

PROPERTIES OF PROJECTION AND PENALTY METHODS FOR DISCRETIZED ELLIPTIC CONTROL PROBLEMS

ANDRZEJ CEGIELSKI

Faculty of Mathematics, Computer Science and Econometrics

University of Zielona Góra

Szafrana 4a, 65-516 Zielona Góra, Poland

AND

CHRISTIAN GROSSMANN

Institute of Numerical Mathematics

Dresden University of Technology

D-01062 Dresden, Germany

Abstract

In this paper, properties of projection and penalty methods are studied in connection with control problems and their discretizations. In particular, the convergence of an interior-exterior penalty method applied to simple state constraints as well as the contraction behavior of projection mappings are analyzed. In this study, the focus is on the application of these methods to discretized control problem.

Keywords: convex programming, control of PDE, projection methods, penalty methods.

2000 Mathematics Subject Classification: 65K10, 49M15, 90C25.

1. INTRODUCTION

The numerical treatment of optimization problems in function spaces, such as optimal control with partial differential equations, requires appropriate discretizations as well as adapted methods for solving the finite dimensional

optimization problems obtained by discretization. As a rule these finite dimensional problems are of huge dimension, but have a specific sparse structure. In the paper, for such optimization problems we study the behavior of two classes of optimization algorithms, namely projection type methods and a specific interior-exterior penalty technique, and combinations of projection and penalty methods where state constraints are handled by penalties and projections are applied to control constraints. In particular, Fejér monotonicity is proved for one projection method applied to augmented problems that occur in penalty methods.

Let $\Omega \subset \mathbb{R}^n$ be some bounded convex polyhedron with the boundary Γ . We denote by $V = H_0^1(\Omega)$ the Sobolev space of functions that have square integrable generalized derivatives and vanishing traces on the boundary. Let $H = L_2(\Omega)$. Both spaces are real Hilbert spaces that together with the dual $V^* = H^{-1}(\Omega)$ of $V = H_0^1(\Omega)$ form a Gelfand triple

$$V \hookrightarrow H \hookrightarrow V^*.$$

Here we identify $H^* = H$. Further, let \leq denote the natural almost everywhere pointwise semi-ordering in H and let $z \in H$ be given.

As an underlying model for our study we consider the weak formulation related to the following optimal control problem:

$$(1) \quad \begin{aligned} J(y, u) &= \frac{1}{2} \int_{\Omega} (y - z)^2 + \frac{\alpha}{2} \int_{\Omega} u^2 \rightarrow \min ! \\ \text{subject to} \quad & -\Delta y = u \text{ in } \Omega, \quad y = 0 \text{ on } \Gamma, \quad a \leq u \leq b, \quad y \leq f. \end{aligned}$$

Here $\alpha > 0$ denotes a given regularization parameter, $a, b \in H$, $f \in W^{2,\infty}(\Omega)$ and we suppose $a < b$ almost everywhere in Ω . Further, we assume that the feasible set of (1) is nonempty.

In relation to the Poisson equation in (1) we define a continuous linear mapping $S : H \rightarrow V$ by $Su = y$, where $y \in V$ is the unique solution of the variational equation

$$(2) \quad a(y, v) = (u, v) \quad \forall v \in V$$

with the bilinear form $a : V \times V \rightarrow \mathbb{R}$ given by

$$(3) \quad a(w, v) = \int_{\Omega} \nabla w \circ \nabla v \quad \forall w, v \in V.$$

The existence and uniqueness of the solution y of (2) for any $u \in H$ follows immediately from Lax-Milgram's lemma (cf. [6]). By using the mapping we can express the weak formulation of the problem (1) as an abstract optimization problem

$$(4) \quad J(u) = \frac{1}{2}(Su - z, Su - z) + \frac{\alpha}{2}(u, u) \rightarrow \min! \quad \text{s.t.} \quad u \in Q, \quad Su \leq f,$$

where

$$Q = \{u \in H : a \leq u \leq b\}.$$

Here (\cdot, \cdot) denotes the inner product in H . Analytical results, in particular optimality conditions, for optimal control problems of the form (4) are given in [12].

Now we apply a discretization technique for the states and controls. As a method of choice we consider piecewise linear C^0 finite elements on a regular triangulation of Ω . We denote the Lagrange basis functions related to all inner grid points of the triangulation by φ_j , $j = 1, \dots, N$. As discrete spaces $V_h \subset V$ and $H_h \subset H$ we defined

$$V_h = \text{span} \{\varphi_j\}_{j=1}^N \quad \text{and} \quad H_h = \text{span} \{\varphi_j\}_{j=1}^N,$$

respectively. Corresponding to this we denote

$$y_h(x) = \sum_{j=1}^N y_j \varphi_j(x) \quad \text{and} \quad u_h(x) = \sum_{j=1}^N u_j \varphi_j(x).$$

As the discrete problem we consider

$$(5) \quad \begin{aligned} J_h(u_h) &= \frac{1}{2}(S_h u_h - z_h, S_h u_h - z_h) + \frac{\alpha}{2}(u_h, u_h) \rightarrow \min! \\ \text{s.t.} \quad &u_h \in Q_h, \quad S_h u_h \leq f_h \end{aligned}$$

with

$$(6) \quad Q_h = \{u_h \in H_h : a_h \leq u_h \leq b_h\},$$

where $S_h : H_h \rightarrow V_h$ is defined by $S_h u_h = y_h$ with $y_h \in V_h$ that satisfies

$$(7) \quad a(y_h, v_h) = (u_h, v_h) \quad \forall v_h \in V_h.$$

Since $V_h \subset V$, variational equation (7) is a conforming finite element discretization of (2) and as in the continuous case Lax-Milgram's lemma guarantees the existence and uniqueness of its solution $y_h \in V_h$. Further,

$$a_h = \sum_{j=1}^N a_j \varphi_j, \quad b_h = \sum_{j=1}^N b_j \varphi_j, \quad f_h = \sum_{j=1}^N f_j \varphi_j, \quad z_h = \sum_{j=1}^N z_j \varphi_j$$

denote the L_2 -projection of the related functions a, b, f, z to the discrete space V_h , i.e.,

$$a_h \in V_h \quad \text{such that} \quad (a_h, v_h) = (a, v_h) \quad \forall v_h \in V_h$$

and similarly for the other L_2 -projections.

As a general property of the given data (as already assumed for the continuous problem (4)) we also suppose that the feasible set of the discrete problem (5) is nonempty. Under relatively weak conditions the feasibility in (5) can be derived asymptotically from the feasibility of (4) for sufficiently fine discretizations.

Theorem 1. *The continuous problem (4) as well as the discrete problem (5) possess unique solutions u and u_h , respectively, and we have*

$$\lim_{h \rightarrow 0+} \|u - u_h\| = 0.$$

For quantitative estimates and a detailed convergence analysis of control constrained problems we refer to e.g., [2, 11]. A more complicated case of state constraints is analyzed in e.g. [1, 4]. The discrete Dirac measures that form the proper discretization of the Lagrangian multiplier used in [4] have a related representation as Lagrangian multipliers for the finite dimensional pointwise inequalities.

Since piecewise linear finite elements are used the occurring inequality constraints $a_h \leq u_h \leq b_h$ reduce to bounds at the grid points of the triangles only, i.e.,

$$a_j \leq u_j \leq b_j \quad j = 1, \dots, N.$$

Similarly, the state constraints are in the discrete case equivalent to

$$y_j \leq f_j, \quad j = 1, \dots, N.$$

All the discrete functions under consideration can be expressed by their representations in finite dimensional coordinates, e.g., by $\mathbf{u} = (u_j)_{j=1}^N \in \mathbb{R}^N$ for the controls u_h and by $\mathbf{y} = (y_j)_{j=1}^N \in \mathbb{R}^N$ for the states y_h . Similarly, $\mathbf{a}, \mathbf{b}, \mathbf{f}$ denote the coordinate vectors related to a_h, b_h and f_h , respectively. In general, we use boldface symbols in the case of finite dimensional representations, e.g.,

$$u_h \in Q_h \iff \mathbf{u} \in \mathbf{Q} = \{\mathbf{u} \in \mathbb{R}^N : \mathbf{a} \leq \mathbf{u} \leq \mathbf{b}\}.$$

With the stiffness matrix $\mathbf{A} = (a_{ij})$ and the mass matrix $\mathbf{B} = (b_{ij})$ given by

$$a_{ij} = a(\varphi_j, \varphi_i) \quad \text{and} \quad b_{ij} = (\varphi_j, \varphi_i), \quad i, j = 1, \dots, N,$$

respectively, we can define the finite dimensional representation $\mathbf{J} : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ of J_h by

$$(8) \quad \mathbf{J}(\mathbf{u}, \mathbf{y}) = \frac{1}{2}(\mathbf{y} - \mathbf{z})^T \mathbf{B}(\mathbf{y} - \mathbf{z}) + \frac{\alpha}{2} \mathbf{u}^T \mathbf{B} \mathbf{u}$$

or in its reduced form $\mathbf{J} : \mathbb{R}^N \rightarrow \mathbb{R}$ by

$$(9) \quad \mathbf{J}(\mathbf{u}) = \frac{1}{2}(\mathbf{A}^{-1} \mathbf{B} \mathbf{u} - \mathbf{z})^T \mathbf{B}(\mathbf{A}^{-1} \mathbf{B} \mathbf{u} - \mathbf{z}) + \frac{\alpha}{2} \mathbf{u}^T \mathbf{B} \mathbf{u}.$$

In this reduced form \mathbf{y} has been eliminated via the Ritz-Galerkin equations $\mathbf{A} \mathbf{y} = \mathbf{B} \mathbf{u}$ which are equivalent to (7).

It should be mentioned that the discrete problems are essentially of the same type as the continuous ones if we replace the continuous spaces and operators by the proper discrete ones. Hence, most of the following properties of (4) hold in a similar way for its discrete version (5).

2. PENALTIES FOR DISCRETE STATE CONSTRAINTS

While the control constraints are easy to handle by projection methods this is not so for state constraints, even for the simple type as considered in (4). Here we apply a specific penalty technique to (5) to include discrete state constraints into the objective functional. We incorporate the discrete state constraints via the penalty like term

$$(10) \quad \Phi_h(y_h; s) = \sum_{j=1}^N \gamma_j \left(y_j - f_j + \sqrt{(y_j - f_j)^2 + s^2} \right),$$

which leads to the auxiliary problems

$$(11) \quad J_h(u_h; s) = J_h(u_h) + \Phi_h(S_h u_h) \rightarrow \min ! \quad \text{s.t.} \quad u_h \in Q_h.$$

Here $s > 0$ denotes a penalty parameter and $\gamma_j > 0$ denote parameters which have to be chosen such that $\gamma_j > \bar{\lambda}_j$, $j = 1, \dots, N$, where $\bar{\lambda}_h = (\bar{\lambda}_j) \in \mathbb{R}^N$ is the optimal Lagrangian multiplier vector related to the state constraints of the discrete problem (5). Then for $s \rightarrow 0+$ problem (11) approximates (5) as shown below. The interior-exterior penalty method used in (11) was introduced and analyzed by Kaplan [8]. Here we study the properties of this methods under the specific structure of the discrete control problem and provide a different convergence proof. In particular, unlike in [8], in this paper we include only some of the constraints, namely state constraints, via the penalty into the auxiliary objective while control constraints as well as the state equations are included unchanged into the penalty problem.

Before we evaluate the derivative of $\Phi_h(S_h u_h)$ let us express Φ_h via an integral, namely

$$(12) \quad \Phi_h(y_h; s) = \int_{\Omega} \sum_{j=1}^N \tilde{\gamma}_j \left(y_j - f_j + \sqrt{(y_j - f_j)^2 + s^2} \right) \varphi_j(x) dx$$

with $\tilde{\gamma}_j = \gamma_j / \int_{\Omega} \varphi_j$. This yields

$$(\Phi'_h(y_h; s), w_h) = \left(\sum_{j=1}^N \tilde{\gamma}_j \left(1 + \frac{y_j - f_j}{\sqrt{(y_j - f_j)^2 + s^2}} \right) \varphi_j, w_h \right),$$

where (\cdot, \cdot) denotes the scalar product in $L_2(\Omega)$. Taking into account $y_h = S_h u_h$, $w_h = S_h d_h$ we obtain

$$(\Phi'_h(S_h u_h; s), d_h) = \left(S_h^* \sum_{j=1}^N \tilde{\gamma}_j \left(1 + \frac{y_j - f_j}{\sqrt{(y_j - f_j)^2 + s^2}} \right) \varphi_j, d_h \right).$$

If we express the discrete problem (5) in finite dimensional coordinates we obtain for the non-reduced form the problem

$$(13) \quad \mathbf{J}(\mathbf{u}, \mathbf{y}) \rightarrow \min ! \quad \text{s.t.} \quad (\mathbf{u}, \mathbf{y}) \in \mathbf{W}, \quad \mathbf{y} \leq \mathbf{f}$$

with \mathbf{J} defined by (8) and

$$\mathbf{W} = \{(\mathbf{u}, \mathbf{y}) \in \mathbb{R}^N \times \mathbb{R}^N : \mathbf{a} \leq \mathbf{u} \leq \mathbf{b}, \mathbf{A}\mathbf{y} = \mathbf{B}\mathbf{u}\}.$$

Similarly, the discrete augmented problem (11) can be expressed in the form

$$(14) \quad \mathbf{J}(\mathbf{u}, \mathbf{y}; s) \rightarrow \min ! \quad \text{s.t.} \quad (\mathbf{u}, \mathbf{y}) \in \mathbf{W},$$

where, $\mathbf{J}(\cdot, \cdot; s) : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is defined by

$$(15) \quad \mathbf{J}(\mathbf{u}, \mathbf{y}; s) = \mathbf{J}(\mathbf{u}, \mathbf{y}) + \sum_{j=1}^N \gamma_j \Phi(y_j - f_j; s),$$

and Φ denotes the interior-exterior penalty function $\Phi(t; s) = t + \sqrt{t^2 + s^2}$. Notice that $\Phi(t; 0) = 2 \max\{0, t\}$, i.e., the function $\Phi(\cdot; s)$ for $s \rightarrow 0+$ approximates the well known exact penalty.

Reduced problems, i.e., formulations without an explicit use of the discrete state, are easily obtained if we substitute \mathbf{y} in (13) and (14), respectively, via discrete Poisson's equation by $\mathbf{y} = \mathbf{A}^{-1}\mathbf{B}\mathbf{u}$.

Theorem 2. *For any $s > 0$ the discrete auxiliary problem (14) has a unique solution $(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s))$ and for $s \rightarrow 0+$ we have $(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s)) \rightarrow (\bar{\mathbf{u}}, \bar{\mathbf{y}})$, where $(\bar{\mathbf{u}}, \bar{\mathbf{y}})$ denotes the optimal solution of (13).*

Proof. Problem (13) forms a linearly constrained optimization problem with a strongly convex objective function. Hence it has a unique optimal solution $(\bar{\mathbf{u}}, \bar{\mathbf{y}})$. Let L denote the partial Lagrangian for (13), where only the state constraints are included, i.e.,

$$L(\mathbf{u}, \mathbf{y}, \boldsymbol{\lambda}) = \mathbf{J}(\mathbf{u}, \mathbf{y}) + \boldsymbol{\lambda}^T(\mathbf{y} - \mathbf{f}).$$

There exists a unique $\bar{\boldsymbol{\lambda}} \in \mathbb{R}_+^N$ such that together with $(\bar{\mathbf{u}}, \bar{\mathbf{y}})$ it forms a saddle point of L , i.e.

$$(16) \quad L(\bar{\mathbf{u}}, \bar{\mathbf{y}}, \boldsymbol{\lambda}) \leq L(\bar{\mathbf{u}}, \bar{\mathbf{y}}, \bar{\boldsymbol{\lambda}}) \leq L(\mathbf{u}, \mathbf{y}, \bar{\boldsymbol{\lambda}}) \quad \forall (\mathbf{u}, \mathbf{y}) \in \mathbf{W}, \boldsymbol{\lambda} \in \mathbb{R}_+^N.$$

The objective function of the augmented problem (14) is strongly convex with a modulus that is independent of the penalty parameter. The set $\mathbf{W} \subset \mathbb{R}^N \times \mathbb{R}^N$ is by assumption non-empty. Since it is closed one can show

that for any $s > 0$ problem (14) possesses a unique solution $(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s)) \in W$ and that some constant $c > 0$ exists such that

$$(17) \quad \|(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s))\| \leq c \quad \forall s > 0.$$

The optimality of $(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s))$ for problem (14) together with the monotonicity of $\Phi(\cdot; s)$ and with $\bar{\mathbf{y}} \leq \mathbf{f}$ lead to

$$(18) \quad \begin{aligned} \mathbf{J}(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s); s) &\leq \mathbf{J}(\bar{\mathbf{u}}, \bar{\mathbf{y}}; s) = \mathbf{J}(\bar{\mathbf{u}}, \bar{\mathbf{y}}) + \sum_{j=1}^N \gamma_j \Phi(\bar{y}_j - f_j; s) \\ &\leq \mathbf{J}(\bar{\mathbf{u}}, \bar{\mathbf{y}}) + s \sum_{j=1}^N \gamma_j. \end{aligned}$$

On the other hand, the saddle point inequalities (16) imply

$$\mathbf{J}(\bar{\mathbf{u}}, \bar{\mathbf{y}}) = L(\bar{\mathbf{u}}, \bar{\mathbf{y}}, \bar{\boldsymbol{\lambda}}) \leq L(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s), \bar{\boldsymbol{\lambda}}) = \mathbf{J}(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s)) + \bar{\boldsymbol{\lambda}}^T (\tilde{\mathbf{y}}(s) - \mathbf{f})$$

and with (18) this yields

$$\sum_{j=1}^N \gamma_j \Phi(\tilde{y}_j(s) - f_j; s) \leq \bar{\boldsymbol{\lambda}}^T (\tilde{\mathbf{y}}(s) - \mathbf{f}) + s \sum_{j=1}^N \gamma_j.$$

Now, the properties

$$\Phi(t; s) \geq t, \quad \Phi(t; s) \geq 0 \quad \forall t \in \mathbb{R}$$

of the penalty function and $\gamma_j > \bar{\lambda}_j \geq 0$, $j = 1, \dots, N$, imply that

$$(\gamma_l - \bar{\lambda}_l) \Phi(\tilde{y}_l(s) - f_l; s) \leq \sum_{j=1}^N (\gamma_j - \bar{\lambda}_j) \Phi(\tilde{y}_j(s) - f_j; s) \leq s \sum_{j=1}^N \gamma_j$$

holds for any $l \in \{1, \dots, N\}$. This yields

$$\lim_{s \rightarrow 0+} \Phi(\tilde{y}_l(s) - f_l; s) = 0, \quad l \in \{1, \dots, N\}.$$

With the boundedness (17) and from the properties of Φ we obtain that the state $\tilde{\mathbf{y}}$ as a part of any accumulation point $(\tilde{\mathbf{u}}, \tilde{\mathbf{y}})$ for $s \rightarrow 0+$ satisfies

the constraints

$$\tilde{\mathbf{y}} \leq \mathbf{f}$$

which have been included via penalties. Hence, $(\tilde{\mathbf{u}}, \tilde{\mathbf{y}})$ is feasible for (13). With

$$\mathbf{J}(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s)) \leq \mathbf{J}(\tilde{\mathbf{u}}(s), \tilde{\mathbf{y}}(s); s) \leq \mathbf{J}(\bar{\mathbf{u}}, \bar{\mathbf{y}}; s) \leq \mathbf{J}(\bar{\mathbf{u}}, \bar{\mathbf{y}}) + s \sum_{j=1}^N \gamma_j$$

and with the continuity of \mathbf{J} we obtain that $(\tilde{\mathbf{u}}, \tilde{\mathbf{y}})$ is optimal. Because of (17) and of the uniqueness of the solution of (13) this proves the theorem. ■

3. PROPERTIES OF THE OBJECTIVE FUNCTIONAL

Here and in the following sections we analyze properties of the objective functional, possibly augmented by penalty terms, and study the contraction behavior of projection methods that rest upon these functionals. It turns out that some properties are restricted to the finite dimensional case while other properties do also hold in the infinite dimensional case of the underlying Hilbert space. Naturally, the properties shown for the more general infinite dimensional case are also similarly valid in the case of conforming discretization. To mark the finite dimensional case as before we write the discretization parameter h as a subscript at functionals, solutions etc.

Lemma 3. *Let $\Phi : H \rightarrow \mathbb{R}$ denote some differentiable convex functional. Then the augmented objective*

$$(19) \quad J_{\Phi}(u) = \frac{1}{2}(Su - z, Su - z) + \frac{\alpha}{2}(u, u) + \Phi(u)$$

is differentiable, convex and we have

$$(20) \quad J'_{\Phi}(u) = S^*(Su - z) + \alpha u + \Phi'(u).$$

Furthermore, J_{Φ} is strongly convex with the modulus

$$(21) \quad c = \lambda_{\inf}^2(S) + \alpha,$$

where $\lambda_{\inf}(S)$ denotes the lower limit of the spectrum of the operator S .

Proof. The first part follows immediately from the definition of J_Φ . Furthermore, by the convexity and differentiability of Φ , and by (20), we have

$$\begin{aligned}
J_\Phi(u+v) &= \frac{1}{2}(Su-z+ Sv, Su-z+ Sv) + \frac{\alpha}{2}(u+v, u+v) + \Phi(u+v) \\
&\geq \frac{1}{2}(Su-z, Su-z) + \frac{1}{2}(Sv, Sv) + (Su-z, Sv) \\
&\quad + \frac{\alpha}{2}(u, u) + \frac{\alpha}{2}(v, v) + \alpha(u, v) + \Phi(u) + (\Phi'(u), v) \\
&= J_\Phi(u) + (S^*(Su-z), v) + \alpha(u, v) + (\Phi'(u), v) \\
&\quad + \frac{1}{2}(Sv, Sv) + \frac{\alpha}{2}(v, v) \\
&= J_\Phi(u) + (J'_\Phi(u), v) + \frac{1}{2}(Sv, Sv) + \frac{\alpha}{2}(v, v) \\
&\geq J_\Phi(u) + (J'_\Phi(u), v) + \frac{1}{2}(\lambda_{\inf}^2(S) + \alpha)\|v\|^2,
\end{aligned}$$

i.e., J_Φ is strongly convex with the modulus $c = \lambda_{\inf}^2(S) + \alpha$. ■

Remark 4. In the infinite dimensional case for the operator S defined by (2) we have $\lambda_{\inf}(S) = 0$ (see e.g., [14]).

When the discrete state equations (7) define the finite dimensional operator S_h we obtain the related functional

$$(22) \quad J_{\Phi,h} = \frac{1}{2}(S_h u - z_h, S_h u - z_h) + \frac{\alpha}{2}(u, u) + \Phi(u)$$

and we have $\lambda_{\inf}(S_h) > 0$, i.e., the minimal eigenvalue is positive. The derivative of $J_{\Phi,h}$ has the form (20) with the operator S_h instead of S .

4. PROBLEM WITH BOUNDS ON CONTROLS ONLY

In this section, we consider the continuous problem (4) without state constraints $Su \leq f$, i.e., we deal with the problem

$$(23) \quad J(u) \rightarrow \min ! \quad u \in Q.$$

Let us define the operator $R : H \rightarrow H$ by

$$(24) \quad Ru = u - \rho J'(u),$$

where $\rho > 0$ denotes some fixed parameter.

From the optimality conditions we obtain that $\tilde{u} \in Q$ is the optimal solution of (4) if and only if

$$(25) \quad (J'(\tilde{u}), u - \tilde{u}) \geq 0 \text{ for all } u \in Q$$

holds. This condition is equivalent to

$$(26) \quad P_Q(\tilde{u} - \rho J'(\tilde{u})) = \tilde{u} \text{ for any fixed } \rho > 0,$$

i.e.,

$$(27) \quad \tilde{u} \in \text{Fix}(P_Q \circ R) \text{ for any fixed } \rho > 0,$$

Next, we study the contraction behavior of the operator R and its consequences.

Lemma 5. *The operator R is contractive for any fixed $\rho \in (0, 2\alpha[\lambda_{\max}^2(S) + \alpha]^2)$.*

Proof. By the structure (24) and by the strong convexity of J , we have

$$\begin{aligned} \|Ru - Rv\|^2 &= \|u - v - \rho(J'(u) - J'(v))\|^2 \\ &= \|u - v\|^2 + \rho^2\|J'(u) - J'(v)\|^2 \\ &\quad - 2\rho\langle u - v, J'(u) - J'(v) \rangle \\ (28) \quad &\leq \|u - v\|^2 + \rho^2\|J'(u) - J'(v)\|^2 - 2\rho c\|u - v\|^2 \\ &\quad \forall u, v \in H, \end{aligned}$$

where $c = \alpha > 0$ (due to $\lambda_{\inf}(S) = 0$ in the infinite dimensional case) is a modulus of the strong convexity of J . Since

$$\begin{aligned} \|J'(u) - J'(v)\|^2 &= \|S^*(Su - z) + \alpha u - S^*(Sv - z) - \alpha v\|^2 \\ &= \|(S^*S + \alpha I)(u - v)\|^2 \\ &\leq [\lambda_{\max}^2(S) + \alpha]^2\|u - v\|^2 \end{aligned}$$

from (28) we obtain

$$\|Ru - Rv\|^2 \leq (1 - 2\rho c + \rho^2[\lambda_{\max}^2(S) + \alpha]^2)\|u - v\|^2 \quad \forall u, v \in H.$$

Hence, R is contractive if

$$1 - 2\rho c + \rho^2[\lambda_{\max}^2(S) + \alpha]^2 < 1$$

holds. With $\lambda_{\inf}(S) = 0$ this is equivalent to $\rho \in (0, 2\alpha[\lambda_{\max}^2(S) + \alpha]^2)$. ■

Using the contraction properties of R the following iterative treatment of the control problem under consideration can be applied.

With an arbitrary $u^1 \in H$ we generate $\{u^k\} \subset V$ recursively by

$$(29) \quad u^{k+1} = Pu^k, \quad k = 1, 2, \dots,$$

where $P : H \rightarrow H$ is given by

$$(30) \quad P = P_Q \circ R.$$

Theorem 6. *If $\rho \in (0, 2\alpha[\lambda_{\max}^2(S) + \alpha]^2)$ then for any $u^1 \in H$ the iterative procedure (29) generates a sequence (u^k) which converges to the solution \tilde{u} of problem (23).*

Proof. Since the projection operator P_Q is nonexpansive, the theorem follows immediately from Lemma 5. ■

Remark 7. Theorem 6 holds in a similar way for the semi-discrete

$$(31) \quad u^1 \in H, \quad u^{k+1} = P_Q(u^k - \rho J'_h(u^k)), \quad k = 1, 2, \dots,$$

and the fully-discrete version

$$u_h^1 \in H_h, \quad u_h^{k+1} = P_{Q_h}(u_h^k - \rho J'_h(u_h^k)), \quad k = 1, 2, \dots,$$

of the iteration process (29).

As proposed by Hinze [7] for problems of type (23) the structure of the derivative of J_h can be exploited efficiently. For the choice $\rho = \frac{1}{\alpha}$ the

optimality condition

$$u = P_Q(u - \rho J'_h(u)),$$

which is the semidiscrete version of (26) used in (31), has just the form

$$(32) \quad u = P_Q\left(-\frac{1}{\alpha} S_h^*(y_h - z)\right).$$

Here y_h denotes the solution of the discrete state equation for a given $u \in H$. This enables us to eliminate the control u via the optimality condition in the semidiscrete case. This means that only states and adjoint states have to be discretized explicitly. Since the operators S_h and S_h^* possess good smoothing properties, optimal convergence rates can be achieved by this control reduced discretization. An approximate version of control reduction in the continuous case by means of logarithmic barriers has been investigated in [13].

5. PROBLEM WITH BOUNDS ON CONTROL AND STATE

In this section, we focus on the fully discrete control problem (5) and we apply the penalty term Φ_h given by (10) to treat the discrete state constraints.

Let the operator $T_h : H_h \rightarrow H_h$ be defined by

$$(33) \quad T_h u_h = u_h - \frac{J_{\Phi,h}(u_h) - \gamma_h}{\|J'_{\Phi,h}(u_h)\|^2} J'_{\Phi,h}(u_h),$$

where $\gamma_h < J_{\Phi,h}(u_h)$, and let $T_{h,\lambda} = (1 - \lambda)I + \lambda T_h$ denote a relaxation of T_h for the relaxation parameter $\lambda \in [0, 2]$.

5.1. The Fejér monotonicity of the operator $P_{Q_h} \circ T_h$

Theorem 8. *Let $\lambda \in [0, 2]$. Then for any $u_h \in Q_h$ and $\gamma_h \in [\tilde{J}_{\Phi,h}, J_{\Phi,h}(u_h)]$ there holds the inequality*

$$\|P_{Q_h} T_{h,\lambda} u_h - \tilde{u}_h\|^2 \leq \|u_h - \tilde{u}_h\|^2 - \lambda(2 - \lambda) \|T_h u_h - u_h\|^2,$$

where \tilde{u}_h and $\tilde{J}_{\Phi,h} = J_{\Phi,h}(\tilde{u}_h)$ denote the solution of (11) and the related optimal value, respectively. Consequently, $P_{Q_h} \circ T_{h,\lambda}$ is Fejér monotone with respect to the solution \tilde{u}_h of problem (11) for $\gamma_h \in [\tilde{J}_{\Phi,h}, J_{\Phi,h}(u_h)]$.

Proof. (See also [9, 10] or [3] for similar results). Let $u_h \in H_h$. Denote by J_{lin} the linearization of $J_{\Phi,h}$ in the point u_h , i.e.,

$$J_{lin}(v_h) = (J'_{\Phi,h}(u_h), v_h - u_h) + J_{\Phi,h}(u_h), \quad \text{for all } v_h \in H_h.$$

Let

$$W(G, \kappa) = \{v_h \in H_h : G(v) \leq \kappa\}$$

denote the sublevel set of a functional $G : H_h \rightarrow \mathbb{R}$ at a given level $\kappa \in \mathbb{R}$. Observe that

$$W(J_{lin}, \gamma_h) = \{v_h \in H_h : (J'_{\Phi,h}(u_h), v_h) \leq \beta\},$$

where $\beta = (J'_{\Phi,h}(u_h), u_h) - J_{\Phi,h}(u_h) + \gamma_h$. Consequently,

$$T_h u_h = P_{W(J_{lin}, \gamma_h)}(u_h) \quad \text{and} \quad T_{h,\lambda} u_h = (1 - \lambda)u_h + \lambda P_{W(J_{lin}, \gamma_h)}(u_h)$$

for $\gamma_h \leq J_{\Phi,h}(u_h)$. Furthermore, we have $\tilde{u}_h \in W(J_{lin}, \gamma_h)$ for $\gamma_h \geq \tilde{J}_{\Phi,h}$.

With the nonexpansivity of P_{Q_h} this yields

$$\begin{aligned} \|P_{Q_h} T_{h,\lambda} u_h - \tilde{u}_h\|^2 &= \|P_{Q_h}((1 - \lambda)u_h + \lambda P_{W(J_{lin}, \gamma_h)}(u_h)) - P_{Q_h} \tilde{u}_h\|^2 \\ &\leq \|(1 - \lambda)u_h + \lambda P_{W(J_{lin}, \gamma_h)}(u_h) - \tilde{u}_h\|^2 \\ &= \|(u_h - \tilde{u}_h) + \lambda(P_{W(J_{lin}, \gamma_h)}(u_h) - u_h)\|^2 \\ &= \|u_h - \tilde{u}_h\|^2 + \lambda^2 \|P_{W(J_{lin}, \gamma_h)}(u_h) - u_h\|^2 \\ &\quad - 2\lambda (\tilde{u}_h - u_h, P_{W(J_{lin}, \gamma_h)}(u_h) - u_h). \end{aligned}$$

By the Kolmogorov characterization of the metric projection we have

$$(\tilde{u}_h - u_h, P_{W(J_{lin}, \gamma_h)}(u_h) - u_h) \geq \|P_{W(J_{lin}, \gamma_h)}(u_h) - u_h\|^2.$$

Hence, we obtain

$$\begin{aligned} \|P_{Q_h} T_{h,\lambda} u_h - \tilde{u}_h\|^2 &\leq \|u_h - \tilde{u}_h\|^2 - \lambda(2 - \lambda) \|P_{W(J_{lin}, \gamma_h)}(u_h) - u_h\|^2 \\ &= \|u_h - \tilde{u}_h\|^2 - \lambda(2 - \lambda) \|T_h u_h - u_h\|^2, \end{aligned}$$

where $\gamma_h \in [\tilde{J}_{\Phi,h}, J_{\Phi,h}(u_h)]$. ■

5.2. An estimation of the optimal value of the control problem

In the projection method presented in the next section we need a lower bound for the optimal objective value as well as an upper bound for the distance of the starting point to the solution set. For the first bound we have the following result.

Lemma 9. *The following inequality holds*

$$(34) \quad \min_{u_h \in Q_h} J_{\Phi,h}(u_h) \geq \frac{\alpha}{2(\lambda_{\max}^2(S_h) + \alpha)} \|z_h\|^2,$$

where z_h is the projection of the given target function.

Proof. Let $\{v_j\}_{j=1}^N$ denote a complete orthonormal system of eigen vectors of S_h , i.e.,

$$S_h v_j = \lambda_j v_j, \text{ for } j = 1, \dots, N$$

and $(v_i, v_j) = \delta_{ij}$, $i, j = 1, \dots, N$. Let us represent u_h, z_h by $\{v_j\}$, i.e.,

$$u_h = \sum_j \xi_j v_j \quad \text{and} \quad z_h = \sum_j \zeta_j v_j.$$

Due to the orthonormality the following holds

$$\xi_j = (u_h, v_j), \quad \zeta_j = (z_h, v_j) \quad j = 1, \dots, N.$$

Using the property $\Phi_h \geq 0$ of the penalty function, we have

$$\begin{aligned} J_{\Phi,h}(u_h) &\geq J_h(u_h) = \frac{1}{2} \sum_j [(\lambda_j \xi_j - \zeta_j)^2 + \alpha \xi_j^2] \\ &= \frac{1}{2} \sum_j [(\lambda_j^2 + \alpha) \xi_j^2 - 2\lambda_j \xi_j \zeta_j + \zeta_j^2] \\ (35) \quad &\geq \frac{1}{2} \sum_j \left(1 - \frac{\lambda_j^2}{\lambda_j^2 + \alpha}\right) \zeta_j^2 = \frac{1}{2} \sum_j \left(\frac{\alpha}{\lambda_j^2 + \alpha}\right) \zeta_j^2 \\ &\geq \frac{1}{2} \frac{\alpha}{\lambda_{\max}^2(S_h) + \alpha} \sum_j \zeta_j^2. \end{aligned}$$

Pythagoras' theorem $\|z_h\|^2 = \sum_j \zeta_j^2$ completes the proof. ■

Remark 10. The lower bound provided in Lemma 9 for the discrete objective J_h can also be extended by the same arguments to the continuous objective J because the operator S has a complete orthonormal eigensystem $\{v_j\}_{j=1}^\infty$ (see, e.g., Zeidler [14]).

The distance of a point $u_h \in Q_h$ to the solution \tilde{u}_h of the auxiliary problem (11) can be obviously estimated in the following ways

$$(36) \quad \begin{aligned} \|u_h - \tilde{u}_h\|_{L_2(\Omega)} &\leq \sup_{v_h, w_h \in Q} \|v_h - w_h\|_{L_2(\Omega)} \\ &\leq \left(\int_{\Omega} (b - a)^2 d\mu \right)^{1/2} \end{aligned}$$

or

$$(37) \quad \|u_h - \tilde{u}_h\|_{L_1(\Omega)} \leq \sup_{v_h \in Q_h} \|v_h - u_h\|_{L_1(\Omega)} = \int_{\Omega} \max\{u_h - a_h, b_h - u_h\} d\mu.$$

Notice that $\Omega_h = \Omega$ due to the assumption that $\Omega \subset \mathbb{R}^2$ is a convex polyhedron for an appropriate triangularization. In particular, for $u_h = \frac{1}{2}(a_h + b_h)$ this yields $\|u_h - \tilde{u}_h\| \leq \frac{1}{2}\|b_h - a_h\|$ for both norms. Since $J_{\Phi,h}$ is strongly convex we can also obtain some other upper approximation of $\|u_h - \tilde{u}_h\|_{L_2(\Omega)}$.

Lemma 11. *The following inequalities hold*

$$(38) \quad \|u_h - \tilde{u}_h\|_{L_2(\Omega)} \leq \sqrt{2 \frac{J_{\Phi,h}(u_h) - \underline{\alpha}}{c_h}},$$

$$(39) \quad \|u_h - \tilde{u}_h\|_{L_2(\Omega)} \leq \frac{\|J'_{\Phi,h}(u_h)\|_{L_2(\Omega)}}{c_h},$$

where $\underline{\alpha} \leq \tilde{J}_{\Phi,h}$ and $c_h > 0$ is a modulus of the strong convexity of $J_{\Phi,h}$.

For the proof we refer to [9].

5.3. A projection method with level control

Now consider the iterative procedure which generates $\{u_h^k\} \subset H_h$ recursively by

$$(40) \quad u_h^{k+1} = P_{Q_h} \left((1 - \lambda)u_h^k - \lambda \frac{J_{\Phi,h}(u_h^k) - \gamma_k}{\|J'_{\Phi,h}(u_h^k)\|^2} J'_{\Phi,h}(u_h^k) \right),$$

where $u_h^1 \in H_h$ is an arbitrary starting point and $\gamma_k = (1 - \nu)\bar{J}^k + \nu\underline{J}^k$ is a convex combination of upper and lower bounds of the minimal objective value $\tilde{J}_{\Phi,h}$ for $\nu \in (0, 1)$. We can take, e.g.,

$$\bar{J}^k = \min\{J_{\Phi_h}(u_h^i) : i = 1, 2, \dots, k\}.$$

As \underline{J}^1 we can take the estimation given in Lemma 9. Furthermore, the lower bound \underline{J}^k of the minimal value $\tilde{J}_{\Phi,h}$ can be updated in an iteration if a lower bound of the optimal value $\tilde{J}_{\Phi,h}$ is available and one detects that $\gamma_k < \tilde{J}_{\Phi,h}$. To do it one needs an upper bound of the distance $\|u_h^k - \tilde{u}_h\|$ which can be obtained by Lemma 11. In this case, one can set $\underline{J}^{k+1} = \gamma_k$. The details as well as the proof of the convergence to a solution (\tilde{u}_h in our case) of the algorithm can be found in [9, 10] or [3].

Of course, the projection method with level control (40) can also be applied to problems with bounds on controls only. In this case, J_{Φ_h} and $J'_{\Phi,h}$ in (40) should be replaced by J_h and J'_h , respectively.

Now we compare the maximal step-length for which we can secure contractivity and the Fejér monotonicity of operators R and $P_{Q_h} \circ T_{\lambda,h}$, respectively. This enables us to compare the convergence of these methods. In the following comparison of the two methods we restrict ourselves to discrete problems with bounds on controls only. Similar to Lemma 5 the maximal limit step-length for operator R to be contractive is equal to

$$\rho_1 = \frac{2\alpha}{[\lambda_{\max}^2(S_h) + \alpha]^2}.$$

By Theorem 8 the maximal step length for operator $P_{Q_h} \circ T_{\lambda,h}$ possessing the Fejér monotonicity is equal to

$$\rho_2 = 2 \frac{J_h(u_h) - \gamma_h}{\|J'_h(u_h)\|^2}.$$

Since the step-length in the second case depends on γ_h , for the sake of comparison we consider the case $\gamma_h = (1 - \nu)\bar{J}_h + \nu\underline{J}_h$, where

$$\nu \in (0, 1), \quad \underline{J}_h \leq \tilde{J}_h \leq \bar{J}_h \leq J_h(u_h),$$

as was made in the variable target value method [9]. We have, by the strong

convexity of J_h , by the optimality condition (25) and by (21)

$$\begin{aligned} \rho_2 &= 2 \frac{J_h(u_h) - \gamma_h}{\|J'_h(u_h)\|^2} \geq 2\nu \frac{J_h(u_h) - \underline{J}_h}{\|J'_h(u_h)\|^2} \geq 2\nu \frac{J_h(u_h) - \tilde{J}_h}{\|J'_h(u_h)\|^2} \\ &\geq 2\nu \frac{\langle J_h(\tilde{u}_h), u_h - \tilde{u}_h \rangle + \frac{1}{2}c\|u_h - \tilde{u}_h\|^2}{\|J'_h(u_h)\|^2} \geq \nu c \frac{\|u_h - \tilde{u}_h\|^2}{\|J'_h(u_h)\|^2} \\ &\geq \nu \frac{\alpha}{[\lambda_{\max}^2(S_h) + \alpha]^2} = \frac{\nu}{2} \rho_1. \end{aligned}$$

Furthermore, one can easily check that

$$\alpha \in [(1 + \sqrt{2})\lambda_{\max}^2(S_h), +\infty) \implies \rho_1 = \frac{2\alpha}{[\lambda_{\max}^2(S_h) + \alpha]^2} \geq \frac{1}{\alpha}.$$

This implies, that the algorithmic operator given by (33) generates the step-length at least as long as the one proposed in [7].

REFERENCES

- [1] N. Arada, E. Casas and F. Tröltzsch, *Error estimates for the numerical approximation of a semilinear elliptic control problem*, Comput. Optim. Appl. **23** (2002), 201–229.
- [2] E. Casas and F. Tröltzsch, *Error estimates for linear-quadratic elliptic control problems*, in: Barbu, V. (ed.) *et al.*, Analysis and optimization of differential systems. IFIP TC7/WG 7.2 International Working Conference. Kluwer, Boston (2003), 89–100.
- [3] A. Cegielski, *A method of projection onto an acute cone with level control in convex minimization*, Math. Programming **85** (1999), 469–490.
- [4] K. Deckelnick and M. Hinze, *Convergence of a finite element approximation to a state constrained elliptic control problem*, Preprint MATH-NM-01-2006, TU Dresden 2006.
- [5] K. Goebel and W.A. Kirk, *Topics in Metric Fixed Point Theory*, Cambridge Univ. Press, Cambridge 1990.
- [6] Ch. Grossmann and H.-G. Roos, *Numerische Behandlung partieller Differentialgleichungen* (3-rd edition), B.G. Teubner, Stuttgart 2005.
- [7] M. Hinze, *A variational discretization concept in control constrained optimization: The linear-quadratic case*, Comput. Optim. Appl. **30** (2005), 45–61.

- [8] A.A. Kaplan, *Convex programming algorithms using the smoothing of exact penalty functions*, (Russian), Sib. Mat. Zh. **23** (1982), 53–64.
- [9] S. Kim, H. Ahn and S.-C. Cho, *Variable target value subgradient method*, Math. Programming **49** (1991), 359–369.
- [10] K.C. Kiwiel, *The efficiency of subgradient projection methods for convex optimization, part I: General level methods*, SIAM J. Control Optim. **34** (1996), 660–676.
- [11] A. Rösch, *Error estimates for linear-quadratic control problems with control constraints*, Optim. Methods Softw. **21** (2006), 121–134.
- [12] F. Tröltzsch, *Optimale Steuerung partieller Differentialgleichungen. Theorie, Verfahren und Anwendungen*, Vieweg, Wiesbaden 2005.
- [13] M. Weiser, T. Gänzler and A. Schiela, *A control reduced primal interior point method for PDE constrained optimization*, ZIB Report 04-38, Zuse-Zentrum Berlin 2004.
- [14] E. Zeidler, *Nonlinear Functional Analysis and Its Applications. II/A: Linear Monotone Operators*, Springer, New York 1990.

Received 24 July 2006