Als Manuskript gedruckt

Technische Universität Dresden
Herausgeber: Der Rektor

A generalized discretization concept for optimal control problems with control constraints

Michael Hinze

MATH-NM-02-2003
April 2003

Contents

1 Introduction **3**

2 The discrete concept **5**

3 Application to control of elliptic equations **7**

4 Numerical realization **11**

4.1 Discrete solution and numerical algorithms 11

4.2 Numerical example 15

A generalized discretization concept for optimal control problems with control constraints

Michael Hinze
Institut für Numerische Mathematik,
Technische Universität Dresden
D–01069 Dresden, Germany,
email: hinze@math.tu-dresden.de

Abstract

A new discretization concept for optimal control problems with control constraints is introduced which utilizes for the discretization of the control variable the relation between adjoint state and control. Its key feature is not to discretize the space of admissible controls but to implicitly utilize the first order optimality conditions and the discretization of the state and adjoint equations for the discretization of the control. For discrete controls obtained in this way an optimal error estimate is proved. The application to control of elliptic equations is discussed. Finally it is shown that the new concept is numerically implementable with only slight increase in program management. A numerical test confirms the theoretical investigations.

Key words. Error estimates, control of pdes, control constraints
AMS(MOS) subject classification. 65J10, 65K10, 49M05, 49N10.

1 Introduction

Consider the linear quadratic abstract optimal control problem;

$$\min_{(y,u)\in Y\times U} J(y,u) \text{ s.t. } y = Su \text{ and } u \in U_{\mathrm{ad}}, \tag{1}$$

where $U = U^*$ denotes the Hilbert space of controls, Y the state space, $S : U \to Y$ the linear, bounded control-to-state solution operator, and $U_{\mathrm{ad}} \subseteq U$ the convex, closed set of admissible controls. For $\alpha > 0$ further let

$$J(y,u) = \frac{1}{2}\|y - z\|_Z^2 + \frac{\alpha}{2}\|u\|_U^2,$$

where $Z = Z^*$ denotes a Hilbert space, $z \in Z$ and $Y \hookrightarrow Z \hookrightarrow Y^*$. Further let $\hat{J}(u) := J(Su,u)$. With this setting an equivalent formulation of problem (1) then is given by

$$\min_{u\in U_{\mathrm{ad}}} \hat{J}(u). \tag{2}$$

Example 1.1. $Z = L^2(\Omega)$, $S := (-\Delta)^{-1}$, $Y = H_0^1(\Omega)$, $U = L^2(\Omega)$ and $U_{\mathrm{ad}} = \{u \in U; u_a \le u \le u_b \text{ a.e. in } \Omega\}$ fit into the abstract frame presented above, see also the subsequent sections.

The proof of the following Theorem can be obtained by standard techniques and therefore is omitted.

Theorem 1.2. Problem (2) admits a unique solution $u^* \in U_{\rm ad}$, which satisfies the variational inequality

$$(\hat{J}'(u^*), v - u^*)_U \geq 0 \quad \forall\, v \in U_{\rm ad}. \tag{3}$$

There holds $\hat{J}'(u) = \alpha u + S^*(Su - z)$, with S^* denoting the adjoint of S.

In the present work a new discretization concept for the control problem (1) is introduced which is based on the discretization of the state space alone. It guarantees conformability of continuous and discrete admissible sets since these sets in fact coincide. Furthermore, in applications to control of partial differential equations with finite element discretizations the approach decouples the approximation of the active set from the nodes of the finite element grid. In the approach presented discrete controls are defined in terms of the discretized state and co-state variables, i.e. through the discrete optimality condition. **Main result:** Let S_h denote a discretization of S which satisfies Assumption 2.4 below. Further let $u_h^* \in U_{\rm ad}$, $u_h^* := \operatorname{argmin}_{u \in U_{\rm ad}} J(S_h u, u)$. Then

$$\|u^* - u_h^*\|_U = \mathcal{O}(h^2),$$

where u^* denotes the solution of (1). Moreover, in practical applications, such as control of partial differential equations u_h^* is numerically computable with only slight increase in program management compared to the conventional methods.

In the literature the common concept for characterizing and computing approximate solutions to problem (1) is based on discretizations of the set of admissible controls $U_{\rm ad}$. Specifically when S denotes the solution operator of an elliptic partial differential equation and S_h its finite element discretization on a finite element grid with gridsize h the approximation of controls often is also related to this mesh. Typical Ansätze are piecewise constant or piecewise linear, globally continuous controls, see Arada/Raymond/Tröltzsch [1]. The authors for piecewise constant control approximations prove optimal error estimates of the form

$$\|u^* - u_h^*\|_U = \mathcal{O}(h),$$

also show convergence of the same order in L^∞. A close inspection of their proof shows that the order of convergence can not be improved even in the presence of better L^∞ error estimates for the state and co-state variables. These observations extend to piecewise linear, globally continuous control approximations where the error of the control has the size $\|u^* - u_h^*\|_U = \mathcal{O}(h^{\frac{3}{2}})$, see Rösch [12] for a proof under special assumptions on the continuous solutions, compare also Casas/Tröltzsch [4]. These authors propagate the concept *First discretize the control space, then discretize the state space* and emphasize that state space approximation should be adapted to the needs of control space approximation. However, the discretization method in the presence of control constraints no longer is conform in the sense that projections of discrete controls onto the continuous admissible set need not be contained in the discrete admissible set of controls. Furthermore, relating the Ansatz for the controls to the finite element grid only allows to resolve the active set on this grid, which certainly is an additional drawback of the conventional approach. The discrete approach presented here circumvents these shortcomings. Moreover, it yields a discrete control u_h^* which also in the situations sketched above satisfies $\|u^* - u_h^*\|_U = \mathcal{O}(h^2)$ and, most important is computable with only slight increase in program management compared to the conventional methods.

The paper is organized as follows. In the following section the new discretization concept is introduced in the abstract setting. It is applied to linear quadratic control problems for elliptic equations is Section 3. Section 4 is devoted to numerical implementation of the new discretization concept and presents numerical results which confirm the theoretical considerations.

2 The discrete concept

In order to define discrete optimal controls let $S_h : U \to Y_h \subset Y$ be the linear, bounded control-to-discretized state solution operator, where $Y_h \subset Y$ is a finite dimensional subspace equipped with the norm of Y.

Definition 2.1. $u_h^* \in U_{\mathrm{ad}}$ is called discrete optimal control: $\Leftrightarrow$

$$(4) \qquad (\hat{J}_h'(u_h^*), v - u_h^*)_U \geq 0 \quad \forall\, v \in U_{\mathrm{ad}},$$

where $\hat{J}_h(u) := J(S_h u, u)$.

Note that $\hat{J}_h'(u) = \alpha u + S_h^*(S_h u - z)$.

Remark 2.2. In the case of linear quadratic control problems with control constraints (as is the case in the present work) the variational inequality in Definition 2.1 is the necessary and sufficient optimality condition of

$$u_h^* = \arg\min_{u \in U_{\mathrm{ad}}} J(S_h u, u).$$

The definition of optimal controls in terms of the first order optimality condition allows greater flexibility. For example, a more general approximation concept would be given by

$$(5) \qquad (\alpha u_h^* + R_h(S_h u_h^* - z), v - u_h^*)_U \geq 0 \quad \forall\, v \in U_{\mathrm{ad}},$$

where R_h denotes an approximation of the adjoint operator S^*.

Remark 2.3. The key idea of the discretization concept becomes more transparent in the case $U_{\mathrm{ad}} = U$. Then the discrete optimal control u_h^* necessarily satisfies

$$u_h^* = -\frac{1}{\alpha} S_h^*(S_h u_h^* - z),$$

where the right hand side lives in the space U but is a discrete object. In fact the discrete control $u_h^* \in U$ is implicitly discretized in terms of the discretized adjoint operator and therefore is discrete. In the following section S_h denotes the solution operator of a finite element discretization with linear elements of an elliptic equation. The control u_h^* in that case is a continuous, piecewise linear finite element function. In Section 4 it is shown that this concept in the presence of control constraints also allows an efficient numerical implementation.

In order to prove the main theorem it is assumed that

Assumption 2.4. S_h, S_h^* satisfy

- $\|(S^* - S_h^*)z\|_U \leq Ch^2\|z\|_Z$, and
- $\|(S^*S - S_h^*S_h)u^*\|_U \leq Ch^2\|u^*\|_U$.

Theorem 2.5. For $h > 0$ small enough the variational inequality (4) admits a unique solution $u_h^* \in U_{\mathrm{ad}}$, which satisfies

$$\|u^* - u_h^*\|_U \leq Ch^2\{\|u^*\|_U + \|z\|_Z\} \tag{6}$$

Here, $u^* \in U_{\mathrm{ad}}$ denotes the unique solution of problem (1).

Proof: Existence of u_h^* follows from the fact that $u_h^* = \mathrm{argmin}_{u \in U_{\mathrm{ad}}} J_h(S_h u, u)$. To argue uniqueness let $u_1 \neq u_2$ two distinguish solutions to (4). Utilizing u_i as test control for $u_j, (i, j = 1, 2)$ and adding the corresponding inequalities one gets

$$(\hat{J}_h'(u_1) - \hat{J}_h'(u_2), u_2 - u_1)_U \geq 0.$$

This implies

$$\begin{aligned} \alpha\|u_1 - u_2\|_U^2 &\leq (S_h^*(S_h u_1 - z) - S_h^*(S_h u_2 - z), u_2 - u_1)_U \\ &\qquad = -\|S_h(u_1 - u_2)\|_Y^2, \end{aligned}$$

and thus $u_1 = u_2$. To prove the error estimate note that $u_h^* \in U_{\mathrm{ad}}$ and therefore is admissible as test function for the variational inequality (3) for u^*. Analogously to the proof of uniqueness one has

$$(\hat{J}'(u^*) - \hat{J}_h'(u_h^*), u_h^* - u^*) \geq 0.$$

Straightforward estimation yields

$$\begin{aligned} \alpha\|u^* - u_h^*\|_U^2 &\leq (S^*(Su^* - z) - S_h^*(S_h u_h^* - z), u_h^* - u^*)_U \\ &= ((S^*S - S_h^*S_h)u^* + (S_h^* - S^*)z + \underbrace{S_h^*S_h(u^* - u_h^*), u_h^* - u^*}_{-\|S_h(u^* - u_h^*)\|_Y^2})_U \leq \\ &\qquad ((S^*S - S_h^*S_h)u^* + (S_h^* - S^*)z, u_h^* - u^*)_U \leq \\ &\qquad\qquad Ch^2\{\|u^*\|_U + \|z\|_Z\}\|u^* - u_h^*\|_U. \end{aligned}$$

Assumption 2.4 and $\alpha > 0$ finally give

$$\|u^* - u_h^*\|_U^2 \le Ch^2\{\|u^*\|_U + \|z\|_Z\}.$$

From the considerations above one may also infer some kind of H^{-1}-*estimate* for the error in the controls in the case $\alpha = 0$.

Corollary 2.6. Let $\alpha = 0$ and let the assumptions of Theorem 2.5 be satisfied. Assume further that $U_{\rm ad}$ is also bounded and that $\|(S - S_h)v\|_Y \le Ch\|v\|_U$ for all $v \in U$. Then problems (3) and (4) admit unique solutions u^* and u_h^*, respectively which satisfy the estimate

$$\|S(u^* - u_h^*)\|_Y^2 \le Ch^2 \left\{ (\|u^*\|_U + \|z\|_Z) \|u^* - u_h^*\|_U + \|u^* - u_h^*\|_U^2 \right\}. \tag{7}$$

Proof: Existence and uniqueness of solutions may be argued as in the case $\alpha > 0$. To obtain the error estimate observe that $\|S(u^* - u_h^*)\|_Y^2 \le 2\|(S - S_h)(u^* - u_h^*)\|_Y + 2\|S_h(u^* - u_h^*)\|_Y^2$. The first addend due to the assumptions is bounded by $Ch^2\|(u^* - u_h^*)\|_U^2$. The estimate for the second addend can be deduced from the proof of the previous theorem.

3 Application to control of elliptic equations

The approach presented in the previous section applies to a large class of control problems for partial differential equations [6, 9]. In the present section its application to control problems for linear elliptic partial differential equations is shown.

To begin with let $\Omega \subset \mathbb{R}^n$ (n=1,2,3) be a bounded, open convex domain with (a) $C^{1,1}$ or (b) polygonal boundary. Further let $A = \{a_{i,j}\}_{i,j=1,n} \in H^{1,\infty}(\Omega)^{3,3}$ be a symmetric matrix which is uniformly positive definite on $\bar\Omega$, i.e.

$$\xi^t A(x)\xi \ge c_e|\xi|^2 \text{ for all } \xi \in \mathbb{R}^n$$

with some positive constant c_e independent of $x \in \bar\Omega$. Further let $c_0 \in L^\infty(\Omega)$, $c_0 \ge 0$. Let $Y := H_0^1(\Omega)$. Then for $f \in Y^* = H^{-1}(\Omega)$ the action of the solution operator $S : Y^* \to Y$ is defined through

(8) $y = Sf :\Longleftrightarrow$

$$a(y, v) := \int_\Omega A\nabla y\nabla v + c_0\, y\, v dx = \langle f, v\rangle_{Y^*,Y} \text{ for all } v \in Y.$$

The solution operator S is the bounded inverse of a self-adjoint operator. For him standard existence and regularity theory [7, 8] yields

Proposition 3.1. For every $f \in Y^*$ (8) admits a unique solution $y \in Y$ which depends continuously on the data f, i.e. with some positive constant C there holds

$$\|y\|_Y \le C\|f\|_{Y^*}.$$

If moreover $f \in L^2(\Omega)$, then $y \in Y \cap H^2(\Omega) \cap H_0^1(\Omega)$ and

$$\|y\|_2 \le C\|f\|_0.$$

If in case (a) $f \in L^\infty(\Omega)$, then $y \in Y \cap H^{2,p}(\Omega) \cap H_0^1(\Omega)$ for all $p > n$ and

$$\|y\|_{2,p} \le C_p\|f\|_\infty,$$

see [3].

In order to formulate the control problem further let $z \in Z := L^2(\Omega)$, $U = L^2(\Omega) = U^*$ and $U_{\rm ad} = \{u \in U; u_a \le u \le u_b \text{ a.e. in } \Omega\}$, where $u_a \le u_b$ denote constants. Note that $U_{\rm ad}$ is only weakly closed in U. Clearly, $U \subset Y^*$ so that the definition of S in problem (1) makes sense.

In order to apply the discrete concept introduced in Section 2 to the control problem

$$\min_{u \in U_{\rm ad}} \hat{J}(u) = J(Su, u) \tag{9}$$

it remains to supply a discretization concept for the operator S. Here standard finite element discretizations of S with piecewise linear, globally continuous Ansatz functions on uniformly stable sequences $\{\tau_h\}_h$ of uniformly quasi-uniform triangulations are considered, where $h := \max\{\text{diam}(T); T \in \tau_h\}$. One has in case (b) $\Omega := \cup_{T\in\tau_h} T$ for all h. In case (a) it is assumed that all boundary vertices of $\Omega_h := \cup_{T\in\tau_h} T \subseteq \Omega$ are contained in $\partial\Omega$ so that $|\Omega \setminus \Omega_h| = \mathcal{O}(h^2)$. Now set

$$Y_h := \{v_h \in C(\bar{\Omega}), v_{h_{|_T}} \text{linear for all } T \in \tau_h, v_h = 0 \text{ in } \bar{\Omega} \setminus \Omega_h\}$$

For $f \in Y^*$ the action of S_h is defined by

$$y_h = S_h f :\Longleftrightarrow \tag{10}$$

$$a(y_h, v_h) := \int_\Omega A\nabla y_h \nabla v_h + c_0\, y_h\, v_h dx = \langle f, v_h\rangle_{Y^*,Y} \text{ for all } v_h \in Y_h.$$

The discrete solution operator S_h is selfadjoint. Moreover, for $f \in Y^*$ the function $y_h \in H^1(\Omega)$ and $\|y_h\|_1 \le C\|f\|_{Y^*}$. The following error estimates are are well known and can be found for example in [5].

Proposition 3.2. Let $f \in L^2(\Omega)$. Further denote by y, y_h the unique solutions of

$$a(y, v) = \int_\Omega f v dx \text{ for all } v \in Y, \quad a(y_h, v_h) = \int_\Omega f v_h dx \text{ for all } v_h \in Y_h.$$

Then $y \in H^2(\Omega) \cap Y$ and

$$\|y - y_h\|_0 + h\|\nabla y - \nabla y_h\|_0 \le Ch^2\|f\|_0.$$

The discrete analogon to problem (9) reads

$$\min_{u \in U_{\mathrm{ad}}} \hat{J}_h(u) = J(S_h u, u), \tag{11}$$

and admits a unique solution $u_h^* \in U_{\mathrm{ad}}$, which satisfies the variational inequality (4). Recall that problem (9) also admits a unique solution $u^* \in U_{\mathrm{ad}}$ which satisfies (3). Since with the previous proposition the requirements of Theorem 2.5 are satisfied for the setting of this section one immediately gets

Theorem 3.3. Let $u^*, u_h^* \in L^2(\Omega)$ denote the solutions to (9) and (11), respectively. Then, for $h > 0$ small enough there holds

$$\|u^* - u_h^*\|_0 \le Ch^2\{\|u^*\|_0 + \|z\|_0\} \tag{12}$$

where C denotes a positive constant independent of h.

Remark 3.4. It should be noted that for piecewise constant control approximations only $\|u^* - u_h^*\|_0 = \mathcal{O}(h)$ and for piecewise linear, continuous approximations only $\|u^* - u_h^*\|_0 = \mathcal{O}(h^{3/2})$ can be expected. The latter is observed in numerical experiments. The estimate of Theorem 3.3 therefore improves these results for the unique solution u_h^* of problem (11). To anticipate discussion note that this control in fact is computable. For example the projected gradient method for every initial control $u^0 \in U_{\mathrm{ad}}$ yields a sequence $\{u^k\}_k \subset U_{\mathrm{ad}}$, where

$$u^k = P_{[u_a,u_b]}\{u^{k-1} - \rho \hat{J}_h'(u^{k-1})\}.$$

Here, $P_{[u_a,u_b]}$ denotes the projection onto the admissible set U_{ad} and $\rho > 0$ the stepsize of the gradient step. Now let $\rho = \frac{1}{\alpha}$. Then $u^{k-1} - \rho \hat{J}_h'(u^{k-1}) = S_h^*(S_h u^{k-1} - z)$ and the number of connected components of the active set defined by $S_h^*(S_h u^{k-1} - z)$ on each element of the triangulation is bounded from above by three. Moreover, this number can not increase with the iteration counter. Therefore u^k is computable for every iteration index k with the same amount of computational overhead. More details are given in the next section.

In order to provide error estimates in the L^∞ norm the following discrete Sobolev embeddings are utilized. A prove for example can be found in [14], see also [13] for the case $n = 2$.

Proposition 3.5. Let τ_h denote a quasi-uniform, regular triangulation of $\Omega \subset \mathbb{R}^n$ ($n = 1, 2, 3$). Then for every piecewise linear, continuous finite element function $v_h \in H_0^1(\Omega)$ there holds

$$\|v_h\|_\infty \le C \left\{ \begin{array}{c} 1 \\ |\ln h|^{\frac{1}{2}} \\ h^{-\frac{1}{2}} \end{array} \right\} |v_h|_1 \text{ for } \left\{ \begin{array}{c} n = 1 \\ n = 2 \\ n = 3 \end{array} \right\}, \tag{13}$$

where $C > 0$ is a generic constant and $|\cdot|_1$ denotes the H^1 semi-norm.

Theorem 3.6. Let $z \in L^2(\Omega)$ and let u^*, u_h^* denote the solutions of problems (9) and (11), respectively. Then there holds

$$
(14) \quad \|u^* - u_h^*\|_\infty \le C \left\{ \|(S^* - S_h^*)z\|_\infty + \|(S^* - S_h^*)Su^*\|_\infty + \left\{ \begin{array}{c} h^2 \\ h^2|\ln h|^{\frac{1}{2}} \\ h^{\frac{3}{2}} \end{array} \right\} \|u^*\|_0 \text{ for } \left\{ \begin{array}{c} n=1 \\ n=2 \\ n=3 \end{array} \right\} \right\}.
$$

Proof: Let $\mu^* := S^*(Su^* - z)$ and $\mu_h^* := S_h^*(S_h u_h^* - z)$ denote the Lagrange multipliers associated to u^*, u_h^*. Now write $\mu^* - \mu_h^* = S^*Su^* - S_h^*S_h u_h^* + (S_h^* - S^*)z$. Since U_{ad} is defined through box constraints one gets

$$
\|u^* - u_h^*\|_\infty \le \frac{1}{\alpha}\|\mu^* - \mu_h^*\|_\infty \le \frac{1}{\alpha}\left\{\|(S^* - S_h^*)Su^*\|_\infty + \|(S^* - S_h^*)z\|_\infty + \|S_h^*Su^* - S_h^*S_hu^*\|_\infty + \|S_h^*S_hu^* - S_h^*S_hu_h^*\|_\infty\right\}.
$$

To estimate the third and fourth addend utilize Proposition 3.5. For the third addend one gets in the case $n = 2$

$$
\|S_h^*Su^* - S_h^*S_hu^*\|_\infty \le C|\ln h|^{\frac{1}{2}}|S_h^*Su^* - S_h^*S_hu^*|_1 \le C|\ln h|^{\frac{1}{2}}\|Su^* - S_hu^*\|_0 \le C|\ln h|^{\frac{1}{2}}h^2\|u^*\|_0.
$$

Similarly for the fourth addend

$$
\|S_h^*S_hu^* - S_h^*S_hu_h^*\|_\infty \le C|\ln h|^{\frac{1}{2}}|S_h^*S_hu^* - S_h^*S_hu_h^*|_1 \le C|\ln h|^{\frac{1}{2}}\|u^* - u_h^*\|_0 \le C|\ln h|^{\frac{1}{2}}h^2\{\|z\|_0 + \|u^*\|_0\},
$$

where Theorem 3.3 was utilized. The exposition for the cases $n = 1, 3$ is similar. This completes the proof. $\square$

Remark 3.7. To finalize the L^∞ error estimate for $u^* - u_h^*$ it remains to provide estimates for $e1 := \|(S^* - S_h^*)Su^*\|_\infty$ and $e2 := \|(S^* - S_h^*)z\|_\infty$. However, the approximation order for these terms is restricted by 2. In this sense estimate (14) is optimal. For example there holds

$e_i = \mathcal{O}(h^{2-\frac{n}{2}})$, $i = 1, 2$, $n = 1, 2, 3$,

$e_i = \mathcal{O}(h)$, $i = 1, 2$, $n = 2, 3$, if a discrete maximum principle is satisfied for the finite element spaces, and

$e_i = \mathcal{O}(h^2|\ln h|)$, $i = 1, 2$, $n = 1, 2$, if $Su^*, S^*z \in H^{2,\infty}(\Omega)$,

see [5].

4 Numerical realization

In the present section the discretization concept is described in more detail. Firstly a characterization of the discrete solution of problem (9) and the related discrete active set is provided. Then two numerical algorithms are discussed which allow to implement the discretization strategy proposed in the present work. Finally a numerical example is presented which confirms the theoretical investigations.

4.1 Discrete solution and numerical algorithms

In order to start the exposition of this section recall that U_{ad} is defined through constant box constraints with bounds $u_a < u_b$. Further recall that $J(y,u) = \frac{1}{2}\int_\Omega |y-z|^2 dx + \frac{\alpha}{2}\int_\Omega |u|^2 dx$, so that for given $u \in \mathcal{U}$ there holds $\hat{J}'(u) = \alpha u + p$, where $p = S^*(Su - z)$. To anticipate discussion note, that the subsequent considerations are also valid for bounds which are linear on every simplex of the triangulation. They may also be extended to the case of smooth spatially varying bounds u_a, u_b and to more general cost functionals. Since the latter cases only introduce technical difficulties their exposition is omitted.

Theorem 4.1. Let $u_h^* \in U_{\text{ad}}$ denote the unique solution to (P_h). Then there exists a partition $\kappa_h = \{K_1, \dots K_{l(h)}\}$ of Ω such that u_h^* restricted to K_j $(j = 1, \dots, l(h))$ is a polynomial either of degree zero or one. For $l(h)$ there holds

$$l(h) \le Cnt(h),$$

with a positive constant $C \le 3$ and $nt(h)$ denoting the number of simplexes in τ_h. In particular, the vertices of the discrete active set associated to u_h^* need not coincide with finite element nodes.

Proof: The solution u_h^* ôf (P_h) satisfies

$$u_h^* = P_{[u_a,u_b]}\{-\frac{1}{\alpha}\mu_h^*\}, \tag{15}$$

compare Fig. 1. Now abbreviate $\xi_h^a := -\frac{1}{\alpha}\mu_h^* - u_a$, $\xi_h^b := u_b - \frac{1}{\alpha}\mu_h^*$ and investigate the zero level sets 0_h^a and 0_h^b of ξ_h^a and ξ_h^b, respectively.
Case $n = 1$: $0_h^a \cap T_i$ is either empty or a point $S_i^a \in T_i$. Every point S_i^a subdivides T_i into two sub-intervals. Analogously $0_h^b \cap T_i$ is either empty or a point $S_i^b \in T_i$. Further $S_i^a \ne S_i^b$ since $u_a < u_b$. The maximum number of sub-intervals of T_i induced by 0_h^a and 0_h^b therefore is equal to three. Therefore, $l(h) \le 3nt(h)$, i.e. $C = 3$.
Case $n = 2$: $0_h^a \cap T_i$ is either empty or a vertex of τ_h or a line $L_i^a \subset T_i$, analogously $0_h^b \cap T_i$ is either empty or a vertex of τ_h or a line $L_i^b \subset T_i$. Since $u_a < u_b$ the lines L_i^a and L_i^b do not intersect. Therefore, similar considerations as in the case $n = 1$ yield $C = 3$.
Case $n = 3$: $0_h^a \cap T_i$ is either empty or a vertex of τ_h or an edge or a part of a plane $L_i^a \subset T_i$, analogously $0_h^b \cap T_i$ is either empty or a vertex or an edge of τ_h or a part of a plane $L_i^b \subset T_i$.

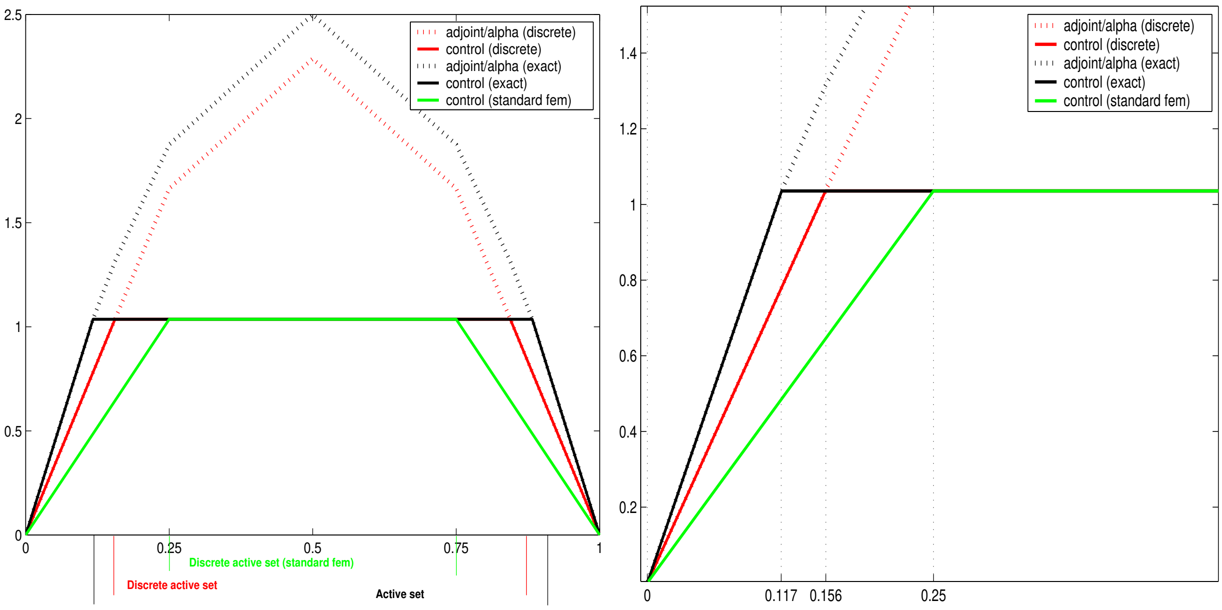

Figure 1: Solutions u^* and u_h^* together with their active sets for $h = 1/4$ in the case $n = 1$ (left). Zoom of the same (right). The decoupling of discrete active set and finite element grid is obvious.

Since $u_a < u_b$ the surfaces L_i^a and L_i^b do not intersect. Therefore, similar considerations as in the case $n = 2$ yield $C = 3$. This completes the proof.

It follows now from the considerations above that commonly utilized convergent solution algorithms for problem (P_h) with discrete initial controls generate sequences of discrete iterates. In the following this is illustrated for the projected gradient method formulated next, and also for a special variant of the primal-dual active set method [2, 10, 11].

Algorithm 4.2. (Projected gradient algorithm)

1. $\tilde{u}^0 \in U$, $u^0 := P_{[u_a,u_b]}\tilde{u}^0$, k = 0, tol > 0.
2. $d^k := \hat{J}_h'(u^k)$
3. $\rho = \text{argmin}_{s \geq 0} J_h(u^k - sd^k)$
4. $u^{k+1} := P_{[u_a,u_b]}(u^k - \rho d^k)$
5. If $|u^{k+1} - u^k| \leq$ tol, STOP
6. k = k+1, goto 2.

A close inspection of this algorithm shows that

$$u^k - \rho d^k = -\rho S_h^*(S_h u^k - z)$$

in step 4. if $\rho = \frac{1}{\alpha}$. Since $S_h^*(S_h u^k - z)$ is a finite element function the determination of its active part follows the lines of the proof of the previous Theorem 4.1. This in turn proves

Corollary 4.3. Let τ_h denote a triangulation of Ω and let $u^0 \in U$ such that $u^0_{|T_i}$ for all $T_i \in \tau_h$ is either a constant or a linear function. Further set $\rho = \frac{1}{\alpha}$ in step 3. of Algorithm 4.2. Then for every iterate u^k of Algorithm 4.2 there exists a partition $\kappa_h^k = \{K_1^k, \dots K_{l(h)}^k\}$ of Ω such that u^k restricted to K_j^k $(j = 1, \dots, l(h))$ is a polynomial either of degree zero or one. The number $l(h)$ is independent of the iteration index k and satisfies the estimate

$$l(h) \leq Cnt(h)$$

with a positive constant $C \leq 3$, where $nt(h)$ again denotes the number of simplexes in τ_h.

It is now easy to verify that with the settings of Corollary 4.3 the numerical overhead for Algorithm 4.2 compared to the standard approaches (a-priori discretization of $\mathcal{U}_{ad}$) consists in managing varying grid points similar to S_i^a, S_i^b in the proof of Theorem 4.1. Of course these grid points depend on the iteration counter k of Algorithm 4.2 and may alter with every iteration of the algorithm, but there are at most two such points on every edge of a finite element simplex.

Next investigate the primal-dual active set strategy [2, 10, 11]. Its algorithmical realization relies on the numerical treatment of the complementary system associated with problem (1). Both, the state equation $y = Su$ and the control constraints $u \in U_{\text{ad}}$ are considered as hard constraints and are supplied with Lagrange multipliers. For the discrete optimal control problem (11) considered in the present work the complementarity system reads

$$(\alpha Id + S_h^* S_h)u + \hat{\lambda} = S_h^* z =: -r,$$

$$\Psi(u, \hat{\lambda}) := \max(\hat{\lambda} + \sigma(u - u_b), 0) + \min(\hat{\lambda} + \sigma(u - u_a), 0) = \hat{\lambda},$$

where $\hat{\lambda} = -\lambda_a + \lambda_b$ with λ_a and λ_b denoting the Lagrange multipliers associated with the box constraints, and $\sigma > 0$ is arbitrary. For the numerical solution of this complementarity system the primal-dual active set strategy works as follows.

Algorithm 4.4. (Primal-dual active set strategy)

1. Initialize $u^0 \in U_{\text{ad}}$, $\hat{\lambda}^0 = -r$; set $l = 1$, $\epsilon > 0$ small.

2. Loop l

   (a) $\mathcal{A}_{u_a}^l := \{\hat{\lambda}^{l-1} + \sigma(u^{l-1} - u_a) < 0\}$ $(= \{-r - S_h^* S_h u^{l-1} - \alpha u_a < 0\}$, if $\sigma = \alpha)$,

   (b) $\mathcal{A}_{u_b}^l := \{\hat{\lambda}^{l-1} + \sigma(u^{l-1} - u_b) > 0\}$ $(= \{-r - S_h^* S_h u^{l-1} - \alpha u_b > 0\}$, if $\sigma = \alpha)$,

   (c) $\mathcal{I}^l := \Omega \setminus (\mathcal{A}_{u_a}^l \cup \mathcal{A}_{u_b}^l)$.

3. $l \geq 2$, $\mathcal{A}_{u_a}^l = \mathcal{A}_{u_a}^{l-1}$, $\mathcal{A}_{u_b}^l = \mathcal{A}_{u_b}^{l-1}$, or $\|\Psi(u^{l-1}, \hat{\lambda}^{l-1}) - \hat{\lambda}^{l-1}\| \leq \epsilon$: $u = u^{l-1}$, $\hat{\lambda} = \hat{\lambda}^l$, RETURN.

4. Otherwise

   (a) $u^l = u_a$ on $\mathcal{A}^l_{u_a}$, $u^l = u_b$ on $\mathcal{A}^l_{u_b}$, $\hat{\lambda}^l = 0$ on $\mathcal{I}^l$

   (b) Solve for $u^l|_{\mathcal{I}^l}$, $\hat{\lambda}^l|_{\mathcal{A}^l_{u_a}\cup\mathcal{A}^l_{u_b}}$

   $(\alpha + S_h^* S_h)u^l + \hat{\lambda}^l = -r$

5. $l := l + 1$.

To solve the system in step 4. (a)(b) of the algorithm the conjugate gradient method of Algorithm 4.5 may be utilized. Let $H := \alpha Id + S_h^* S_h$

Algorithm 4.5. (CG to solve 4.)

1. Initialize

$$v^0_{|\mathcal{I}^l} := 0, \quad v^0_{|\mathcal{A}^l_{u_a}} = u_{a_{|\mathcal{A}^l_{u_a}}}, \quad v^0_{|\mathcal{A}^l_{u_b}} = u_{b_{|\mathcal{A}^l_{u_b}}},$$
$$d^0_{|\mathcal{I}^l} = r_{|\mathcal{I}^l} - (Hv^0)_{|\mathcal{I}^l} =: g^0_{|\mathcal{I}^l}, \quad k := 0.$$

2. Do until convergence

$$\begin{array}{ll}
(a) & t_k := \frac{\|g^k_{|\mathcal{I}^l}\|^2}{(d^k, Hd^k)}, \\
(b) & v^{k+1} = v^k + t_k d^k, \\
(c) & g^{k+1}_{|\mathcal{I}^l} = g^k_{|\mathcal{I}^l} + t_k (Hd^k)_{|\mathcal{I}^l}, \\
(d) & \beta_k = \frac{\|g^{k+1}_{|\mathcal{I}^l}\|^2}{\|g^k_{|\mathcal{I}^l}\|^2}, \\
(e) & d^{k+1}_{|\mathcal{I}^l} = -g^{k+1}_{|\mathcal{I}^l} + \beta_k d^k_{|\mathcal{I}^l}, \\
(f) & d^{k+1}_{|\mathcal{A}^l} = 0, \\
(g) & k = k + 1.
\end{array}$$

3. Compute $\hat{\lambda}^l = -r - Hv^k$.

As already observed for the projected gradient method of Algorithm 4.2 the particular choice of a parameter, here namely $\sigma = \alpha$ allows the numerical implementation of this algorithm with only slightly increased numerical overhead when compared to the implementation of the conventional discretization approaches. Indeed, for $\sigma = \alpha$ in step 2. (a),(b) of the algorithm the active set in iteration step l can be identified by means of the finite element function $-r - S_h^* S_h u^{l-1} - \alpha u_a$, so that on each simplex of the triangulation the active set in step l contains at most two connected components. These components again may be managed by varying grid points similar to S_i^a, S_i^b in the proof of Theorem 4.1. Again note that these grid points may alter with every iteration of Algorithm 4.4, but there are at most two such points on every edge of an finite element simplex.

Both algorithms presented perfectly mimic the decoupling of the discrete active set and the finite element grid. As the following numerical example illustrates the approximation of the active set with the method presented in the present work already is very accurate for coarse discretizations. Moreover, the boundary of the discrete active set seems to converge quadratically to that of the continuous active set.

4.2 Numerical example

Consider the following optimal control problem of tracking type;

$$\min_{y,u} \frac{1}{2}\int_0^1 |y-z|^2 dx + \frac{\alpha}{2}\int_0^1 u^2 dx,$$

where $u \in U_{\text{ad}} := \{v \in L^2(0,1); u \leq u_b\}$ and y and u are related via the second order boundary value problem

$$y = Su \Longleftrightarrow -y'' + y = u + e \text{ in } (0,1), \quad y(0) = y(1) = 0.$$

Note that in the setting of the previous sections $u_a = -\infty$. With the choice of $e := -2 + x^2 - x - \min\left(u_b, -\frac{x^2-x}{\alpha}\right)$ and $z(x) \equiv 2$ the unique optimal solution of the control problem is given by

$$y(x) = x^2 - x, \text{ and } u(x) = \min\left(-\frac{x^2-x}{\alpha}, u_b\right),$$

where the associated Lagrange multiplier of the equality constraint is given by $p(x) = x^2 - x$.

The solution operator S is discretized with piecewise linear, continuous finite elements on an equidistant grid $x_i = ih$ with grid width $h = \frac{1}{n+1}$. The related discrete solution operator is denoted by S_h. In all numerical computations presented $\tilde{u}^0 := 0$, $\alpha = 0.1$ and $u_b = \frac{1}{2}(\sqrt{2} - 1.)/(2\alpha)$, so that the boundary points $p^l = \frac{1}{2}(1 - \sqrt{1 - 4\alpha u_b})$ and $p^r = \frac{1}{2}(1 + \sqrt{1 - 4\alpha u_b})$ of the continuous active set can never coincide with a point of the finite element grid. As numerical solution algorithm the projected gradient method of Algorithm 4.2 is utilized, where step 3. is replaced by $\rho = \frac{1}{\alpha}$. The iteration is stopped if the relative difference of two consecutive iterates and the distance of two consecutive active sets is smaller than $1.e-6$. The method for this termination criterion converges after five iterations, where this number is independent of the finite element grid size. Recall that for this choice of ρ the projection step 4. consists of projecting on each element the finite element function $-\rho S_h^*(S_h u^k - z)$ onto the admissible set.

In Fig. 2 (left) a comparison is shown of the discrete control on a grid containing 5 points and the exact solution. As further reference to the performance of the generalized discretization concept also the numerical solution is shown which is obtained by the projected gradient method, where the projection in step 4. is replaced by $u^{k+1} = Q_{[u_a,u_b]}(u^k - \rho d^k)$. Here $v = Q_{[u_a,u_b]}(w)$ iff $v \in U_{\text{ad}}$ is the finite element function with nodal values $v_i = \max\{u_a, \min\{w(x_i), u_b\}\}$. To anticipate discussion note that for both discrete methods numerically quadratic convergence of the error in both the L^2 and L^∞ norm is observed, see the tables below. However, the overall size of the errors on the different refinement levels is much smaller for the generalized approach. The right picture in Fig. 2 shows a zoom on the controls at the right contact point p^r for the numerical solutions obtained on a grid with 5 points. As one can see the generalized method approximates

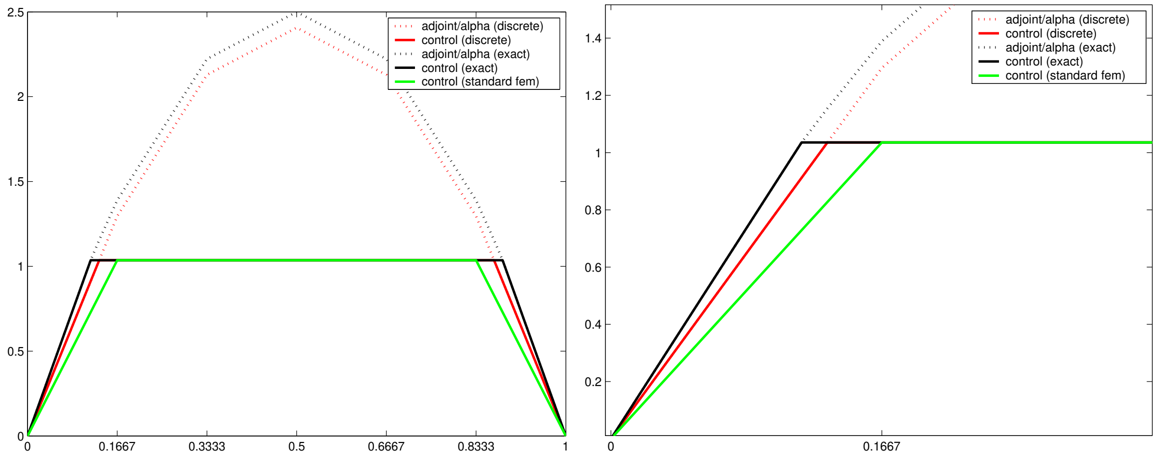

Figure 2: Comparison of optimal continuous control (black), control obtained from the generalized discretization concept (red) and by nodal projections (green) for 5 grid points (left), and zoom of the same at the right contact point (right).

the exact control already pretty well. The figures illustrate that the numerical solution obtained by the generalized discrete concept already on this coarse grid provides a very good approximation of the continuous solution. Moreover, the active set is approximated very accurately as well, see also the tables below illustrating the convergence behavior of the methods. The numerical solution obtained with projection $Q_{[u_a,u_b]}$ only is capable of resolving the active set on the nodes of the finite element grid which results in linear convergence of the active set only. For the generalized approach numerically quadratic convergence is observed.

The experimental order of convergence for positive error functionals $E(h)$ with $h > 0$ in this context is defined as follows: Given two grid sizes $h_1 \neq h_2$, let

$$\mathrm{EOC} := \frac{\ln E(h_1) - \ln E(h_2)}{\ln h_1 - \ln h_2}.$$

It follows from this definition that if $E(h) = \mathcal{O}(h^\xi)$ $(h \to 0)$ then EOC $\approx \xi$. The error functionals investigated in the present section are given by

$$E_2(h) := \|u - u_h\|_0, \quad , E_\infty(h) := \|u - u_h\|_\infty \text{ and } E_a(h) := dist(p^r, p_h^r).$$

Table 1 shows the values of the error functionals for the generalized discretization approach for $\alpha = 0.1$. The functional value and the values of the error functionals for the numerical method with projection $Q_{[u_a,u_b]}$ on the finest grid for u^* replaced by its finite element interpolation are $J = 2.39486559$, $E_2 = 6.278354e-08$, $E_\infty = 1.98161808e-07$ and $E_a = 5.7761306e-05$. In particular, compared to the generalized approach the error for E_a is three orders of magnitude larger. Table 2 shows the experimental order of convergence for the generalized approach in the case $\alpha = 0.1$. As one can see the error estimates (12)

Grid size h	$J(y_h^*, u_h^*)$	E_2	E_∞	E_a
1/9	2.11076146	8.3251513e-3	3.34160317e-2	6.8213674e-3
1/17	2.24179789	3.887878e-3	1.02853562e-2	1.8477263e-3
1/33	2.31563168	1.11582392e-3	2.87882858e-3	3.76347497e-4
1/65	2.35465602	3.38640613e-4	7.67719593e-4	1.56126651e-4
1/129	2.37488443	9.27430945e-05	1.97963645e-4	4.06101135e-05
1/257	2.3851495	2.39257839e-05	5.02863055e-05	1.04427161e-05
1/513	2.39031998	6.07732619e-06	1.26785945e-05	2.69926219e-06
1/1025	2.39291475	1.52991372e-06	3.17714837e-06	7.03630621e-07
1/2049	2.39421448	3.84411362e-07	7.97616384e-07	1.81939892e-07
1/4097	2.39486495	9.61490376e-08	1.99034055e-07	4.00552543e-08

Table 1: Values of cost and error functionals on different levels, $\alpha = 0.1$

and (14) for the errors in the L^2 and L^∞ norms, respectively are numerically confirmed. Moreover, the discrete active set numerically converges quadratically to its continuous counterpart. Note that with projection $Q_{[u_a,u_b]}$ the numerically observed convergence of the L^2 and the L^∞ norm is also quadratic. However, the numerically observed rate of convergence of the discrete active set only is linear, compare the rightmost column in Table 2. Further note that with increasing α the experimental order of convergence for the L^2 and L^∞ norms of the method with projection $Q_{[u_a,u_b]}$ deteriorates, whereas the numerically observed convergence of the generalized approach is robust w.r.t. α.

Grid size h	E_2	E_∞	E_a	E_a^Q
1/9	-	-	-	-
1/17	1.09849354	1.69994866	1.8843099	0.7445453
1/33	1.80087357	1.83703796	2.29561379	0.730538903
1/65	1.72028247	1.90683052	1.26934855	2.93261382
1/129	1.86844323	1.95534396	1.94280586	-0.564329479
1/257	1.95467367	1.97699804	1.95934205	1.02104377
1/513	1.97705756	1.98777074	1.95185998	1.06100344
1/1025	1.98998643	1.99659045	1.93967495	1.14040453
1/2049	1.99272741	1.99396551	1.95135637	1.33634283
1/4097	1.99930666	2.00267975	2.18339849	2.08111609

Table 2: Experimental order of convergence, $\alpha = 0.1$

Acknowledgement

The author acknowledges support of the Special Reserach Grant SFB 609 *Elektormagnetische Strömungsbeeinflussung in Metallurgie, Kristallzüchtung und Elektrochemie* sponsored by the Deutsche Forschungsgemeinschaft

References

[1] **Arada, N.; Casas, E. & Tröltzsch, F.** Error estimates for the numerical approximation of a semilinear elliptic control problem. *Computational Optimization and Applications*, 23:201–229, 2002.

[2] **Bergounioux, M.; Ito, K. & Kunisch, K.** Primal-dual strategy for constrained optimal control problems. *Siam J. Control Optim.*, 37:1176–1194, 1999.

[3] **Bonnans, F. & Casas, E.** An extension of Pontryagin's maximum principle for state constrained optimal control of semilinear elliptic equations and variational inequalities. *SIAM J. Control and Optim.*, 33:274–298, 1995.

[4] **Casas, E. & Tröltzsch, F.** Error estimates for linear-quadratic elliptic control problem. Preprint, 2002. Institut für Mathematik, Technische Universität Berlin, Germany.

[5] **Ciarlet, P.G.** *The Finite Element Method for Elliptic Problems.* North Holland, 1978.

[6] **Deckelnick, K. & Hinze, M.** Error estimates in space and time for tracking-type control of the instationary Stokes system. *ISNM*, 143:87–103, 2002.

[7] **Grisvard, P.** *Elliptic Operators in Nonsmooth Domains.* Pitman, London, 1985.

[8] **Hackbusch, W.** *Theory und Numerik elliptischer Differentialgleichungen.* Teubner, 1995.

[9] **Hintermüller, M. & Hinze, M.** SQP-methods in optimal control of the time-dependent Navier-Stokes system with control constraints. Preprint, to appear, 2003. Institut für Mathematik, Technische Universität Dresden, Germany.

[10] **Hintermüller, M.; Ito, K. & Kunisch, K.** The primal-dual active set method as a semi-smooth Newton method. *SIAM J. Control and Optim.*, 2003. to appear.

[11] **Kunisch, K. & Rösch, A.** Primal-dual active set strategy for a general class of constrained optimal control problems. *SIAM J. Optim.*, 13:321–334, 2002.

[12] **Rösch, A.** Error estimates for linear-quadratic control problems with control constraints. Preprint 0740-02, 2002. Institut für Mathematik, Technische Universität Berlin, Germany.

[13] **Thomée, V.** *Galerkin Finite Element Methods for Parabolic Problems.* Springer, 1997.

[14] **Xu, J.** & **Zou, J.** . Some nonoverlapping domain decomposition methods. *SIAM Rev.*, 40:857–914, 1998.