as follows: in §2 we describe our discretization and establish bounds on the relevant discrete quantities which are uniform in the discretization parameter. These bounds are used in §3 in order to prove the following error bounds

$$\|u - u_h\|_U,\ \|y - y_h\|_{H^1} = \begin{cases} O(h^{\frac{1}{2}}), & \text{if } d = 2, \\ O(h^{\frac{1}{4}}), & \text{if } d = 3, \end{cases}$$

where u_h and y_h are the discrete control and state respectively. If in addition $Bu \in W^{1,s}(\Omega)$ we obtain

$$\|u - u_h\|_U,\ \|y - y_h\|_{H^1} \le Ch^{\frac{3}{2}-\frac{d}{2s}}\sqrt{|\log h|}.$$

Roughly speaking, the idea is to test (1.5) with u_h and (2.8), the discrete counterpart of (1.5), with the continuous solution u. This is feasible since controls are not discretized explicitly. An important tool in the analysis is the use of L^∞–error estimates for finite element approximations of the Neumann problem developed in [14]. The need for uniform estimates is due to the presence of the measure μ in (1.4).

2 Finite element discretization

Let $\mathcal{T}_h$ be a triangulation of Ω with maximum mesh size $h := \max_{T\in\mathcal{T}_h} \operatorname{diam}(T)$ and vertices $x_1, \ldots, x_m$. We suppose that $\bar{\Omega}$ is the union of the elements of $\mathcal{T}_h$ so that element edges lying on the boundary are curved. In addition, we assume that the triangulation is quasi-uniform in the sense that there exists a constant $\kappa > 0$ (independent of h) such that each $T \in \mathcal{T}_h$ is contained in a ball of radius $\kappa^{-1}h$ and contains a ball of radius κh. Let us define the space of linear finite elements,

$$X_h := \{v_h \in C^0(\bar{\Omega}) \,|\, v_h \text{ is a linear polynomial on each } T \in \mathcal{T}_h\}$$

with the appropriate modification for boundary elements. In what follows it is convenient to introduce a discrete approximation of the operator $\mathcal{G}$. For a given function $v \in L^2(\Omega)$ we denote by $z_h = \mathcal{G}_h(v) \in X_h$ the solution of the discrete Neumann problem

$$a(z_h, v_h) \;=\; \int_\Omega v v_h \quad \text{for all } v_h \in X_h.$$

It is well–known that for all $v \in L^2(\Omega)$

$$\|\mathcal{G}(v) - \mathcal{G}_h(v)\| \;\leq\; Ch^2\|v\|, \tag{2.1}$$

$$\|\mathcal{G}(v) - \mathcal{G}_h(v)\|_{L^\infty} \;\leq\; Ch^{2-\frac{d}{2}}\|v\|. \tag{2.2}$$

The estimate (2.2) can be improved provided one strengthens the assumption on v.

Lemma 2.1. *Suppose that $v \in W^{1,s}(\Omega)$ for some $1 < s < \frac{d}{d-1}$. Then*

$$\|\mathcal{G}(v) - \mathcal{G}_h(v)\|_{L^\infty} \leq Ch^{3-\frac{d}{s}}|\log h|\, \|v\|_{W^{1,s}}.$$

Proof. Let $z = \mathcal{G}(v)$, $z_h = \mathcal{G}_h(v)$. Elliptic regularity theory implies that $z \in W^{3,s}(\Omega)$ from which we infer that $z \in W^{2,q}(\Omega)$ with $q = \frac{ds}{d-s}$ using a well–known embedding theorem. Furthermore, we have

$$\|z\|_{W^{2,q}} \leq c\|z\|_{W^{3,s}} \leq c\|v\|_{W^{1,s}}. \tag{2.3}$$

Using Theorem 2.2 and the following Remark in [14] we have

$$\|z - z_h\|_{L^\infty} \leq c|\log h| \inf_{\chi\in X_h} \|z - \chi\|_{L^\infty}, \tag{2.4}$$

which, combined with a well–known interpolation estimate, yields

$$\|z - z_h\|_{L^\infty} \leq ch^{2-\frac{d}{q}}|\log h|\|z\|_{W^{2,q}} \leq ch^{3-\frac{d}{s}}|\log h|\|v\|_{W^{1,s}}$$

in view (2.3) and the relation between s and q. ∎

Problem (1.3) is now approximated by the following sequence of control problems depending on the mesh parameter h:

$$\begin{aligned} &\min_{u\in U_{ad}} J_h(u) := \frac{1}{2}\int_\Omega |y_h - y_0|^2 + \frac{\alpha}{2}\|u - u_{0,h}\|_U^2 \\ &\text{subject to } \; y_h = \mathcal{G}_h(Bu) \text{ and } y_h(x_j) \leq b(x_j) \text{ for } j = 1, \ldots, m. \end{aligned} \tag{2.5}$$

Here, $u_{0,h}$ denotes an approximation to u_0 which is assumed to satisfy

$$\|u_0 - u_{0,h}\| \leq Ch. \tag{2.6}$$

Problem (2.5) represents a convex infinite-dimensional optimization problem of similar structure as problem (1.3), but with only finitely many equality and inequality constraints for the state, which form a convex admissible set. Again we can apply [3, Theorem 5.2] which together with [2, Corollary 1] yields (compare also the analysis of problem (P) in [4])

Lemma 2.2. *Problem (2.5) has a unique solution $u_h \in U_{ad}$. There exist $\mu_1, \dots, \mu_m \in \mathbb{R}$ and $p_h \in X_h$ such that with $y_h = \mathcal{G}_h(Bu_h)$ and $\mu_h = \sum_{j=1}^m \mu_j \delta_{x_j}$ we have*

$$a(v_h, p_h) = \int_\Omega (y_h - y_0)v_h + \int_{\bar\Omega} v_h d\mu_h \qquad \forall v_h \in X_h, \tag{2.7}$$

$$(B^* p_h + \alpha(u_h - u_{0,h}), v - u_h)_U \geq 0 \qquad \forall v \in U_{ad}, \tag{2.8}$$

$$\mu_j \geq 0,\ y_h(x_j) \leq b(x_j), j = 1, \dots, m \text{ and } \int_{\bar\Omega} (I_h b - y_h) d\mu_h = 0. \tag{2.9}$$

Here, δ_x denotes the Dirac measure concentrated at x and I_h is the usual Lagrange interpolation operator.

Remark 2.3. Problem (2.5) is still an infinite–dimensional optimization problem, but with finitely many state constraints. This is reflected by the well known fact that the variational inequalities (1.5) and (2.8) can be rewritten in the form

$$u = \Pi_{U_{ad}}\left(-\frac{1}{\alpha}B^* p + u_0\right) \text{ and } u_h = \Pi_{U_{ad}}\left(-\frac{1}{\alpha}B^* p_h + u_{0,h}\right) \tag{2.10}$$

respectively, where $\Pi_{U_{ad}} : U \to U_{ad}$ denotes the orthogonal projection onto U_{ad}. Due to the presence of Π_{Uad} the function u_h will in general not belong to X_h even in the case $U = L^2(\Omega), B = Id$. This is different for the purely state constrained problem, for which $\Pi_{U_{ad}} \equiv Id$, so that $u_h = -\frac{1}{\alpha}p_h + u_{0,h} \in X_h$ by (2.10). In that case the space $U = L^2(\Omega)$ in (2.5) may be replaced by X_h to obtain the same discrete solution u_h, which results in a finite–dimensional discrete optimization problem instead. However, we emphasize, that the infinite–dimensional formulation of (2.5) is crucial for our numerical analysis in §3.

As a first result for (2.5) we prove that the sequence of optimal controls, states and the measures μ_h are uniformly bounded.

Lemma 2.4. *Let $u_h \in U_{ad}$ be the optimal solution of (2.5) with corresponding state $y_h \in X_h$ and adjoint variables $p_h \in X_h$ and $\mu_h \in \mathcal{M}(\bar\Omega)$. Then there exists $\bar h > 0$ so that*

$$\|y_h\|,\ \|u_h\|_U,\ \|\mu_h\|_{\mathcal{M}(\bar\Omega)} \leq C \qquad \textit{for all } 0 < h \leq \bar h.$$

Proof. Since $\mathcal{G}(B\tilde u)$ is continuous, Assumption 1.1 implies that there exists $\delta > 0$ such that

$$\mathcal{G}(B\tilde u) \leq b - \delta \qquad \text{in } \bar\Omega. \tag{2.11}$$

It follows from (2.2) that there is $h_0 > 0$ with

$$\mathcal{G}_h(B\tilde u) \leq b \quad \text{in } \bar\Omega \text{ for all } 0 < h \leq h_0$$

so that $J_h(u_h) \leq J_h(\tilde u) \leq C$ uniformly in h giving

$$\|u_h\|_U,\ \|y_h\| \leq C \qquad \text{for all } h \leq h_0. \tag{2.12}$$

Next, let u denote the unique solution to problem (1.3). We infer from (2.11) and (2.2) that $v := \frac{1}{2}u + \frac{1}{2}\tilde u$ satisfies

$$\begin{aligned} \mathcal{G}_h(Bv) &\leq \frac{1}{2}\mathcal{G}(Bu) + \frac{1}{2}\mathcal{G}(B\tilde u) + Ch^{2-\frac{d}{2}}(\|Bu\| + \|B\tilde u\|) \\ &\leq b - \frac{\delta}{2} + Ch^{2-\frac{d}{2}}(\|u\|_U + \|\tilde u\|_U) \leq b - \frac{\delta}{4} \qquad \text{in } \bar\Omega \end{aligned} \tag{2.13}$$

provided that $h \le \bar{h}, \bar{h} \le h_0$. Since $v \in U_{ad}$, (2.8), (2.7), (2.12) and (2.13) imply

$$\begin{aligned}
0 &\le (B^* p_h + \alpha(u_h - u_{0,h}), v - u_h)_U = \int_\Omega B(v-u_h)p_h + \alpha(u_h - u_{0,h}, v - u_h)_U \\
&= a(\mathcal{G}_h(Bv) - y_h, p_h) + \alpha(u_h - u_{0,h}, v - u_h)_U \\
&= \int_\Omega (\mathcal{G}_h(Bv) - y_h)(y_h - y_0) + \int_{\bar{\Omega}} (\mathcal{G}_h(Bv) - y_h) d\mu_h + \alpha(u_h - u_{0,h}, v - u_h)_U \\
&\le C + \sum_{j=1}^m \mu_j \big(b(x_j) - \frac{\delta}{4} - y_h(x_j)\big) = C - \frac{\delta}{4}\sum_{j=1}^m \mu_j
\end{aligned}$$

where the last equality is a consequence of (2.9). It follows that

$$\|\mu_h\|_{\mathcal{M}(\bar{\Omega})} \le C$$

and the lemma is proved. ∎

3 Error analysis

An important ingredient in our analysis is an error bound for a solution of a Neumann problem with a measure valued right hand side. Let A as above and consider

$$\begin{aligned}
A^* q &= \tilde{\mu}\llcorner\Omega \quad \text{in } \Omega \\
\sum_{i=1}^d \big(\sum_{j=1}^d a_{ij} q_{x_j} + b_i q\big)\nu_i &= \tilde{\mu}\llcorner\partial\Omega \quad \text{on } \partial\Omega.
\end{aligned} \tag{3.14}$$

Theorem 3.1. Let $\tilde{\mu} \in \mathcal{M}(\bar{\Omega})$. Then there exists a unique weak solution $q \in L^2(\Omega)$ of (3.14), i.e.

$$\int_\Omega qAv = \int_{\bar{\Omega}} v d\tilde{\mu} \qquad \forall v \in H^2(\Omega) \text{ with } \sum_{i,j=1}^d a_{ij} v_{x_i} \nu_j = 0 \text{ on } \partial\Omega.$$

Furthermore, q belongs to $W^{1,s}(\Omega)$ for all $s \in (1, \frac{d}{d-1})$. For the finite element approximation $q_h \in X_h$ of q defined by

$$a(v_h, q_h) = \int_{\bar{\Omega}} v_h d\tilde{\mu} \qquad \text{for all } v_h \in X_h.$$

the following error estimate holds:

$$\|q - q_h\| \le C h^{2-\frac{d}{2}} \|\tilde{\mu}\|_{\mathcal{M}(\bar{\Omega})}. \tag{3.15}$$

Proof. A corresponding result is proved in [1] for the case of an operator A without transport term subject to Dirichlet conditions, but the arguments can be adapted to our situation. We omit the details. ∎

We are now prepared to prove our main theorem for the optimal controls.

Theorem 3.2. *Let u and u_h be the solutions of (1.3) and (2.5) respectively. Then*

$$\|u - u_h\|_U + \|y - y_h\|_{H^1} \le C h^{1-\frac{d}{4}}.$$

If in addition $Bu \in W^{1,s}(\Omega)$ for some $s \in (1, \frac{d}{d-1})$ then

$$\|u - u_h\|_U + \|y - y_h\|_{H^1} \le C h^{\frac{3}{2}-\frac{d}{2s}} \sqrt{|\log h|}.$$

Proof. We test (1.5) with u_h, (2.8) with u and add the resulting inequalities. This gives

$$(B^*(p-p_h) - \alpha(u_0 - u_{0,h}) + \alpha(u-u_h), u_h - u)_U \geq 0,$$

which in turn yields

$$\alpha\|u-u_h\|_U^2 \leq \int_\Omega B(u_h-u)(p-p_h) - \alpha\,(u_0-u_{0,h}, u_h-u)_U\,. \tag{3.16}$$

Let $y^h := \mathcal{G}_h(Bu) \in X_h$ and denote by $p^h \in X_h$ the unique solution of

$$a(w_h, p^h) = \int_\Omega (y-y_0)w_h + \int_{\bar\Omega} w_h d\mu \qquad \text{for all } w_h \in X_h.$$

Applying Theorem 3.1 with $\tilde\mu = (y-y_0)dx + \mu$ we infer

$$\|p-p^h\| \leq Ch^{2-\frac{d}{2}}\big(\|y-y_0\| + \|\mu\|_{\mathcal{M}(\bar\Omega)}\big). \tag{3.17}$$

Recalling that $y_h = \mathcal{G}_h(Bu_h), y^h = \mathcal{G}_h(Bu)$ and observing (2.7) as well as the definition of p^h we can rewrite the first term in (3.16)

$$\begin{aligned}
\int_\Omega B(u_h-u)(p-p_h) &= \int_\Omega B(u_h-u)(p-p^h) + \int_\Omega B(u_h-u)(p^h-p_h)\\
&= \int_\Omega B(u_h-u)(p-p^h) + a(y_h-y^h, p^h-p_h) \qquad (3.18)\\
&= \int_\Omega B(u_h-u)(p-p^h) + \int_\Omega (y-y_h)(y_h-y^h) + \int_{\bar\Omega}(y_h-y^h)d\mu - \int_{\bar\Omega}(y_h-y^h)d\mu_h\\
&= \int_\Omega B(u_h-u)(p-p^h) - \|y-y_h\|^2 + \int_\Omega (y-y_h)(y-y^h)\\
&\quad + \int_{\bar\Omega}(y_h-y^h)d\mu + \int_{\bar\Omega}(y^h-y_h)d\mu_h.
\end{aligned}$$

After inserting (3.18) into (3.16) and using Young's inequality we obtain in view of (3.17), (2.1) and (2.6)

$$\begin{aligned}
&\frac{\alpha}{2}\|u-u_h\|_U^2 + \frac{1}{2}\|y-y_h\|^2 \qquad (3.19)\\
&\leq C\big(\|p-p^h\|^2 + \|y-y^h\|^2 + \|u_0-u_{0,h}\|^2\big) + \int_{\bar\Omega}(y_h-y^h)d\mu + \int_{\bar\Omega}(y^h-y_h)d\mu_h\\
&\leq Ch^{4-d} + \int_{\bar\Omega}(y_h-y^h)d\mu + \int_{\bar\Omega}(y^h-y_h)d\mu_h.
\end{aligned}$$

It remains to estimate the integrals involving the measures μ and μ_h. Since

$$y_h - y^h \leq (I_hb-b) + (b-y) + (y-y^h) \qquad \text{in } \bar\Omega$$

we deduce with the help of (1.6)

$$\int_{\bar\Omega}(y_h-y^h)d\mu \leq \|\mu\|_{\mathcal{M}(\bar\Omega)}\left(\|I_hb-b\|_\infty + \|y-y^h\|_\infty\right).$$

Similarly, (2.9) implies

$$\int_{\bar\Omega}(y^h-y_h)d\mu_h \leq \|\mu_h\|_{\mathcal{M}(\bar\Omega)}\left(\|b-I_hb\|_\infty + \|y-y^h\|_\infty\right).$$

Inserting the above estimates into (3.19) and using Lemma 2.4 as well as an interpolation estimate we infer

$$\|u-u_h\|_U^2 + \|y-y_h\|^2 \leq Ch^{4-d} + C\|y-y^h\|_{L^\infty}. \tag{3.20}$$

The estimates on $\|u-u_h\|_U$ now follow from (2.2) and Lemma 2.1 respectively. Finally, in order to bound $\|y-y_h\|_{H^1}$ we note that

$$a(y-y_h, v_h) = \int_\Omega B(u-u_h)v_h$$

for all $v_h \in X_h$, from which one derives the desired estimates using standard finite element techniques and the bounds on $\|u-u_h\|_U$. ∎

Remark 3.3. Let us note that the approximation order of the controls and states in the presence of control and state constraints is the same as in the purely state constrained case, if $Bu \in W^{1,s}(\Omega)$. This assumption holds for the important example $U = L^2(\Omega)$, $B = Id$ and $u_{0_h} = P_h u_0$, with $u_0 \in H^1(\Omega)$ and $P_h : L^2(\Omega) \to X_h$ denoting the L^2–projection, and subsets of the form

$$U_{ad} = \{v \in L^2(\Omega), a_l \leq v \leq a_u \text{ a.e. in } \Omega\},$$

with bounds $a_l, a_u \in W^{1,s}(\Omega)$, since $u_0 \in H^1(\Omega)$, and $p \in W^{1,s}(\Omega)$.

4 Implementation issues

The numerical computation of solutions to problem (2.5) is more involved than in the purely state constrained case, i.e. $U_{ad} \equiv U$. The latter is treated in [6] for the special case $U = L^2(\Omega)$. As already mentioned in Remark 2.3 the purely state constrained problem may be substituted by a finite–dimensional one which yields the same solution u_h. Therefore, common solution techniques for finite–dimensional optimization problems with equality and inequality constraints can be applied for its numerical computation.
In the present situation this is definitely different, since problem (2.5) really is infinite–dimensional, with finitely many state constraints. It is not clear at the first instance, whether an iterative solution algorithm for this problem can be implemented on a computer without further discretization steps, while keeping the management of information overhead bounded with increasing number of iterations. We now sketch that this is indeed possible.
To simplify the exposition we now assume $a_{ij} = \delta_{ij}$, $B = Id$, $U = L^2(\Omega)$ and $u_0 = 0, y_0 = 0$. We rewrite (2.5) as nonlinear, nonsmooth operator equation. To begin with, we define

$$A = \tilde{A} + \tilde{M} := \left[\int_\Omega \nabla\phi_i \nabla\phi_j\right]_{i,j=1,}^m + \left[\int_\Omega \phi_i\phi_j\right]_{i,j=1,}^m \quad U := \left[\int_\Omega u\phi_i\right]_{i=1,}^m \quad B := [b(x_i)]_{i=1,}^m$$

and

$$Y := [y_i]_{i=1,}^m \quad P := [p_i]_{i=1,}^m \quad M := [\mu_i]_{i=1,}^m$$

where

$$\left\{\begin{array}{c} y_h \\ p_h \\ \mu_h \end{array}\right\} = \sum_{i=1}^m \left\{\begin{array}{c} y_i\phi_i \\ p_i\phi_i \\ \mu_i\delta_{x_i} \end{array}\right\}$$

denote the corresponding finite element representations of Y, P and M, respectively, and $u \in L^2(\Omega)$. We recall that (2.8) in the present situation is equivalent to the nonsmooth equation

$$u_h = P_{U_{ad}}\left(-\frac{1}{\alpha}p_h\right),$$

and that the complementarity system in (2.9) can be equivalently rewritten in the form

$$M = \max\{0, M + c(Y - B)\} \text{ component wise,} \tag{4.21}$$

where $c > 0$ is arbitrary. Let us define the mapping $G : \mathbb{R}^m \times \mathbb{R}^m \times L^2(\Omega) \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m \times L^2(\Omega) \times \mathbb{R}^m$ by

$$G(Y, P, u, M) := \begin{bmatrix} AY - U \\ AP - \tilde{M}Y - M \\ u - P_{U_{ad}}\left(-\frac{1}{\alpha}p_h\right) \\ M - \max\{0, M + c(Y - B)\} \end{bmatrix}.$$

Then, the first–order necessary optimality conditions $y_h = \mathcal{G}_h(u)$ together with (2.7)–(2.9) are equivalent to the nonsmooth system

$$G(Y, P, u_h, M) = 0,$$

where u_h denotes the solution of (2.5), and $Y, P,$ and M the nodal representations of the corresponding state y_h, adjoint state p_h, and multiplier μ_h, respectively. Let us apply a generalized Newton step to this system. For this purpose let $[Y, P, u, M]^t \in \mathbb{R}^m \times \mathbb{R}^m \times L^2(\Omega) \times \mathbb{R}^m \to \mathbb{R}^m \times \mathbb{R}^m \times L^2(\Omega) \times \mathbb{R}^m$ be given. Let us define $fmax := \max(0, M + c(Y - B))$ and $dmax := diag\left(\max'(0, M + c(Y - B))\right)$. For the generalized Jacobian of G at $[Y, P, u, M]^t$ we obtain

$$DG(Y, P, u, M) =$$

$$= \begin{bmatrix} A & 0 & -\left[\int\limits_\Omega \bullet\phi_i\right]_{i=1}^m & 0 \\ -\tilde{M} & A & 0 & -Id_M \\ 0 & \frac{1}{\alpha}P'_{U_{ad}}\left(-\frac{1}{\alpha}p_h\right)\bullet & Id_{L^2(\Omega)} & 0 \\ -c\,dmaxId_Y & 0 & 0 & Id_M - dmaxId_M \end{bmatrix},$$

so that one generalized Newton step for the computation of $[Y^n, P^n, u^n, M^n]^t$ amounts to solving

$$DG(Y, P, u, M)\begin{bmatrix} Y^n \\ P^n \\ u^n \\ M^n \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \frac{1}{\alpha}P'_{U_{ad}}\left(-\frac{1}{\alpha}p_h\right)p_h + P_{U_{ad}}\left(-\frac{1}{\alpha}p_h\right) \\ -c\,dmaxY - dmaxM + fmax \end{bmatrix}$$

From the first equation we deduce $Y^n = A^{-1}U^n$, the second yields $M^n = AP^n - \tilde{M}A^{-1}U^n$, and from the third equation we obtain $U^n = -\frac{1}{\alpha}CP^n + R$, where the matrix C and the vector R are defined by

$$C = C(p_h) := \left[\int\limits_{\mathcal{I}(-\frac{1}{\alpha}p_h)} \phi_j\phi_i\right]_{i,j=1}^m \text{ and } R = \frac{1}{\alpha}\left[\int\limits_{\mathcal{I}(-\frac{1}{\alpha}p_h)} p_h\phi_i\right]_{i=1}^m + \left[\int\limits_\Omega P_{U_{ad}}\left(-\frac{1}{\alpha}p_h\right)\phi_i\right]_{i=1}^m.$$

Here, $\mathcal{I}\left(-\frac{1}{\alpha}p_h\right)$ denotes the subset of Ω, where $-\frac{1}{\alpha}p_h$ is inactive. Its description clearly relies on the structure of U_{ad}. Finally, with the substitutions from above the fourth equation leads to

$$\left(A - dmaxA + \frac{1}{\alpha}\left(\tilde{M} + dmax\left(c\,Id_Y - \tilde{M}\right)\right)A^{-1}C\right)P^n =$$
$$= \tilde{M}A^{-1}R + \frac{1}{\alpha}\left(c\,Id_Y - \tilde{M}\right)A^{-1}R - dmaxY - dmaxM + fmax, \quad (4.22)$$

which represents a linear system for the computation of P^n. We conclude, that a step of the generalized Newton method is feasible, iff the matrix

$$A - dmaxA + \frac{1}{\alpha}\left(\tilde{M} + dmax\left(c\,Id_Y - \tilde{M}\right)\right)A^{-1}C$$

is regular. In this case, once P^n is computed, the vectors Y^n, M^n and the function u^n are obtained by performing the corresponding re–substitutions. Let us briefly comment on the cases i.) $U_{ad} \equiv U$ and ii.) $b = \infty$, i.e. no state constraints are present. In the first case the system matrix in (4.22) takes the form

$$A - dmaxA + \frac{1}{\alpha}\left(\tilde{M} + dmax\left(c\,Id_Y - \tilde{M}\right)\right)A^{-1}\tilde{M}$$

and is positive definite if we choose c large enough. In the second case we obtain the system matrix

$$A + \frac{1}{\alpha}\tilde{M}A^{-1}C$$

which is positive definite since $\frac{1}{\alpha}\tilde{M}A^{-1}C$ is positive semi–definite. In both cases the Newton step is feasible. This reflects the fact that in these cases the adjoint variables p, p_h and the multipliers μ, μ_h are unique. Let us finally note that the constant c in (4.21) may also be replaced by an vector $c \in \mathbb{R}^m$ containing positive components $c_i > 0 (i = 1, \ldots, m)$. If one now replaces the mass matrix $\tilde{M}$ by its mass–lumped version $\bar{M}$, the choice $c := diag(\bar{M})$ leads to the system matrix

$$\left(A - dmaxA + \frac{1}{\alpha}\bar{M}A^{-1}C\right)$$

in (4.22).

We note that the adjoint states p, p_h and the multipliers μ, μ_h need not be unique if state and control constraints occur simultaneously. Necessary and sufficient conditions for the uniqueness of multipliers in the presence of cone constraints are provided by Shapiro in [15].

The exposition in this section shows that the infinite dimensional problem (2.5) can numerically be implemented on a computer by a generalized Newton method, say, without further discretization steps, and with keeping the management of information overhead constant in each iteration step. Further details inc. convergence analysis for generalized Newton methods applied to the numerical solution of (2.5) will be given in a forthcoming paper.

Acknowledgements

The authors gratefully acknowledge the support of the DFG Priority Program 1253 entitled Optimization With Partial Differential Equations.

References

[1] Casas, E.: *L^2 estimates for the finite element method for the Dirichlet problem with singular data*, Numer. Math. **47**, 627–632 (1985).

[2] Casas, E.: *Control of an elliptic problem with pointwise state constraints*, SIAM J. Cont. Optim. **24**, 1309–1322 (1986).

[3] Casas, E.: *Boundary control of semilinear elliptic equations with pointwise state constraints*, SIAM J. Cont. Optim. **31**, 993–1006 (1993).

[4] Casas, E.: *Error Estimates for the Numerical Approximation of Semilinear Elliptic Control Problems with Finitely Many State Constraints*, ESAIM, Control Optim. Calc. Var. **8**, 345–374 (2002).

[5] Casas, E., Mateos, M.: *Uniform convergence of the FEM. Applications to state constrained control problems*. Comp. Appl. Math. **21** (2002).

[6] Deckelnick, K., Hinze, M.: *Convergence of a finite element approximation to a state constrained elliptic control problem*, MATH-NM-01-2006, Institut fr Numerischen Mathematik, TU Dresden (2006).

[7] Hintermüller, M., K., Kunisch, K.: *Path following methods for a class of constrained minimization methods in function spaces*, Report RICAM2004-07, RICAM Linz (2004).

[8] Hintermüller, M., K., Kunisch, K.: *Feasible and non-interior path following in constrained minimization with low multiplier regularity*, Report, Universität Graz (2005).

[9] Hinze, M.: *A variational discretization concept in control constrained optimization: the linear-quadratic case*, J. Computational Optimization and Applications **30**, 45-63 (2005).

[10] Hinze, M., Meyer, C.: *Numerical analysis of Lavrentiev-regularized state constrained elliptic control problems*, WIAS Preprint 1168, Berlin (2006).

[11] Meyer, C.: *Error estimates for the finite element approximation of an elliptic control problem with pointwise constraints on the state and the control*, WIAS Preprint 1159 (2006).

[12] Meyer, C., Prüfert, U., Tröltzsch, F.: *On two numerical methods for state-constrained elliptic control problems*, Technical Report 5-2005, Institut für Mathematik, TU Berlin (2005).

[13] Meyer, C., Rösch, A., Tröltzsch, F.: *Optimal control problems of PDEs with regularized pointwise state constraints*, Preprint 14, Institut für Mathematik, TU Berlin, to appear in Computational Optimization and Applications (2004).

[14] Schatz, A.H.: *Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids. I: Global estimates*, Math. Comput. **67**, No.223, 877–899 (1998).

[15] Shapiro, A.: *On uniqueness of Lagrange multipliers in optimization problems subject to cone constraints*, SIAM J. Optim. **7**, 508–518 (1997).

[16] Tröltzsch, F.: *Optimale Steuerung mit partiellen Differentialgleichungen*. Vieweg, 2005.