

Primal-dual Strategy for Constrained Optimal Control Problems

Maitine Bergounioux, Kazufumi Ito, Karl Kunisch

► **To cite this version:**

Maitine Bergounioux, Kazufumi Ito, Karl Kunisch. Primal-dual Strategy for Constrained Optimal Control Problems. *SIAM Journal on Control and Optimization*, Society for Industrial and Applied Mathematics, 1999, 37, no 4, pp. 1176-1194. hal-00023011

HAL Id: hal-00023011

<https://hal.archives-ouvertes.fr/hal-00023011>

Submitted on 18 Apr 2006

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Primal-dual Strategy for Constrained Optimal Control Problems

MAÏTINE BERGOUNIOUX¹ KAZUFUMI ITO² KARL KUNISCH³

Abstract. An algorithm for efficient solution of control constrained optimal control problems is proposed and analyzed. It is based on an active set strategy involving primal as well as dual variables. For discretized problems sufficient conditions for convergence in finitely many iterations are given. Numerical examples are given and the role of strict complementarity condition is discussed.

Keywords: Active Set, Augmented Lagrangian, Primal-dual method, Optimal Control.

AMS subject classification. 49J20, 49M29

1. Introduction and formulation of the problem. In the recent past significant advances have been made in solving efficiently nonlinear optimal control problems. Most of the proposed methods are based on variations of the sequential quadratic programming (SQP) technique, see for instance [HT, KeS, KuS, K, T] and the references given there. The SQP-algorithm is sequential and each of its iterations requires the solution of a quadratic minimization problem subject to linearized constraints. If these auxiliary problems contain inequality constraints with infinite dimensional image space then their solution is still a significant challenge.

In this paper we propose an algorithm for the solution of infinite dimensional quadratic problems with linear equality constraints and pointwise affine inequality constraints. It is based on an active set strategy involving primal and dual variables. It thus differs significantly from conventional active set strategies that involve primal variables only, see [Sch] for example. In practice the proposed algorithm behaves like an infeasible one. The iterates of the algorithm violate the constraints up to the next-to-the-last iterate. The algorithm stops at a feasible and optimal solution.

Within this paper we do not aim for generality but rather we treat as a model problem an unilateral control constraint optimal control problem related to elliptic partial differential equations. The distributed nature of this problem, which is reflected in the fact that it behaves like an obstacle problem for the biharmonic equation, makes it difficult to analyze.

Let us briefly outline the contents of the paper. The algorithm will be presented in Section 2. We prove that if the algorithm produces the same active set in two consecutive iterates then the optimal solution has been obtained. In Section 3 we shall give sufficient conditions which guarantee that an augmented Lagrangian functional behaves as a decreasing merit function for the algorithm. In practice this implies finite step convergence of the discretized problem. Section 4 is devoted to showing that for a minor modification of the algorithm the cost functional is increasing until the

¹UMR-CNRS 6628, Université d'Orléans, U.F.R. Sciences, B.P. 6759, F-45067 Orléans Cedex 2, France. E-mail: Maitine.Bergounioux@labomath.univ-orleans.fr. This work was supported in part by EEC, HCM Contract CHRX-CT94-0471

²Department of Mathematics, North Carolina State University, Raleigh, NC27695, USA.

³Institut für Mathematik, Universität Graz, A-8010 Graz, Austria, E-mail: Kunisch@kfunigraz.ac.at. Work supported in part by EEC, HCM Contract CHRX-CT94-0471 and Fonds zur Förderung der wissenschaftlichen Forschung, UF8, "Optimization and Control".

feasible optimal solution is reached. In Section 5 several numerical examples are given. For most examples the algorithm behaves extremely efficient and typically converges in less than five iterations. Thus, to present interesting cases the majority of the test examples is in some sense extreme: Either the strict complementarity condition is violated or the cost of the control is nearly zero.

To describe the problem, let Ω be an open, bounded subset of \mathbb{R}^N , $N \leq 3$, with smooth boundary Γ and consider the following distributed optimal control problem :

$$\min J(y, u) = \frac{1}{2} \int_{\Omega} (y - z_d)^2 dx + \frac{\alpha}{2} \int_{\Omega} (u - u_d)^2 dx , \quad (\mathcal{P})$$

$$(1.1) \quad -\Delta y = u \text{ in } \Omega , \quad y = 0 \text{ on } \Gamma ,$$

$$(1.2) \quad u \in U_{ad} \subset L^2(\Omega) ,$$

where $z_d, u_d \in L^2(\Omega)$, $\alpha > 0$ and $U_{ad} = \{ u \in L^2(\Omega) \mid u(x) \leq b(x) \text{ a.e. in } \Omega \}$, $b \in L^\infty(\Omega)$.

It is well known that, for every $u \in L^2(\Omega)$ system (1.1) has a unique solution $y = \mathcal{T}(u)$ in $H^2(\Omega) \cap H_o^1(\Omega)$.

REMARK 1.1. *To emphasis the basic ideas of the proposed approach we treated the rather simple problem (\mathcal{P}) . Many generalizations are possible. In particular, the analysis of this paper can be extended to the case where $-\Delta$ in (1.1) is replaced by any strictly elliptic second order differential operator. The algorithm itself can be easily adapted to other optimal control problems involving, for example ordinary differential equations. Its numerical efficiency as well as the convergence analysis require some additional research.*

It is standard that problem (\mathcal{P}) has a unique solution (y^*, u^*) characterized by the following optimality system :

$$\begin{cases} -\Delta y^* = u^* \text{ in } \Omega, & y^* \in H_o^1(\Omega) , \\ -\Delta p^* = z_d - y^* \text{ in } \Omega, & p^* \in H_o^1(\Omega) , \\ (\alpha(u^* - u_d) - p^*, u - u^*) \geq 0 & \text{for all } u \in U_{ad}, \end{cases}$$

where (\cdot, \cdot) denotes the $L^2(\Omega)$ -inner product.

Let us give an equivalent formulation for this optimality system which is essential to motivate the forthcoming algorithm:

THEOREM 1.1. *The unique solution (y^*, u^*) to problem (\mathcal{P}) is characterized by*

$$(S) \quad \begin{cases} -\Delta y^* = u^* \text{ in } \Omega, & y^* \in H_o^1(\Omega) , \\ -\Delta p^* = z_d - y^* \text{ in } \Omega, & p^* \in H_o^1(\Omega) , \\ u^* = u_d + \frac{p^* - \lambda^*}{\alpha} , \\ \lambda^* = c[u^* + \frac{\lambda^*}{c} - \Pi(u^* + \frac{\lambda^*}{c})] = c \max(0, u^* + \frac{\lambda^*}{c} - b) , \end{cases}$$

for every $c > 0$. Here Π denotes the projection of $L^2(\Omega)$ onto U_{ad} .

Proof - We refer to [IK]. ■

We point out that the last equation in (\mathcal{S})

$$(1.3) \quad \lambda^* = c[u^* + \frac{\lambda^*}{c} - \Pi(u^* + \frac{\lambda^*}{c})]$$

is equivalent to

$$(1.4) \quad \lambda^* \in \partial I_{U_{ad}}(u^*) ,$$

where ∂I_C denotes the subdifferential of the indicator function I_C of a convex set C . This follows from general properties of convex functions (see [IK] for example) and can also easily be verified directly for the convex function $I_{U_{ad}}$. The replacement of the well known differential inclusion (1.4) [B] in the optimality system for (\mathcal{P}) by (1.3) is an essential ingredient of the algorithm that we shall propose.

Here and below, order relations like “max” and “ \leq ” between elements of $L^2(\Omega)$ are understood in the pointwise almost everywhere sense.

Let us interpret the optimality system (\mathcal{S}) . From $-\Delta y^* = u_d + \frac{p^* - \lambda^*}{\alpha}$ it follows that $p^* = \alpha[-\Delta y^* - u_d] + \lambda^*$ and hence

$$-\alpha \Delta y^* - \Delta^{-1} y^* + \lambda^* = \alpha u_d - \Delta^{-1} z_d .$$

It follows that

$$\begin{aligned} \alpha u^* + \Delta^{-2} u^* + \lambda^* &= \alpha u_d - \Delta^{-1} z_d , \\ \lambda^* &= c \max(0, u^* + \frac{\lambda^*}{c} - b) \quad \text{for all } c > 0 \end{aligned}$$

which implies the highly distributed nature of the optimal control. Setting $\mathcal{H} = \alpha I + \Delta^{-2}$ and $f = \alpha u_d - \Delta^{-1} z_d$, system (\mathcal{S}) can be expressed as

$$(\mathcal{S})_1 \quad \begin{cases} \mathcal{H}u^* + \lambda^* &= f , \\ \lambda^* &= c \max(0, u^* + \frac{\lambda^*}{c} - b) \quad \text{for all } c > 0 \end{cases}$$

We observe that by setting $u = -\Delta y$, system (\mathcal{S}) constitutes an optimality system for the variational inequality

$$\begin{cases} \min & \frac{\alpha}{2} \int_{\Omega} |\Delta y|^2 dx + \frac{1}{2} \int_{\Omega} |y - (z_d - \alpha \Delta u_d)|^2 dx \\ y \in H_o^1(\Omega) \cap H^2(\Omega) \\ -\Delta y \leq b \end{cases}$$

the regularity of which was studied in [BS].

2. Presentation of the Algorithm. In this section we present the primal-dual active set algorithm and discuss some of its basic properties. Let us introduce the active and inactive sets for the solution to (\mathcal{P}) and define

$$\mathcal{A}^* = \{ x \mid u^*(x) = b \text{ a.e. } \} \text{ and } \mathcal{I}^* = \{ x \mid u^*(x) < b \text{ a.e. } \} .$$

The proposed strategy is based on (1.3). Given (u_{n-1}, λ_{n-1}) the active set for the current iterate is chosen as

$$\mathcal{A}_n = \left\{ x \mid u_{n-1}(x) + \frac{\lambda_{n-1}(x)}{c} > b \text{ a.e.} \right\}.$$

We recall that $\lambda^* \geq 0$ and in the case of strict complementarity $\lambda^* > 0$ on \mathcal{A}^* . The complete algorithm is specified next

Algorithm

1. Initialization : choose y_o , u_o and λ_o and set $n = 1$.
2. Determine the following subsets of Ω :

$$\mathcal{A}_n = \left\{ x \mid u_{n-1}(x) + \frac{\lambda_{n-1}(x)}{c} > b \right\}, \quad \mathcal{I}_n = \left\{ x \mid u_{n-1}(x) + \frac{\lambda_{n-1}(x)}{c} \leq b \right\}.$$

3. If $n \geq 2$ and $\mathcal{A}_n = \mathcal{A}_{n-1}$ then STOP.
4. Else, find $(y_n, p_n) \in H_o^1(\Omega) \times H_o^1(\Omega)$ such that

$$\begin{aligned} -\Delta y_n &= \begin{cases} b & \text{in } \mathcal{A}_n \\ u_d + \frac{p_n}{\alpha} & \text{in } \mathcal{I}_n, \end{cases} \\ -\Delta p_n &= z_d - y_n \text{ in } \Omega. \end{aligned}$$

and set

$$u_n = \begin{cases} b & \text{in } \mathcal{A}_n \\ u_d + \frac{p_n}{\alpha} & \text{in } \mathcal{I}_n, \end{cases}$$

5. Set $\lambda_n = p_n - \alpha(u_n - u_d)$, update $n = n + 1$ and goto 2.

The existence of the triple (y_n, u_n, p_n) satisfying the system of step 4 of the Algorithm follows from the fact that it constitutes the optimality system for the auxiliary problem

$$(\mathcal{P}_{aux}) \quad \min \{ J(y, u) \mid y \in H_o^1(\Omega), \quad -\Delta y = u \text{ in } \Omega, \quad u = b \text{ on } \mathcal{A}_n \}$$

which has (y_n, u_n) as unique solution.

We may use different initialization schemes. The one that was used most frequently is the following one:

$$(2.1) \quad \begin{cases} u_o = b, \\ -\Delta y_o = u_o, \quad y_o \in H_o^1(\Omega), \\ -\Delta p_o = z_d - y_o, \quad p_o \in H_o^1(\Omega), \\ \lambda_o = \max(0, \alpha(u_d - b) + p_o). \end{cases}$$

This choice of initialization has the property of feasibility. Alternatively, we tested the algorithm with initialization as the solution of the unconstrained problem, i.e.

$$(2.2) \quad \begin{cases} \lambda_o = 0 \\ -\Delta y_o = u_d + \frac{p_o}{\alpha}, \quad y_o \in H_o^1(\Omega), \\ -\Delta p_o = z_d - y_o, \quad p_o \in H_o^1(\Omega), \\ u_o = u_d + \frac{p_o}{\alpha}. \end{cases}$$

For all examples the first initialization behaved better or equal to the second.

The initialization process (2.1) has the property that the first set \mathcal{A}_1 is always included in the active set \mathcal{A}^* of problem (\mathcal{P}) . More precisely we have

LEMMA 2.1. *If (u_o, y_o, λ_o) are given by (2.1) with $u_o \geq u^*$; then $\lambda_o \leq \lambda^*$.*

In addition, if $u_o = b$ then $\mathcal{A}_1 \subset \mathcal{A}^$.*

Proof - By construction

$$\lambda_o = \max(0, \alpha(u_d - u_o) + p_o) = \max(0, \alpha(u_d - u_o) + \Delta^{-1}(y_o - z_d)) ,$$

and as a consequence of (\mathcal{S})

$$\lambda^* = \alpha(u_d - u^*) + p^* = \alpha(u_d - u^*) + \Delta^{-1}(y^* - z_d) = \alpha(u_d - u^*) - \Delta^{-2}u^* - \Delta^{-1}z_d \geq 0.$$

It follows that

$\lambda^* - \lambda_o = \lambda^* \geq 0$ if $\alpha(u_d - u_o) + \Delta^{-1}(y_o - z_d) \leq 0$, and

$\lambda^* - \lambda_o = \alpha(u_o - u^*) + \Delta^{-2}(u_o - u^*) + \alpha(u_d - u_o) + \Delta^{-1}(y_o - z_d)$ else .

If $u_o \geq u^*$ the maximum principle yields $\Delta^{-2}(u_o - u^*) \geq 0$ and

$$\lambda^* - \lambda_o \begin{cases} = \lambda^* \geq 0 & \text{if } \alpha(u_d - u_o) + \Delta^{-1}(y_o - z_d) \leq 0 \\ \geq \alpha(u_d - u_o) + \Delta^{-1}(y_o - z_d) \geq 0 & \text{else .} \end{cases}$$

Therefore $\lambda_o \leq \lambda^*$.

In addition, if $u_o = b$ then $u_o + \frac{\lambda_o}{c} = b + \frac{\lambda_o}{c} > b$ on \mathcal{A}_1 . Consequently $\lambda_o > 0$ on \mathcal{A}_1 and $\lambda^* > 0$. It follows that $\mathcal{A}_1 \subset \mathcal{A}^*$ and the proof is complete. \blacksquare

A first convergence result which also justifies the stopping criterion in Step 3 is given in the following theorem.

THEOREM 2.1. *If there exists $n \in \mathbb{N} - \{0\}$ such that $\mathcal{A}_n = \mathcal{A}_{n+1}$ then the algorithm stops and the last iterate satisfies*

$$(\mathcal{S}_n) \quad \begin{cases} -\Delta y_n = u_n = \begin{cases} b & \text{in } \mathcal{A}_n \\ u_d + \frac{p_n}{\alpha} & \text{in } \Omega - \mathcal{A}_n , \end{cases} \\ -\Delta p_n = z_d - y_n & \text{in } \Omega . \\ \lambda_n = p_n - \alpha(u_n - u_d) , \quad u_n \in U_{ad} \end{cases}$$

with

$$(2.3) \quad \lambda_n = 0 \text{ on } \mathcal{I}_n \text{ and } \lambda_n > 0 \text{ on } \mathcal{A}_n .$$

Therefore, the last iterate is the solution of the original optimality system (\mathcal{S}) .

Proof - If there exists $n \in \mathbb{N} - \{0\}$ such that $\mathcal{A}_n = \mathcal{A}_{n+1}$, then it is clear that algorithm stops and the last iterate satisfies (\mathcal{S}_n) by construction except possibly for $u_n \in U_{ad}$.

Thus we have to prove $u_n \in U_{ad}$ and (2.3).

- On \mathcal{I}_n we have $\lambda_n = 0$ by step 5 of the Algorithm. Moreover $u_n + \frac{\lambda_n}{c} = u_n \leq b$, since $\mathcal{I}_n = \mathcal{I}_{n+1}$.
- On \mathcal{A}_n we get $u_n = b$ and $u_n + \frac{\lambda_n}{c} > b$ since $\mathcal{A}_n = \mathcal{A}_{n+1}$. Therefore $\lambda_n > 0$ on \mathcal{A}_n and $u_n \in U_{ad}$.

To prove that the last iterate is a solution of the original optimality system (\mathcal{S}) , it remains to show that

$$\lambda_n = c[u_n + \frac{\lambda_n}{c} - \Pi(u_n + \frac{\lambda_n}{c})] .$$

- On \mathcal{I}_n we have $\lambda_n = 0$ and $u_n + \frac{\lambda_n}{c} = u_n \leq b$. It follows that

$$u_n + \frac{\lambda_n}{c} - \Pi(u_n + \frac{\lambda_n}{c}) = u_n - \Pi(u_n) = 0 = \lambda_n .$$

- On \mathcal{A}_n we get $u_n = b$, $\lambda_n > 0$ and therefore

$$c[u_n + \frac{\lambda_n}{c} - \Pi(u_n + \frac{\lambda_n}{c})] = c[b + \frac{\lambda_n}{c} - b] = \lambda_n .$$

■

Now we give a structural property of the algorithm :

LEMMA 2.2. *If u_n is feasible for some $n \in \mathbb{N} - \{0\}$ (i.e. $u_n \leq b$) then $\mathcal{A}_{n+1} \subset \mathcal{A}_n$.*

Proof - On \mathcal{I}_n we get $\lambda_n = 0$ by construction, so that $u_n + \frac{\lambda_n}{c} = u_n \leq b$ (because of feasibility). This implies $\mathcal{I}_n \subset \mathcal{I}_{n+1}$ and consequently $\mathcal{A}_{n+1} \subset \mathcal{A}_n$. ■

Note that Theorem 2.1 and in particular (2.3) does not utilize or imply strict complementarity. In fact, if (2.3) holds, then the set of x for which $u_n(x) = b$ and $\lambda_n(x) = 0$ is contained in \mathcal{I}_n .

We end this section with “simple cases” where we may conclude easily that the algorithm is convergent.

THEOREM 2.2. *For initialization (2.1), the Algorithm converges in one iteration in the following cases*

1. $z_d \leq 0$, $u_d = 0$, $b \geq 0$ and the solution to $-\alpha\Delta u - \Delta^{-1}u = z_d$ is nonpositive.
2. $z_d \geq 0$, $b \leq 0$, $u_d > b$ or

$z_d \geq 0$, $b \leq 0$, $u_d \geq b$ and $z_d + \Delta^{-1}b$ is not zero as element in $L^2(\Omega)$.

Proof - Let us first examine case 1. The maximum principle implies that $-\Delta^{-1}u_o \geq 0$. Consequently $z_d + \Delta^{-1}u_o \leq 0$ and by a second application of the maximum principle

$$-\Delta^{-1}(z_d + \Delta^{-1}u_o) \leq 0 .$$

Together with the fact that $u_d - b = -b \leq 0$, this implies

$$\lambda_o = \max(0, \alpha(u_d - b) - \Delta^{-1}(z_d + \Delta^{-1}u_o)) = 0 .$$

Therefore $\mathcal{A}_1 = \emptyset$ and $\mathcal{I}_1 = \Omega$.

Using the first iteration we obtain $u_1 = \frac{p_1}{\alpha}$ in Ω . Moreover $-\Delta y_1 = u_1$ and $-\Delta p_1 = z_d - y_1$ imply that

$$-\alpha\Delta u_1 - \Delta^{-1}u_1 = z_d .$$

By assumption u_1 is feasible. Therefore $\mathcal{A}_2 = \mathcal{A}_1 = \emptyset$ and by Theorem 2.1 the algorithm stops at the solution to (\mathcal{P}) .

Now we consider case 2. By assumption and due to (2.1) we have $z_d \geq 0$, $b \leq 0$, $\lambda_o \geq 0$ and $\mathcal{A}_1 = \{ \lambda_o > 0 \}$. Due to the maximum principle $-\Delta^{-1}u_o \leq 0$ and

$$p_o = -\Delta^{-1}(z_d - y_o) = -\Delta^{-1}[z_d - (-\Delta^{-1}u_o)] \geq 0 .$$

Moreover if $z_d + \Delta^{-1}b$ is not the zero element in $L^2(\Omega)$, then $p_o > 0$ in Ω and $\alpha(u_d - b) + p_o > \alpha(u_d - b)$. If $u_d > b$ or ($u_d = b$ and $z_d + \Delta^{-1}b \neq 0$) then $\lambda_o = \max(0, \alpha(u_d - b) + p_o) > 0$ in Ω (and $\lambda_o = 0$ on $\partial\Omega$). Consequently $\mathcal{A}_1 = \Omega$ and $u_1 = b$, $\lambda_1 = -\Delta^{-1}(z_d + \Delta^{-1}b) + \alpha(u_d - b) > 0$. This yields $\mathcal{A}_2 = \mathcal{A}_1 = \Omega$ and the algorithm stops. \blacksquare

3. Convergence analysis.

3.1. The Continuous Case. The convergence analysis of the Algorithm is based on the decrease of appropriately chosen merit functions. For that purpose we define the following augmented Lagrangian functions

$$L_c(y, u, \lambda) = J(y, u) + (\lambda, \hat{g}_c(u, \lambda)) + \frac{c}{2} \|\hat{g}_c(u, \lambda)\|^2 \quad \text{and} \quad \hat{L}_c(y, u, \lambda) = L_c(y, u, \lambda^+),$$

where (\cdot, \cdot) is the $L^2(\Omega)$ -inner product, $\|\cdot\|$ is the $L^2(\Omega)$ -norm, $\lambda^+ = \max(\lambda, 0)$ and $\hat{g}_c(u, \lambda) = \max(g(u), -\frac{\lambda}{c})$ with $g(u) = u - b$. Further $(\cdot, \cdot)_{|_S}$ and $\|\cdot\|_{|_S}$ denote the L^2 -inner product and norm on a measurable subset $S \subset \Omega$. Note that the mapping

$$u \mapsto (\lambda, \hat{g}_c(u, \lambda)) + \frac{c}{2} \|\hat{g}_c(u, \lambda)\|^2,$$

is \mathcal{C}^1 , which is not the case for the function given by

$$u \mapsto (\lambda, g(u)) + \frac{c}{2} \|\max(g(u), 0)\|^2.$$

The following relationship between primal and dual variables will be essential.

LEMMA 3.1. *For all $n \in \mathbb{N} - \{0\}$ and $(y, u) \in H_o^1(\Omega) \times L^2(\Omega)$ satisfying $-\Delta y = u$ we have*

$$(3.1) \quad J(y_n, u_n) - J(y, u) = -\frac{1}{2} \|y - y_n\|^2 - \frac{\alpha}{2} \|u - u_n\|^2 + (\lambda_n, u - u_n)_{|_{\mathcal{A}_n}}$$

Proof - Using $\|a\|^2 - \|b\|^2 = -\|a - b\|^2 + 2(a - b, a)$ and Steps 4 and 5 of the Algorithm, we find that

$$\begin{aligned} J(y_n, u_n) - J(y, u) &= -\frac{1}{2} \|y - y_n\|^2 - \frac{\alpha}{2} \|u - u_n\|^2 + (y_n - y, y_n - z_d) + \alpha (u_n - u, u_n - u_d) \\ &= -\frac{1}{2} \|y - y_n\|^2 - \frac{\alpha}{2} \|u - u_n\|^2 + (\Delta(y_n - y), p_n) + \alpha (u_n - u, u_n - u_d) \\ &= -\frac{1}{2} \|y - y_n\|^2 - \frac{\alpha}{2} \|u - u_n\|^2 + (u_n - u, -p_n + \alpha(u_n - u_d)) \\ &= -\frac{1}{2} \|y - y_n\|^2 - \frac{\alpha}{2} \|u - u_n\|^2 + (u - u_n, \lambda_n). \end{aligned}$$

As $\lambda_n = 0$ on \mathcal{I}_n the result follows. \blacksquare

Let us define

$$\mathcal{S}_{n-1} = \{ x \in \mathcal{A}_{n-1} \mid \lambda_{n-1}(x) \leq 0 \} \quad \text{and} \quad \mathcal{T}_{n-1} = \{ x \in \mathcal{I}_{n-1} \mid u_{n-1}(x) > b(x) \}.$$

These two sets can be paraphrased by calling \mathcal{S}_{n-1} the set of elements that the active set strategy predicts to be active at level $n - 1$ but the Lagrange multiplier indicates that they should be inactive,

and by calling \mathcal{I}_{n-1} the set of elements that was predicted to be inactive but the $n - 1$ st iteration level corrects it to be active. We note that

$$(3.2) \quad \Omega = (\mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1}) \cup \mathcal{T}_{n-1} \cup \mathcal{S}_{n-1} \cup (\mathcal{A}_{n-1} \setminus \mathcal{S}_{n-1})$$

defines a decomposition of Ω in mutually disjoint sets. Moreover we have the following relation between these sets at each level n :

$$(3.3) \quad \mathcal{I}_n = (\mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1}) \cup \mathcal{S}_{n-1}, \quad \mathcal{A}_n = (\mathcal{A}_{n-1} \setminus \mathcal{S}_{n-1}) \cup \mathcal{T}_{n-1}.$$

In fact, as $\Omega = \mathcal{I}_n \cup \mathcal{A}_n$ is sufficient to prove that

$$(\mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1}) \cup \mathcal{S}_{n-1} \subset \mathcal{I}_n \quad \text{and} \quad (\mathcal{A}_{n-1} \setminus \mathcal{S}_{n-1}) \cup \mathcal{T}_{n-1} \subset \mathcal{A}_n,$$

that is

$$\mathcal{S}_{n-1} \subset \mathcal{I}_n \quad \text{and} \quad \mathcal{T}_{n-1} \subset \mathcal{A}_n.$$

Since $\mathcal{S}_{n-1} \subset \mathcal{A}_{n-1}$ we find $u_{n-1} = b$ on \mathcal{S}_{n-1} . From the definition of \mathcal{S}_{n-1} we conclude that $\lambda_{n-1} \leq 0$ so that $u_{n-1} + \frac{\lambda_{n-1}}{c} \leq b$. This implies $\mathcal{S}_{n-1} \subset \mathcal{I}_n$. The verification of $\mathcal{T}_{n-1} \subset \mathcal{A}_n$ is quite similar.

For the convenience of the reader we present these sets in Figure 1.

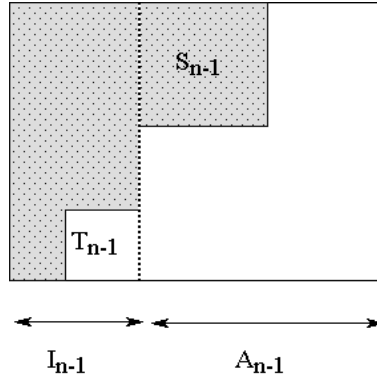


Figure 1: Decomposition of Ω at levels $n - 1$ and n

In Figure 1 the shaded region depicts \mathcal{I}_n and the white region is \mathcal{A}_n . The following table depicts the signs of primal and dual variables for two consecutive iteration levels.

	λ_{n-1}	λ_n	u_{n-1}	u_n
$\mathcal{T}_{n-1} = \mathcal{I}_{n-1} \cap \mathcal{A}_n$	0		$> b$	$= b$
$\mathcal{S}_{n-1} = \mathcal{A}_{n-1} \cap \mathcal{I}_n$	≤ 0	0	$= b$	
$\mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1} (\subset \mathcal{I}_n)$	0	0	$\leq b$	
$\mathcal{A}_{n-1} \setminus \mathcal{S}_{n-1} (\subset \mathcal{A}_n)$	> 0		$= b$	$= b$

Table 1

Below $\|\Delta^{-1}\|$ will denote the operator norm of Δ^{-1} in $\mathcal{L}(L^2(\Omega))$.

THEOREM 3.1. If $\mathcal{A}_n \neq \mathcal{A}_{n-1}$ and

$$(3.4) \quad \alpha + \gamma \leq c \leq \alpha - \frac{\alpha^2}{\gamma} + \frac{\alpha^2}{\|\Delta^{-1}\|^2}$$

for some $\gamma > 0$, then $\hat{L}_c(y_n, u_n, \lambda_n) \leq \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1})$. In addition, if the second inequality of (3.4) is strict then either $\hat{L}_c(y_n, u_n, \lambda_n) < \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1})$ or the Algorithm stops at the solution to (S).

Proof - A short computation gives

$$\begin{aligned} & (\lambda, \hat{g}_c(u, \lambda)) + \frac{c}{2} \|\hat{g}_c(u, \lambda)\|^2 \\ &= \left(\frac{1}{\sqrt{c}} \lambda, \sqrt{c} \hat{g}_c(u, \lambda) \right) + \frac{1}{2} (\sqrt{c} \hat{g}_c(u, \lambda), \sqrt{c} \hat{g}_c(u, \lambda)) \\ &= \frac{1}{2} \|\sqrt{c} \max(g(u), -\frac{\lambda}{c}) + \frac{1}{\sqrt{c}} \lambda\|^2 - \frac{1}{2c} \|\lambda\|^2 \\ &= \frac{1}{2} \|\max(\sqrt{c} g(u), -\frac{\lambda}{\sqrt{c}}) + \frac{1}{\sqrt{c}} \lambda\|^2 - \frac{1}{2c} \|\lambda\|^2 \\ &= \frac{1}{2c} \|\max(c g(u) + \lambda, 0)\|^2 - \frac{1}{2c} \|\lambda\|^2. \end{aligned}$$

Moreover for all (y, u, λ) we find

$$(3.5) \quad L_c(y, u, \lambda) = J(y, u) + \frac{1}{2c} \|\max(c g(u) + \lambda, 0)\|^2 - \frac{1}{2c} \|\lambda\|^2.$$

By assumption $\mathcal{A}_n \neq \mathcal{A}_{n-1}$ and hence $\mathcal{S}_{n-1} \cup \mathcal{T}_{n-1} \neq \emptyset$. Using (3.5) we get

$$\begin{aligned} & \hat{L}_c(y_n, u_n, \lambda_n) - \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1}) = \\ & J(y_n, u_n) - J(y_{n-1}, u_{n-1}) \\ & + \frac{1}{2c} [\|\max(c g(u_n) + \lambda_n^+, 0)\|^2 - \|\lambda_n^+\|^2 - \|\max(c g(u_{n-1}) + \lambda_{n-1}^+, 0)\|^2 + \|\lambda_{n-1}^+\|^2] \end{aligned}$$

and by (3.1)

$$(3.6) \quad \begin{aligned} & \hat{L}_c(y_n, u_n, \lambda_n) - \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1}) = \\ & -\frac{1}{2} \|y_{n-1} - y_n\|^2 - \frac{\alpha}{2} \|u_{n-1} - u_n\|^2 + (u_{n-1} - u_n, \lambda_n)_{\mathcal{T}_{n-1}} + \\ & \frac{1}{2c} [\|\max(c g(u_n) + \lambda_n^+, 0)\|^2 - \|\lambda_n^+\|^2 - \|\max(c g(u_{n-1}) + \lambda_{n-1}^+, 0)\|^2 + \|\lambda_{n-1}^+\|^2]. \end{aligned}$$

It will be convenient to introduce $d(x) =$

$$|\max(c g(u_n(x)) + \lambda_n^+(x), 0)|^2 - |\lambda_n^+(x)|^2 - |\max(c g(u_{n-1}(x)) + \lambda_{n-1}^+(x), 0)|^2 + |\lambda_{n-1}^+(x)|^2.$$

Let us estimate d on the four distinct subsets of Ω according to (3.2).

- On $\mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1}$ we have $\lambda_n(x) = \lambda_{n-1}(x) = 0$, $u_{n-1}(x) \leq b(x)$ ($g(u_{n-1}(x)) \leq 0$) and

$$d(x) = |\max(c g(u_n(x)), 0)|^2 - |\max(c g(u_{n-1}(x)), 0)|^2 \leq c^2 |u_n(x) - u_{n-1}(x)|^2.$$

Moreover as $\lambda_n = p_n - \alpha(u_n - u_d) = 0$ and $\lambda_{n-1} = p_{n-1} - \alpha(u_{n-1} - u_d) = 0$ we have $u_n(x) - u_{n-1}(x) = \frac{p_n(x) - p_{n-1}(x)}{\alpha}$ so that

$$|u_n(x) - u_{n-1}(x)| \leq \frac{1}{\alpha} |p_n(x) - p_{n-1}(x)| \quad \text{on } \mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1}$$

• On \mathcal{S}_{n-1} , $\lambda_n(x) = 0$, $\lambda_{n-1}(x) \leq 0$, $g(u_{n-1}(x)) = 0$, so that $d(x) = |\max(cg(u_n(x)), 0)|^2$. Here we used the positivity of λ^+ to get $\lambda_{n-1}^+(x) = 0$. To estimate $d(x)$ in detail we consider first the case where $u_n(x) \geq b(x)$. Since $x \in \mathcal{S}_{n-1} \subset \mathcal{I}_n$ we obtain $\lambda_n(x) = p_n(x) - \alpha[u_n(x) - u_d(x)] = 0$ and hence $u_n(x) = \frac{p_n(x)}{\alpha} + u_d(x)$. Moreover, $\lambda_{n-1}(x) = p_{n-1}(x) - \alpha[u_{n-1}(x) - u_d(x)] \leq 0$ so that $u_d(x) - b(x) \leq -\frac{p_{n-1}(x)}{\alpha}$ where we used $u_{n-1}(x) = b(x)$. Since by assumption $u_n(x) \geq b$ these estimates imply

$$|u_n(x) - u_{n-1}(x)| = u_n(x) - b(x) = \frac{p_n(x)}{\alpha} + u_d(x) - b(x) \leq \frac{p_n(x)}{\alpha} - \frac{p_{n-1}(x)}{\alpha} = \frac{1}{\alpha} |p_n(x) - p_{n-1}(x)|.$$

In addition it is clear that on the set \mathcal{I}_n :

$$d(x) = |\max(cg(u_n(x)), 0)|^2 \leq c^2 |u_n(x) - u_{n-1}(x)|^2.$$

In the second case, $u_n(x) < b(x)$ so that $\max(cg(u_n(x)), 0) = 0$ and $d(x) = 0$.

Finally we have a precise estimate on the whole set \mathcal{I}_n . Let us denote

$$\mathcal{I}_n^* = \mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1} \cup \{x \in \mathcal{S}_{n-1} \mid u_n(x) \geq b(x)\} = \mathcal{I}_n \setminus \{x \in \mathcal{S}_{n-1} \mid u_n(x) < b(x)\};$$

then

$$(3.7) \quad \int_{\mathcal{I}_n} d(x) dx = \int_{\mathcal{I}_n^*} d(x) dx = c^2 \|\max(g(u_n), 0)\|_{\mathcal{I}_n^*}^2 \leq c^2 \|u_n - u_{n-1}\|_{\mathcal{I}_n^*}^2.$$

We note that we have proved in addition that

$$(3.8) \quad \|u_n - u_{n-1}\|_{\mathcal{I}_n^*} \leq \frac{\|\Delta^{-1}\|}{\alpha} \|y_n - y_{n-1}\|.$$

• On $\mathcal{A}_{n-1} \setminus \mathcal{S}_{n-1}$, we have $g(u_{n-1}(x)) = g(u_n(x)) = 0$, $\lambda_{n-1}(x) > 0$ and hence

$$(3.9) \quad d(x) = |\max(\lambda_n^+(x), 0)|^2 - |\lambda_{n-1}^+(x)|^2 \leq 0.$$

• On \mathcal{T}_{n-1} we have $\lambda_{n-1}(x) = 0$, $g(u_n(x)) = 0$, $g(u_{n-1}(x)) > 0$ and thus

$$(3.10) \quad d(x) = -c^2 |g(u_{n-1}(x))|^2 = -c^2 |u_n(x) - u_{n-1}(x)|^2.$$

Next we estimate the term $(\lambda_n, u_{n-1} - u_n)_{\mathcal{T}_{n-1}}$ in (3.6):

$$\begin{aligned} (\lambda_n, u_{n-1} - u_n)_{\mathcal{T}_{n-1}} &= (\lambda_n - \lambda_{n-1}, u_{n-1} - u_n)_{\mathcal{T}_{n-1}} \\ &= (p_n - p_{n-1}, u_{n-1} - u_n)_{\mathcal{T}_{n-1}} + \alpha \|u_n - u_{n-1}\|_{\mathcal{T}_{n-1}}^2. \end{aligned}$$

and therefore

$$(3.11) \quad (\lambda_n, u_{n-1} - u_n)_{\mathcal{T}_{n-1}} \leq \|\Delta^{-1}\| \|y_n - y_{n-1}\|_{\Omega} \|u_n - u_{n-1}\|_{\mathcal{T}_{n-1}} + \alpha \|u_n - u_{n-1}\|_{\mathcal{T}_{n-1}}^2.$$

Inserting (3.7-3.11) into (3.6) we find

$$\begin{aligned}
& \hat{L}_c(y_n, u_n, \lambda_n) - \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1}) \leq \\
(3.12) \quad & -\frac{1}{2}\|y_{n-1} - y_n\|^2 - \frac{\alpha}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_n^*}^2 - \frac{\alpha}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_n \setminus \mathcal{I}_n^*}^2 - \frac{\alpha}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_{n-1}}^2 \\
& + \|\Delta^{-1}\| \|y_n - y_{n-1}\| \Omega \|u_n - u_{n-1}\|_{\mathcal{I}_{n-1}} + \alpha \|u_n - u_{n-1}\|_{\mathcal{I}_{n-1}}^2 \\
& + \frac{c}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_n^*}^2 - \frac{c}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_{n-1}}^2 .
\end{aligned}$$

Using $ab \leq \frac{1}{2}(\frac{a^2}{\rho} + \rho b^2)$ for every $\rho > 0$ and relation (3.8), we get for $c \geq \alpha$

$$\begin{aligned}
& \hat{L}_c(y_n, u_n, \lambda_n) - \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1}) \leq \\
& -\frac{1}{2}\|y_{n-1} - y_n\|^2 + \frac{(c-\alpha)}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_n^*}^2 + \frac{(\alpha-c)}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_{n-1}}^2 \\
& + \frac{\|\Delta^{-1}\|}{2\rho}\|y_{n-1} - y_n\|^2 + \frac{\rho\|\Delta^{-1}\|}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_{n-1}}^2 \leq \\
& -\frac{1}{2}\|y_{n-1} - y_n\|^2 + \frac{(c-\alpha)\|\Delta^{-1}\|^2}{2\alpha^2}\|y_{n-1} - y_n\|^2 \\
& + \frac{\alpha-c+\rho\|\Delta^{-1}\|}{2}\|u_{n-1} - u_n\|_{\mathcal{I}_{n-1}}^2 + \frac{\|\Delta^{-1}\|}{2\rho}\|y_{n-1} - y_n\|^2 = \\
& \frac{1}{2}\left[(c-\alpha)\frac{\|\Delta^{-1}\|^2}{\alpha^2} + \frac{\|\Delta^{-1}\|}{\rho} - 1 \right] \|y_{n-1} - y_n\|^2 + \frac{1}{2}(\alpha + \rho\|\Delta^{-1}\| - c)\|u_{n-1} - u_n\|_{\mathcal{I}_{n-1}}^2 .
\end{aligned}$$

Setting $\gamma = \rho\|\Delta^{-1}\|$ then $\hat{L}_c(y_n, u_n, \lambda_n) \leq \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1})$ provided that

$$\left[\left[\frac{(c-\alpha)}{\alpha^2} + \frac{1}{\gamma} \right] \|\Delta^{-1}\|^2 - 1 \right] \leq 0 \quad \text{and} \quad \alpha + \gamma - c \leq 0 .$$

The latter condition is equivalent to

$$(3.4) \quad \alpha + \gamma \leq c \leq \alpha - \frac{\alpha^2}{\gamma} + \frac{\alpha^2}{\|\Delta^{-1}\|^2} .$$

If the second inequality is strict then $\hat{L}_c(y_n, u_n, \lambda_n) < \hat{L}_c(y_{n-1}, u_{n-1}, \lambda_{n-1})$ except if $y_n = y_{n-1}$. In this latter case $u_n = u_{n-1}$ as well and the Algorithm stops at the solution to (S). \blacksquare

REMARK 3.1. Note that for the choice $\gamma = \alpha$ condition (3.4) is equivalent to

$$(3.13) \quad 2\alpha \leq c \leq \frac{\alpha^2}{\|\Delta^{-1}\|^2} .$$

REMARK 3.2. If there exists γ such that (3.4) holds, then necessarily

$$c > \alpha \geq 2\|\Delta^{-1}\|^2$$

holds. Indeed, assume that $\alpha < 2\|\Delta^{-1}\|^2$. Then

$$\alpha + \gamma < \alpha - \frac{\alpha^2}{\gamma} + 2\alpha ,$$

that is

$$\gamma^2 - 2\alpha\gamma + \alpha^2 = (\gamma - \alpha)^2 < 0 ,$$

which is a contradiction.

3.2. The Discrete Case. So far we have given a sufficient condition for \hat{L}_c to act as a merit function for which the Algorithm has a strict descent property. In particular this eliminates the possibility of chattering of the Algorithm: it will not return to the same active set a second time. If the control and state spaces are discretized then the descent property can be used to argue convergence in a finite number of steps. More precisely, assume that a finite difference or finite element based approximation to (\mathcal{P}) results in

$$(P^{N,M}) \quad \begin{aligned} \min \quad & J^{N,M}(Y, U) = \frac{1}{2} \|M_1^{\frac{1}{2}}(Y - Z_d)\|_{\mathbb{R}^N}^2 + \frac{\alpha}{2} \|M_2^{\frac{1}{2}}(U - U_d)\|_{\mathbb{R}^M}^2, \\ & S Y = M_3 U, \\ & U \leq B. \end{aligned}$$

Here Y and Z_d denotes vectors in \mathbb{R}^N corresponding to the discretization of y and z_d , and U , U_d and B denote vectors in \mathbb{R}^M , corresponding to the discretizations of u , u_d and b . Further M_1 , S and M_2 are respectively $N \times N$, $N \times N$ and $M \times M$ positive definite matrices while M_3 is an $N \times M$ matrix. The norms in $(P^{N,M})$ denote Euclidian norms and the inequality is understood coordinatewise. Finally, it is assumed that M_2 is a diagonal matrix. It is simple to argue the existence of a solution (Y^*, U^*) to $(P^{N,M})$. A first order optimality system is given by

$$(3.14) \quad \begin{cases} S Y^* &= M_3 U^* \\ S P^* &= -M_1(Y^* - Z_d) \\ U^* &= U_d + \frac{1}{\alpha} M_2^{-1} (M_3^\top P^* - \Lambda^*) \\ \Lambda^* &= c \max(0, U^* + \frac{1}{c} \Lambda^* - B), \end{cases}$$

with $(P^*, \Lambda^*) \in \mathbb{R}^N \times \mathbb{R}^M$, for every $c > 0$. Here max is understood coordinatewise. The algorithm for the discretized problem is given next.

Discretized Algorithm

1. Initialization : choose Y^o , U^o and Λ^o , and set $n = 1$.
2. Determine the following subsets of $\{1, \dots, M\}$:

$$A_n = \{ i \mid U_i^{n-1} + \frac{1}{c} \Lambda_i^{n-1} > B_i \}, \quad I_n = \{1, \dots, M\} \setminus A_n.$$

3. If $n \geq 2$ and $A_n = A_{n-1}$ then STOP.
4. Else, find $(Y^n, P^n) \in \mathbb{R}^N \times \mathbb{R}^N$ such that

$$\begin{aligned} S Y^n &= M_3 \begin{cases} B & \text{in } A_n \\ U_d + \frac{1}{\alpha} M_2^{-1} M_3^\top P^n & \text{in } I_n, \end{cases} \\ S P^n &= -M_1(Y^n - Z_d) \end{aligned}$$

and set

$$U^n = \begin{cases} B & \text{in } A_n \\ U_d + \frac{1}{\alpha} M_2^{-1} M_3^\top P^n & \text{in } I_n, \end{cases}$$

5. Set $\Lambda^n = M_3^\top P^n - \alpha M_2(U^n - U_d)$, update $n = n + 1$ and goto 2.

The following corollary describing properties of the Discretized Algorithm can be obtained with techniques analogous to those utilized above for analysing the continuous Algorithm. We shall denote

$$\underline{m}_2 = \min_i (M_2)_{i,i}, \quad \overline{m}_2 = \max_i (M_2)_{i,i} \text{ and } K = \|M_2^{-1} M_3^\top\| \|S^{-1} M_1\| .$$

COROLLARY 3.1. *If*

$$(3.15) \quad \overline{m}_2 (\alpha + \gamma) \leq c < \alpha \underline{m}_2 - \frac{\alpha^2}{\gamma} + \frac{\alpha^2 \|M_1\|}{K}$$

holds for some $\gamma > 0$ then the Discretized Algorithm converges in finitely many steps to the solution of (\mathcal{P}^N) .

Proof - First we observe that if the Discretized Algorithm stops in Step 3 then the current iterate gives the unique solution. Then we show with an argument analogous to that of the proof of Theorem 3.1 that with (3.15) holding, we have $L_c^{N,M}(Y_n, U_n, \Lambda_n) < L_c^{N,M}(Y_{n-1}, U_{n-1}, \Lambda_{n-1})$ or $(Y_n, U_n) = (Y_{n-1}, U_{n-1})$, where the discretized merit function is given by

$$L_c^{N,M}(Y, U, \Lambda) = \frac{1}{2} \|M_1^{\frac{1}{2}} (Y - Z_d)\|_{\mathbb{R}^N}^2 + \frac{\alpha}{2} \|M_2^{\frac{1}{2}} (U - U_d)\|_{\mathbb{R}^M}^2 + (\Lambda, \hat{g}_c(U, \Lambda))_{\mathbb{R}^M} + \frac{c}{2} \|\hat{g}_c(U, \Lambda)\|_{\mathbb{R}^M}^2 ,$$

with $\hat{g}_c(U, \Lambda) = (\max(U_1 - B_1, -\frac{\Lambda_1}{c}), \dots, \max(U_M - B_M, -\frac{\Lambda_M}{c}))^\top$. If $(Y_n, U_n) = (Y_{n-1}, U_{n-1})$ then $A_{n+1} = A_n$ and the Discretized Algorithm stops at the solution. The case $L_c^{N,M}(Y_n, U_n, \Lambda_n) < L_c^{N,M}(Y_{n-1}, U_{n-1}, \Lambda_{n-1})$ cannot occur for infinitely many n since there are only finitely many different combinations of active index sets. In fact, assume that there exists $p < n$ such that $A_n = A_p$ and $I_n = I_p$. Since (Y_n, U_n) is a solution of the optimality system of Step 4 if and only if (Y_n, U_n) is the unique solution of

$$\min\{ J^{N,M}(y, u) \mid S Y = M_3 U, U = B \text{ in } A_n \} ,$$

it follows that $Y_n = Y_p$, $U_n = U_p$ and $\Lambda_n = \Lambda_p$. This contradicts $L_c^{N,M}(Y_n, U_n, \Lambda_n) < L_c^{N,M}(Y_p, U_p, \Lambda_p)$ and ends the proof. \blacksquare

REMARK 3.3. *Note that for $\gamma = \frac{\alpha}{\underline{m}_2}$ condition (3.15) is satisfied if*

$$\overline{m}_2 \alpha \left(1 + \frac{1}{\underline{m}_2}\right) \leq c < \frac{\alpha^2 \|M_1\|}{K} .$$

Therefore, one can choose $c = \overline{m}_2 \alpha \left(1 + \frac{1}{\underline{m}_2}\right)$ for any

$$\alpha > \frac{\overline{m}_2 K}{\|M_1\|} \left(1 + \frac{1}{\underline{m}_2}\right) .$$

4. Ascent properties of Algorithm. In the previous section sufficient conditions for convergence of the Algorithm in terms of α , c and $\|\Delta^{-1}\|$ were given. Numerical experiments showed that the Algorithm converges also for values of α , c and $\|\Delta^{-1}\|$ which do not satisfy the conditions of Theorems 3.1. In fact the only possibility of constructing an example for which the Algorithm has some difficulties (which will be made precise in the following section) is based on violating the strict complementarity condition.

Thus one is challenged to further justify theoretically the efficient behavior of the Algorithm. In the tests that were performed it was observed that the cost functional was always increasing so that in practice the Algorithm behaves like an infeasible algorithm. To parallel theoretically this behavior of the Algorithm as far as possible, we slightly modify the Algorithm. For the modified Algorithm an ascent property of the cost J will be shown.

Modified Algorithm

1. Initialization : choose u_o , y_o and λ_o ; set $n = 1$.

2. (a) Determine the following subsets of Ω :

$$\mathcal{A}_n = \{ x \mid u_{n-1}(x) + \frac{\lambda_{n-1}(x)}{c} > b \} , \quad \mathcal{I}_n = \{ x \mid u_{n-1}(x) + \frac{\lambda_{n-1}(x)}{c} \leq b \} ,$$

(b) and find $(\tilde{y}, \tilde{p}) \in H_o^1(\Omega) \times H_o^1(\Omega)$ such that

$$\begin{aligned} -\Delta \tilde{y} &= \begin{cases} b & \text{in } \mathcal{A}_n \\ u_d + \frac{\tilde{p}}{\alpha} & \text{in } \mathcal{I}_n , \end{cases} \\ -\Delta \tilde{p} &= z_d - \tilde{y} \text{ in } \Omega . \end{aligned}$$

and set

$$\tilde{u} = \begin{cases} b & \text{in } \mathcal{A}_n \\ u_d + \frac{\tilde{p}}{\alpha} & \text{in } \mathcal{I}_n , \end{cases}$$

3. $\tilde{\lambda} = \tilde{p} - \alpha(\tilde{u} - u_d)$.

4. Set

$$\tilde{\mathcal{A}} = \{ x \mid \tilde{u}(x) + \frac{\tilde{\lambda}(x)}{c} > b \} .$$

If $\tilde{\mathcal{A}} = \mathcal{A}_n$ then STOP, else goto 5.

5. Check for $J(\tilde{y}, \tilde{u}) > J(y_{n-1}, u_{n-1})$.

(a) If $J(\tilde{y}, \tilde{u}) > J(y_{n-1}, u_{n-1})$ then

$n = n + 1$, $y_n = \tilde{y}$, $u_n = \tilde{u}$, $\lambda_n = \tilde{\lambda}$ and goto 2a.

(b) Otherwise, determine

$$\mathcal{T}_{n-1} = \{ x \in \mathcal{I}_{n-1} \mid u_{n-1}(x) > b \} .$$

- If measure of \mathcal{T}_{n-1} is null then STOP;
- else set

$$\mathcal{A}_n = \mathcal{A}_{n-1} \cup \mathcal{T}_{n-1} , \quad \mathcal{I}_n = \mathcal{I}_{n-1} \setminus \mathcal{T}_{n-1} ,$$

then goto 2b.

THEOREM 4.1. *If the Modified Algorithm stops in Step 4, then $(\tilde{u}, \tilde{y}, \tilde{\lambda})$ is the solution to (S). If it never stops in Step 5b, then the sequence $J(y_n, u_n)$ ($n \geq 2$) is strictly increasing and converges to some J^* .*

Proof - Let us first assume that the algorithm stops in Step 4. In case \mathcal{A}_n is calculated from 2a then $(\tilde{u}, \tilde{y}, \tilde{\lambda})$ is the solution to (S) by Theorem 2.1. If \mathcal{A}_n is determined from 5b then an argument

analogous to that used in the proof of Theorem 2.1 allows to argue that again $(\tilde{u}, \tilde{y}, \tilde{\lambda})$ is the solution to (\mathcal{S}) .

Next we assume that algorithm never stops in Step 4. Let us consider an iteration level, where the check for ascent in Step 5a is not passed. Consequently \mathcal{A}_n and \mathcal{I}_n are redefined according to step 5b and (\tilde{y}, \tilde{u}) are recalculated from 2b. We have already noticed that (\tilde{y}, \tilde{u}) is a solution of the optimality system of Step 2b if and only if (\tilde{y}, \tilde{u}) is the unique solution of

$$(\mathcal{P}_{aux}) \quad \min\{ J(y, u) \mid -\Delta y = u \text{ in } \Omega, y \in H_o^1(\Omega), u = b \text{ in } \mathcal{A}_n \} .$$

Since $\mathcal{A}_n = \mathcal{A}_{n-1} \cup \mathcal{T}_{n-1}$ strictly contains \mathcal{A}_{n-1} it necessary follows that

$$(4.1) \quad J(y_{n-1}, u_{n-1}) \leq J(\tilde{y}, \tilde{u}) .$$

It will next be shown that equality in (4.1) is impossible. In fact if $J(\tilde{y}, \tilde{u}) = J(y_{n-1}, u_{n-1})$ then due to uniqueness of the solution to (\mathcal{P}_{aux}) it follows that $(\tilde{y}, \tilde{u}) = (y_{n-1}, u_{n-1})$ and consequently $\tilde{\lambda} = \lambda_{n-1}$. On $\mathcal{A}_n = \mathcal{A}_{n-1} \cup \mathcal{T}_{n-1}$, we get $\tilde{u} = b = u_{n-1}$. This implies that $u_{n-1} = b$ on \mathcal{T}_{n-1} and gives a contradiction to the assumption that the measure of \mathcal{T}_{n-1} is non null. Hence $J(y_{n-1}, u_{n-1}) = J(\tilde{y}, \tilde{u})$ is impossible. Together with (4.1) it follows that $J(y_{n-1}, u_{n-1}) < J(\tilde{y}, \tilde{u})$ and thus the sequence $\{J(y_n, u_n)\}$ generated by the Modified Algorithm is strictly increasing. The pair (y_b, b) with $-\Delta y_b = b$ in Ω is feasible for all (\mathcal{P}_{aux}) so that $J(y_n, u_n) \leq J(y_b, b)$. It follows that $J(y_n, u_n)$ is convergent to some J^* .

We note, in addition that \tilde{u} is feasible since $\tilde{u} = u_{n-1} = u_{n-1} + \frac{\lambda_{n-1}}{c} \leq b$ on \mathcal{I}_n ($\lambda_{n-1} = \tilde{\lambda} = 0$ on \mathcal{I}_n). ■

The previous result can be strengthened in the case where (\mathcal{P}) is discretized as in subsection 3.1.

COROLLARY 4.1. *If the Modified Algorithm is discretized as described in the previous section and if it never stops in Step 5b, then the (discretized) solution is obtained in finitely many steps.*

Proof - Unless the algorithm stops in Step 4, the values of $J^N(Y_n, U_n)$ ($n \geq 2$) are strictly increasing. As argued in the proof of Corollary 3.1 at each level of the iteration the minimization is carried out over an active set different from all those that have been computed before. As there are only finitely many different possibilities for active sets, the Modified Algorithm terminates in Step 4 at the unique solution of (\mathcal{S}) . ■

We have not found a numerical example in which the Modified Algorithm terminates in Step 5b.

5. Numerical Experiments. In this section we report on numerical tests with the proposed Algorithm. For these tests we chose $\Omega =]0, 1[\times]0, 1[$ and the five-point finite difference approximation of the Laplacian. Unless otherwise specified the discretization was carried out on a uniform mesh with grid size 1/50.

For the chosen dimension $\|\Delta^{-1}\| = \frac{1}{2\pi^2}$ so that $\frac{1}{\|\Delta^{-1}\|^2} = 4\pi^4 \simeq 390$. Relation (3.13) which is required for the applicability of Theorem 3.1 is satisfied if $\alpha \geq 5 \cdot 10^{-3}$ to get the convergence via Theorem 3.1. Nevertheless we have also tested the method for smaller values of α .

The tests were performed on an HP Work station using the MATLAB[©] package.

5.1. Example 1. We set

$$z_d(x_1, x_2) = \sin(2\pi x_1) \sin(2\pi x_2) \exp(2x_1)/6, \quad b \equiv 0 .$$

Several tests for different values for α , c and u_d were performed. We present two of them. For the first one (3.13) is satisfied with strict inequalities.

Table 2

Example 1.a: $u_d \equiv 0$, $\alpha = 10^{-2}$, $c = 10^{-1}$

Iteration	$\max(u_n - b)$	size of \mathcal{A}_n	$J(y_n, u_n)$	$L_c(y_n, u_n, \lambda_n)$	$\hat{L}_c(y_n, u_n, \lambda_n)$
1	4.8708e-02	1250	4.190703e-02	4.190785e-02	4.190785e-02
2	5.8230e-05	1331	4.190712e-02	4.190712e-02	4.190712e-02
3	0.0000e+00	1332	4.190712e-02	4.190712e-02	4.190712e-02
4	0.0000e+00	1332	4.190712e-02	4.190712e-02	4.190712e-02

Let us give plots of the optimal control and state.

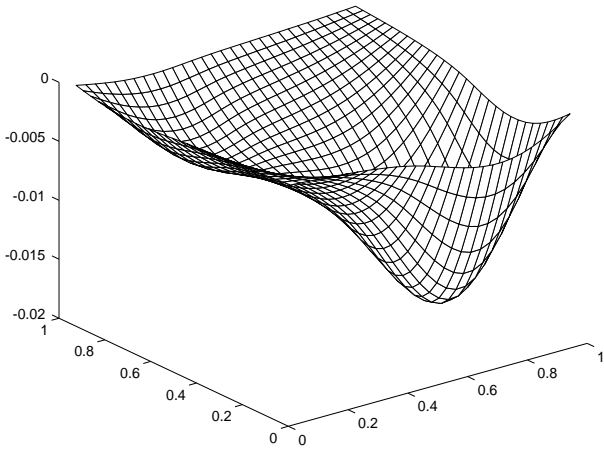


Figure 2: Optimal State

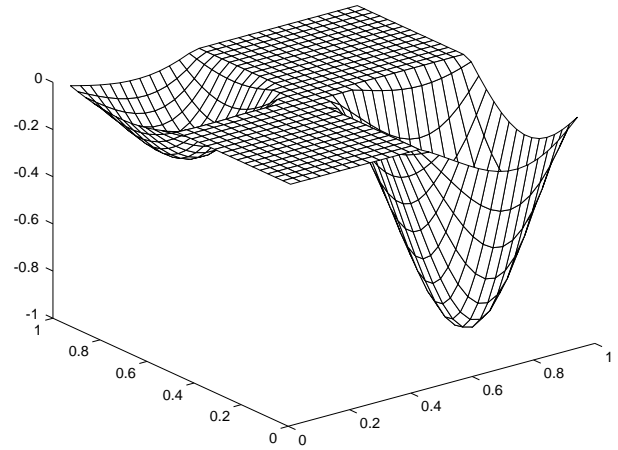


Figure 3: Optimal control

We present below a second example where (3.13) is not fulfilled because α is too small; in addition u_d has been chosen infeasible.

Table 3Example 1.b: $u_d \equiv 1$, $\alpha = 10^{-6}$, $c = 10^{-2}$

Iteration	$\max(u_n - b)$	size of \mathcal{A}_n	$J(y_n, u_n)$	$L_c(y_n, u_n, \lambda_n)$	$\hat{L}_c(y_n, u_n, \lambda_n)$
1	5.0986e+02	1250	1.734351e-02	9.858325e+00	9.858325e+00
2	4.4728e+02	1487	2.089663e-02	7.688683e+00	7.688683e+00
3	3.6796e+02	1677	2.375001e-02	5.612075e+00	5.612075e+00
4	5.8313e+02	1831	2.603213e-02	4.526200e+00	4.526200e+00
5	6.7329e+02	1944	2.782111e-02	3.657995e+00	3.657995e+00
6	5.3724e+02	2039	2.911665e-02	2.402021e+00	2.402021e+00
7	3.6175e+02	2098	2.981378e-02	1.191161e+00	1.191161e+00
8	1.5071e+02	2146	3.011540e-02	3.678089e-01	3.678089e-01
9	6.5928e+01	2178	3.018832e-02	7.796022e-02	7.796022e-02
10	2.3420e+01	2196	3.019715e-02	3.344241e-02	3.344241e-02
11	3.4889e+00	2208	3.019762e-02	3.022994e-02	3.022994e-02
12	0.0000e+00	2210	3.019762e-02	3.019762e-02	3.019762e-02
13	0.0000e+00	2210	3.019762e-02	3.019762e-02	3.019762e-02

Though the size of the set \mathcal{A}_n , in the sense of number of grid points in \mathcal{A}_n is increasing, the sequence \mathcal{A}_n does not increase monotonically. More precisely points in \mathcal{A}_n at iteration n may not belong to \mathcal{A}_{n+1} at iteration $n + 1$.

We observe numerically that the algorithm stops as soon as an iterate is feasible. So the sequence of iterates is not feasible until it reaches the solution. We could say that we have an “outer” method. We must also underline that differently from classical primal active set methods, the primal-dual method that we propose can move a lot of points from one iteration to the next.

We compared the new Algorithm to an Uzawa method for the augmented Lagrangian with Gauss-Seidel splitting. For convenience we recall that algorithm.

Algorithm : UGS

- Step 1. Initialization : Set $n = 1$ and choose $\gamma > 0$.
Choose $q_o \in L^2(\Omega)$ and $u_{-1} \in L^2(\Omega)$.
- Step 2. Choose $k_n \in \mathbb{N}$, set $u_n^{-1} = u_{n-1}$ and for $j = 0, \dots, k_n$

$$\begin{aligned} y_n^j &= \text{Arg min } \{ L_\gamma(y, u_n^{j-1}, q_n) \mid y \in H^2(\Omega) \cap H_o^1(\Omega) \} \\ u_n^j &= \text{Arg min } \{ L_\gamma(y_n^j, u, q_n) \mid u \in U_{ad} \} . \end{aligned}$$

End of the inner loop : $y_n = y_n^{k_n}$, $u_n = u_n^{k_n}$.

- Step 3.

$$q_{n+1} = q_n + \frac{\rho}{k_n + 1} \sum_{j=0}^{k_n} (Ay_n^j - u_n^j), \quad \text{where } \rho \in (0, 2\gamma] ,$$

where

$$L_\gamma(y, u, q) = J(y, u) + (q, Ay - u)_{L^2(\Omega)} + \frac{\gamma}{2} \|Ay - u\|_{L^2(\Omega)}^2.$$

For this algorithm a detailed convergence analysis was given in [BK]. Due to the splitting technique the second constrained minimization in Step 2 can be carried out by a simple algebraic manipulation. Algorithm UGS is an iterative algorithm that approximates the solution (y^*, u^*) , whereas the new Algorithm obtains the exact (discretized) solution. For Example 1a. the computing time was 61 secs whereas the Algorithm UGS with accuracy set at 10^{-3} was stopped after 105 min. At that moment the difference between the Algorithm and Algorithm UGS was

$$|J_{ugs} - J(y^*, u^*)| \approx 4.10^{-8}, \|y_{ugs} - y^*\|_{L^\infty} \approx 8.10^{-7} \text{ and } \|u_{ugs} - u^*\|_{L^\infty} \approx 4.10^{-6},$$

where the index “ugs” refers to the result from Algorithm UGS. For Example 1.b the Algorithm took 191 secs whereas Algorithm UGS was stopped after 120 min.

5.2. Example 2. The desired state z_d , b are set as in the previous example and $\alpha = 10^{-2}$, $c = 10^{-1}$. This example has been constructed such that there is no strict complementarity at the solution. More precisely we have set $u_d = b - \frac{1}{\alpha}[-\Delta^{-1}z_d + \Delta^{-2}b]$ so that the exact solution of problem (\mathcal{P}) is $u^* = b = 0$ and $\lambda^* = 0$ and hence λ^* is not positive where the constraint is active. This example was considered by means of the optimality system (\mathcal{S}) of Theorem 1.1.

Table 4

$$u_o \equiv 0 (\equiv b)$$

Iteration	$\max(u_n - b)$	size of \mathcal{A}_n	$J(y_n, u_n)$	$L_c(y_n, u_n, \lambda_n)$	$\hat{L}_c(y_n, u_n, \lambda_n)$
1	4.4409e-15	1385	4.296739e-02	4.296739e-02	4.296739e-02
2	1.2546e-14	160	4.296739e-02	4.296739e-02	4.296739e-02
3	3.2752e-15	2078	4.296739e-02	4.296739e-02	4.296739e-02
4	4.5519e-15	2308	4.296739e-02	4.296739e-02	4.296739e-02
5	4.5242e-15	1613	4.296739e-02	4.296739e-02	4.296739e-02
6	4.3299e-15	1787	4.296739e-02	4.296739e-02	4.296739e-02

Here the canonical initial guess u_o coincides with the solution u^* . From the Table 3 we observe that u_n , $J(y_n, u_n)$, $L_c(y_n, u_n)$ and $\hat{L}_c(y_n, u_n)$ remain constant while the active sets \mathcal{A}_n chatter. For different initial guesses for u_o the same type of behavior is observed, the Algorithm always reaches the optimal value for u and J in one iteration, and if the stopping criterion of the Algorithm was based on the coincidence of two consecutive values of J it would stop after one iteration. The chattering of active sets is due to lack of strict complementarity and machine precision. Let us briefly consider this phenomenon and note at first that the signs in the Algorithm are set such that at the limit we should have $\Omega = \mathcal{I}^*$ (all inactive with $\lambda^* = u^* = 0$). If $x \in \mathcal{A}_{n-1}$ then $u_{n-1}(x) = 0$ by Step 4 and $\lambda_{n-1}(x) = \pm\varepsilon$, with ε equal to the computer epsilon, will decide whether $x \in \mathcal{A}_n$ or \mathcal{I}_n , although for numerical purposes the exact pair for (u, λ) is already obtained. If $x \in \mathcal{I}_{n-1}$ then $\lambda_{n-1} = 0$ and $u_{n-1}(x) = \pm\varepsilon$ will decide whether $x \in \mathcal{A}_n$ or \mathcal{I}_n , while the influence of this choice on J or L_c is of the

order of ε^2 i.e. it is numerically zero. Therefore we decided to replace “ $> b$ ” in the definition of \mathcal{A}_n by “ $> b - \varepsilon$ ” (and $\mathcal{I}_n = \Omega \setminus \mathcal{A}_n$): the algorithm behaves now as expected and stops after 2 iterations.

5.3. Example 3. We have seen with Example 1. that the augmented Lagrangian function decreases during iterations. We show with this example that the augmented Lagrangian function may not decrease though the method is convergent and provides the exact solution. Let us precise the data :

$$z_d = \begin{cases} 200 x_1 x_2 (x_1 - \frac{1}{2})^2 (1 - x_2) & \text{if } 0 < x_1 \leq 1/2 , \\ 200 x_2 (x_1 - 1)(x_1 - \frac{1}{2})^2 (1 - x_2) & \text{if } 1/2 < x_1 \leq 1 , \end{cases}$$

$$u_d \equiv 0 , b \equiv 1 , c = 10^{-2} .$$

Table 5

Example 3.a: $\alpha = 10^{-6}$, $u_o \equiv 1$ ($\equiv b$)

Iteration	$\max(u_n - b)$	size of \mathcal{A}_n	$J(y_n, u_n)$	$L_c(y_n, u_n, \lambda_n)$	$\hat{L}_c(y_n, u_n, \lambda_n)$
1	4.1995e+02	1100	3.314755e-02	9.645226e+00	9.645226e+00
2	3.8057e+02	1370	3.672870e-02	7.943326e+00	7.943326e+00
3	3.6453e+02	1300	3.963515e-02	7.393744e+00	7.393744e+00
4	3.7512e+02	1400	4.249987e-02	7.809205e+00	7.809205e+00
5	3.8952e+02	1500	4.555558e-02	8.300084e+00	8.300084e+00
6	3.9452e+02	1600	4.880515e-02	8.320358e+00	8.320358e+00
7	3.8004e+02	1700	5.203947e-02	7.485445e+00	7.485445e+00
8	3.3858e+02	1800	5.490267e-02	5.699382e+00	5.699382e+00
9	2.6458e+02	1898	5.701220e-02	3.286759e+00	3.286759e+00
10	1.5311e+02	1986	5.811845e-02	1.093548e+00	1.093548e+00
11	8.3048e+01	2040	5.834162e-02	3.099587e-01	3.099587e-01
12	1.5809e+01	2086	5.839423e-02	5.959874e-02	5.959874e-02
13	0.0000e+00	2098	5.839438e-02	5.839438e-02	5.839438e-02
14	0.0000e+00	2098	5.839438e-02	5.839438e-02	5.839438e-02

The solution was obtained in 210 secs.

The following plot shows the influence of α on the behavior of the Lagrangian function L_c .

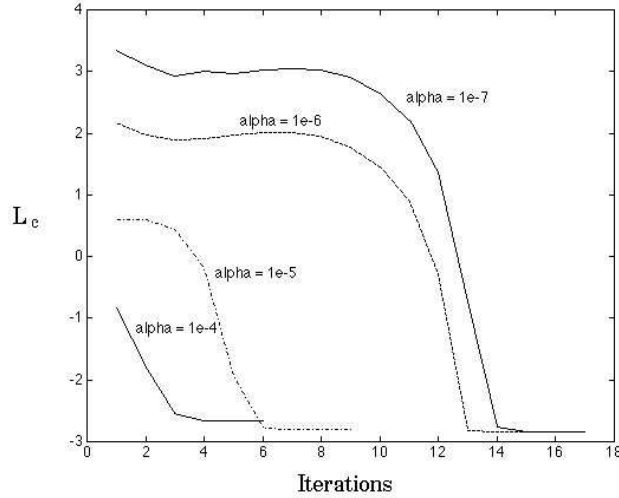


Figure 4: Influence of α on the behavior of L_c (Logarithmic scale)

We see that during the first iterations the augmented Lagrangian function does not decrease if α is too small.

However, if the initialization point is close enough the solution then this function becomes decreasing. We have tested initialization points different from b which were closer to the solution and obtained decrease of L_c . As an example we give in Table 6 the results for $\alpha = 10^{-10}$ with an initialization according to (2.1) but with u_o the solution for $\alpha = 10^{-5}$

Table 6

Example 3.b: $\alpha = 10^{-10}$, u_o given by the solution to (\mathcal{P}) for $\alpha = 10^{-5}$

Iteration	$\max(u_n - b)$	size of \mathcal{A}_n	$J(y_n, u_n)$	$L_c(y_n, u_n, \lambda_n)$
1	1.6605e+03	1986	5.696032e-02	4.889158e+01
2	1.4741e+03	2034	5.750110e-02	2.948470e+01
3	1.1542e+03	2082	5.781067e-02	1.299992e+01
4	6.8931e+02	2130	5.793424e-02	2.631407e+00
5	1.6713e+02	2168	5.795024e-02	2.198494e-01
6	1.1931e+02	2172	5.795048e-02	1.276798e-01
7	7.0091e+01	2176	5.795058e-02	7.857522e-02
8	2.0618e+01	2180	5.795061e-02	5.958497e-02
9	0.0000e+00	2182	5.795061e-02	5.795061e-02
10	0.0000e+00	2182	5.795061e-02	5.795061e-02

Note that the total number of iterations including the initialization with $\alpha = 10^{-5}$ to obtain the solution corresponding for $\alpha = 10^{-10}$ is equal to 18. If one computes the solution with initialization $u_o = b$, the number of iterations is 27 and L_c decreases after iteration 12. Thus a good initial guess can decrease the number of iterations to obtain the solution. This process was repeated successfully for smaller values of α up to $\alpha = 10^{-15}$ as well.

REFERENCES

- [B] **V. Barbu**, *Analysis and Control of Non Linear Infinite Dimensional Systems*, Math in Science and Engineering, Vol. 190, Academic Press 1993.
- [BK] **M. Bergounioux - K. Kunisch**, *Augmented Lagrangian Techniques for Elliptic State Constrained Optimal Control Problems*, SIAM Journal on Control and Optimization, Vol.35, n° 5 (1997).
- [BS] **H. Brézis - G. Stampacchia**, *Remarks on some fourth order variational inequalities*, Annali Scuola Norm. Sup. Pisa, 4 (1977), 363-371.
- [HT] **M. Heinkenschloss - F. Tröltzsch**, *Analysis of the Lagrange-SQP-Newton method for the control of a phase field equation*, Preprint, Virginia Tech.
- [IK] **K. Ito - K. Kunisch**, *Augmented Lagrangian formulation of nonsmooth convex optimization in Hilbert spaces*, Lecture Notes in Pure and Applied Mathematics, Control of Partial Differential Equations and Applications, E. Casas, Ed., Marcel Dekker, Vol. 174 (1995), 107-117.
- [KeS] **C.T. Kelley - E. Sachs**, *Approximate quasi-Newton methods*, Mathematical Programming 48 (1990), 41-70.
- [KuS] **K. Kunisch - E. Sachs**, *Reduced SQP-methods for parameter identification problems*, SIAM Journal Numerical Analysis, 29 (1992), 1793-1820.
- [K] **F.S. Kupfer**, *An infinite-dimensional convergence theory for reduced SQP-methods in Hilbert spaces*, SIAM Journal on Optimization, 6 (1996), 126-163.
- [Sch] **K. Schittkowski**, *On the convergence of a sequential quadratic programming method with an augmented Lagrangian line search function*, Math. Operationsforsch. u. Statist., Ser Optimization 14(1983), 197-216.
- [T] **F. Tröltzsch**, *An SQP-method for optimal control of a nonlinear heat equation*, Control & Cybernetics, 23 (1994), 268-288.