On a SQP–multigrid technique for nonlinear parabolic boundary control problems *

H. Goldberg and F. Tröltzsch†

Technical University of Chemnitz–Zwickau

Faculty of Mathematics

Abstract

An optimal control problem governed by the heat equation with nonlinear boundary conditions is considered. The objective functional consists of a quadratic terminal part and a quadratic regularization term. It is known, that an SQP method converges quadratically to the optimal solution of the problem. To handle the quadratic optimal control subproblems with high precision, very large scale mathematical programming problems have to be treated. The constrained problem is reduced to an unconstrained one by a method due to Bertsekas. A multigrid approach developed by Hackbusch is applied to solve the unconstrained problems. Some numerical examples illustrate the behaviour of the method.

AMS subject classification: 49M40, 49M05

Keywords: optimal control, semilinear parabolic equation, multigrid method, SQP method

1 Introduction

The behaviour of Lagrange–Newton–SQP methods for solving nonlinear optimal control problems has been the subject of several recent publications. For instance, their application to the control of ordinary differential equations was discussed by Alt [1], [2], [3], Alt and Malanowski [5], Machielsen [25]. The case of weakly singular integral equations was considered by Alt, Sontag and Tröltzsch [6]. Control problems for nonlinear partial differential equations were studied by Heinkenschloss [17], Heinkenschloss and Sachs [18], Heinkenschloss and Tröltzsch [19], Kelley and Sachs [20], [21], Kupfer and Sachs [23],

*Supported by Deutsche Forschungsgemeinschaft (DFG), under grant "Tr 302/3–1"

†D–09107 Chemnitz, Germany

Goldberg and Tröltzsch [13], and Tröltzsch [30], [31]. We refer also to a recent paper by Gill and others [10]. It is meanwhile shown in most of the cases mentioned above that the (continuous) SQP method exhibits the expected local q–quadratic convergence in spaces of type L^∞. We refer to [6], [19], [30], [31] for the proof under strong second order sufficient optimality conditions. A detailed convergence analysis assuming weaker second order conditions is contained for a simplified model in [13] and for a general class of control problems governed by semilinear parabolic equations in [32]. Owing to their quadratic convergence, these methods appear to be promising for a high precision numerical solution of control problems.

In the applications to function spaces, the method has to be linked with a discretization. It may appear on a different level. The simplest way is that of discretizing the optimal control problem as a whole, to obtain a large scale finite–dimensional optimization problem. Then the finite–dimensional SQP method is applied. This direct approach was succesfully tested for many control problems governed by ordinary differential equations and for some parabolic control problems with moderate discretization.

Direct methods may have some drawbacks in the case of partial differential equations. There are several reasons for: First of all, even a moderate accuracy for solving the state equation may lead to huge dimensions. For instance, discretizing a parabolic equation in a rectangular 2D–domain with uniform meshes of 100 node points for the time and the two space variables leads to 10^6 state variables. Note that this large number is already needed to solve the state equation with a moderate precision of order 10^{-2}.

A second problem is connected with solving the state equation itself. An explicit discretization is, as a rule, useful only for domains of simple geometry. If domains with curved boundaries have to be used, professional software should be applied. The solver for the equation may work with with irregular and moving grids, so that the discretization of the problem is not known in advance. It is defined by the solver during the process of computation. Therefore, we are particularly interested in methods, which are able to incorporate the best available solvers for the state equation (more or less as a black box).

In this paper, we continue the numerical analysis of [13], where the convergence was shown for a simplified n-dimensional model under weak second order assumptions. Numerical examples were presented there for the one–dimensional heat equation with nonlinear boundary condition. Here, we concentrate on the computational aspects, which are worked out in more detail. Moreover, we regard examples in a domain $\Omega \subset I\!R^2$, where the dimension of the discretized problems is already very large.

We pursue the following strategy: The discretization is applied to the (continuous) linear-quadratic control sub-problems. They are solved through a number of unconstrained problems. This leads to solving a sequence of fixed-point equations by a multigrid technique owing to Hackbusch [14]. We refer also to Hackbusch and Will [16]. Similar fixed point techniques were used also by Kelley and Sachs [20], [21].

Certainly, there is a lot of freedom in choosing the different steps of the method. Moreover, our method is not very effective (cf. the final comments of our paper).

Our main aim is to verify the fast convergence of the *standard* (infinite-dimensional) SQP method for a simplified class of nonlinear parabolic problems in 2D–domains by a technique close to former theoretical investigations.

The way of discretizing the problem and herafter applying the SQP method is not suitable for this purpose. It would confirm the (known) quadratic convergence of the method in finite dimensions, tending to an approximate solution. The distance between approximate and exact solution will have the order of the discretization error.

It is evident that the precision of computed optimal controls cannot be better than that for solving the state equation. Our remarks to this situation in [13] reveal that the (infinite-dimensional) SQP method converges quadratically as long as the precision for solving the linear-quadratic subproblems is compatible with the distance of the current iterate to the exact one. Hence the discretization level has to be increased from step to step. This is the point where the dimension becomes soon astronomical.

Due to this reason, we are not able to report on a sequence of accuracy 10^{-1}, 10^{-2}, 10^{-4}, 10^{-8}, as the reader might expect. We focused our attention to a satisfactory accuracy of the computations and terminated the process at the level of "graphical exactness" (related to a precision of order 10^{-2}). In 2D-domains, this moderate precision leads already to more than $4 \cdot 10^6$ state variables. Our method has to deal with this dimension. Undoubtedly, such high accuracy will not be needed in many practical applications. On the other hand, 3D domains will lead to the same large dimensions for a much lower precision.

For testing the SQP method we consider the optimal boundary control problem to *minimize*

$$\varphi(y,u) = \frac{1}{2}\|y(\cdot,T) - y_T\|^2_{L^2(\Omega)} + \frac{\lambda}{2}\|u\|^2_{L^2(\Sigma)} \tag{1.1}$$

subject to

$$\begin{array}{rcll} y_t(x,t) &=& \Delta_x y(x,t) & \text{in} \quad Q \\ y(x,0) &=& y_I(x) & \text{in} \quad \Omega \\ \partial_\nu y(x,t) &=& b(y(x,t)) + u(x,t) & \text{on} \quad \Sigma \end{array} \tag{1.2}$$

and

$$u_a \le u(x,t) \le u_b \qquad \text{a.e. on } \Sigma. \tag{1.3}$$

We assume that $\Omega \subset I\!R^n$ is a bounded domain with boundary Γ of class $C^{2,\alpha}$, $(0 < \alpha < 1)$ such that Ω is locally at one side of Γ. We put $Q = \Omega \times (0,T)$, $\Sigma = \Gamma \times (0,T)$; $T > 0$, $\lambda > 0$, $u_a \le u_b$ are fixed real numbers, and y_I, $y_T \in C(\overline{\Omega})$ are given functions. By ∂_ν the (outward) normal derivative on Γ is denoted. We assume that $b = b(y)$ belongs to $C^{2,1}(I\!R)$ and is monotone non–increasing. The *control* function $u = u(x,t)$ is looked upon in $L^\infty(\Sigma)$, while the *state* $y = y(x,t)$ is defined as weak solution of (1.2) in $Y = W(0,T) \cap C(\overline{Q})$, where $W(0,T) = \{y \in L^2(0,T;H^1(\Omega)) |\, y_t \in L^2(0,T;H^1(\Omega)')\}$ (cf. Lions and Magenes [24]).

A weak solution y of (1.2) is defined by $y(x,0)=y_0$ and

$$(y_t(t),v)_{(H^1)',H^1} + \int_\Omega \nabla y(t)\nabla v\,dx = \int_\Gamma (b(y(\cdot,t)) + u(\cdot,t))v\,dS \qquad \text{a.e. } t\in(0,T) \tag{1.4}$$

for all $v \in H^1(\Omega)$ (dS: surface measure on Γ).

Let $U_{ad} = \{u \in L^\infty(\Sigma) : u_a \le u(x,t) \le u_b \text{ a. e. on } \Sigma\}$ be the *admissible set.*

2 Necessary and sufficient optimality conditions

First of all we should mention the following result on existence and uniqueness for the state equation (1.2). It is due to Raymond and Zidani [28]:

Theorem 2.1 *For each control $u \in L^\infty(\Sigma)$ the equation (1.2) has a unique solution $y \in C(\bar{Q}) \cap W(0,T)$.*

(cf. [28], Theorem 3.1 and Proposition 3.1).

As an immediate conclusion we obtain by standard methods the existence of at least one optimal control, as u is appearing linearly, φ is convex and continuous and U_{ad} is weakly–star compact. However, we do not focus our method on (globally) optimal controls only. The SQP method will converge in a neighbourhood of any locally optimal control, provided that some natural assumptions are satisfied. To make them precise we state at first a set of standard first and second order optimality conditions.

The *first order necessary optimality conditions* for a pair $(\bar{y},\bar{u})$ consists of the state equation (1.2), the constraint $u \in U_{ad}$, the *adjoint equation*

$$\begin{array}{rcll} -p_t(x,t) &=& \Delta_x p(x,t) & \text{in} \quad Q \\ p(x,T) &=& \bar{y}(x,T) - y_T(x) & \text{in} \quad \Omega \\ \partial_n p(x,t) &=& b'(\bar{y})p(x,t) & \text{on} \quad \Sigma \end{array} \tag{2.1}$$

and of the *variational inequality*

$$\int_\Sigma (\bar{p}(x,t) + \lambda\bar{u}(x,t))(u(x,t) - \bar{u}(x,t))\,dSdt \ge 0. \tag{2.2}$$

Moreover, we shall assume that $(\bar{y},\bar{u})$ satisfies second order sufficient optimality conditions. To formulate them, we introduce for arbitrary small (but fixed) $\sigma > 0$ the set

$$I_\sigma = \{(x,t) \in \Sigma \mid \qquad |\lambda\bar{u}(x,t) + \bar{p}(x,t)| \ge \sigma\} \tag{2.3}$$

of sufficiently strong active inequalities.

To formulate associated second order sufficient optimality conditions, we introduce the *Lagrange function* $\mathcal{L}$,

$$\begin{aligned}\mathcal{L}(y,u,p) &= \phi(y,u) - \int_Q \{y_t\, p + < \nabla y\,,\, \nabla p >\} dxdt \\ &\quad + \int_\Sigma p\,(b(y)+u)\, dSdt.\end{aligned}$$

$\mathcal{L}$ is defined on $Y \times L^\infty(\Sigma) \times Y$. It is twice continuously differentiable w.r. to (y,u) in $Y \times L^\infty(\Sigma)$. Note that this is not true in $W(0,T) \times L^\infty(\Sigma)$. The product $\int_Q y_t\, p\, dxdt$ in the definition of $\mathcal{L}$ is defined in the sense of the pairing between $L^2(0,T;H^1(\Omega)')$ and $L^2(0,T;H^1(\Omega))$. However, this will not be needed in this paper. The second order derivative of $\mathcal{L}$ w.r. to (y,u) is

$$\begin{aligned}\mathcal{L}''(y,u,p)[(y_1,u_1),(y_2,u_2)] &= \int_\Omega y_1(T) y_2(T)\, dx + \\ &\quad + \int_\Sigma (\lambda u_1 u_2 + p\, b''(y) y_1 y_2)\, dSdt.\end{aligned}$$

The *second order sufficient optimality conditions* for $(\bar{y},\bar{u},\bar{p})$ are as follows:

There exist $\delta > 0$ and $\sigma > 0$ such that

$$\text{(SSC)} \quad \left\{ \begin{array}{ll} \mathcal{L}''(\bar{y},\bar{u},\bar{p})[(y,u),(y,u)] \geq \delta \|u\|^2_{L^2(\Sigma)} & (2.4) \\ \text{for all } (y,u) \in W(0,T) \times L^2(\Sigma) \text{ such that } u = 0 \text{ on } I_\sigma \text{ and} & \\ \begin{aligned} y_t &= \Delta_x y \\ y(0) &= 0 \\ \partial_\nu y &= b'(\bar{y}) y + u. \end{aligned} & (2.5) \end{array} \right.$$

Now we assume once and for all that a *reference pair* $(\bar{y},\bar{u})$ is given, which satisfies together with an associated adjoint state $\bar{p}$ the optimality system and the second order sufficient optimality condition.

It can be shown that $\bar{u}$ is under these assumptions *locally optimal* in the sense of $L^\infty(\Sigma)$ (this can be even proved in $L^p(\Sigma)$ for $p > N+1$): For a weaker version of second order sufficient conditions we refer to Goldberg and Tröltzsch [12]. In the case of an elliptic equation of state this is shown in Casas, Tröltzsch and Unger [8]. Their technique can easily be transferred to the parabolic case considered here.

3 The Sequential Quadratic Programming method

In this section we recall the (continuous) SQP method. Let $w_0 = (y_0, p_0, u_0)$ be a starting triplet (we shall assume that w_0 is close to the reference triplet $\bar{w} = (\bar{y},\bar{u},\bar{p})$). Then the Sequential Quadratic Programming (SQP) method determines a sequence $w_k = (y_k, p_k, u_k)$

as follows. Let $\varepsilon > 0$ be given. Initiating from w_k, the next iterate w_{k+1} is obtained from solving the linear–quadratic control problem:

Minimize

$$\varphi'(y_k, u_k)(y - y_k, u - u_k) + \frac{1}{2}\mathcal{L}''(y_k, u_k, p_k)[(y - y_k, u - u_k)]^2 \tag{3.1}$$

subject to

$$\begin{array}{rcll} y_t &=& \Delta y & \text{in} \quad Q \\ y(0) &=& y_I & \text{in} \quad \Omega \\ \partial_\nu y &=& b(y_k) + b'(y_k)(y - y_k) + u & \text{on} \quad \Sigma, \\ u &\in& U_{ad}. & \end{array} \tag{3.2}$$

The solution is (y_{k+1}, u_{k+1}), while the next iterate p_{k+1} of the adjoint state is obtained from the adjoint equation

$$\begin{array}{rcll} -p_t &=& \Delta p & \text{in} \quad Q \\ p(T) &=& y_{k+1}(T) - y_T & \text{in} \quad \Omega \\ \partial_\nu p &=& b'(y_k)p + p_k b''(y_k)(y_{k+1} - y_k) & \text{on} \quad \Sigma. \end{array} \tag{3.3}$$

For convenience we indicate the explicit expressions of φ' and $\mathcal{L}''$:

$$\begin{array}{rcl} \varphi'(y_k, u_k)(y - y_k, u - u_k) &=& \int_\Omega y_k(T)(y(T) - y_k(T))\, dx + \\ && + \int_\Sigma \lambda\, u_k\, (u - u_k)\, dS dt \\ \mathcal{L}''(y_k, u_k, p_k)[(y - y_k, u - u_k)]^2 &=& \int_\Omega (y(T) - y_k(T))^2\, dx + \\ && + \int_\Sigma (\lambda(u - u_k)^2 + p_k\, b''(y_k)(y - y_k)^2)\, dS dt. \end{array}$$

Unfortunately, the linear-quadratic optimal control problem above is not necessarily convex. Our second order sufficient optimality condition imposed on $\bar{w}$ is too weak to guarantee convexity. Therefore, we cannot expect that the SQP method converges locally to $\bar{w}$, unless $\bar{w}$ belongs to a unique global minimum. Note that our method determines global minima of the quadratic sub-problems.

This is the reason to restrict the optimization in (3.1)–(3.2) to a neighborhood U_{ad}^r of the starting element u_o (containing $\bar{(w)}$ in its interior), where

$$U_{ad}^r = \{u \in U_{ad} \mid \|u - u_o\|_{L^\infty} \leq r\}.$$

The following convergence result can be shown, cf. Goldberg and Tröltzsch [13] for the simplified problem discussed here and to Tröltzsch [32] for a detailed analysis in the case of a general class of nonlinear parabolic control problems. Let $B_\rho(\bar{w})$ denote the open ball around $\bar{w}$ in the natural norm of $C(\bar{Q})^2 \times L^\infty(\Sigma)$.

Theorem 3.1 *Let $\rho > 0$ be sufficiently small and $r := 2\rho$. If the search in (3.1)-(3.2) is restricted to U^r_{ad}, then the SQP-method generates for any starting point $w_0 = (y_0, p_0, u_0) \in B_\rho(\bar{w})$ a unique sequence $\{(y_k, p_k, u_k)\}$ such that*

$$\|(y_{k+1}, p_{k+1}, u_{k+1}) - (\bar{y}, \bar{p}, \bar{u})\|_{C(\bar{Q})^2 \times L^\infty(\Sigma)} \le c_\rho \|(y_k, p_k, u_k) - (\bar{y}, \bar{p}, \bar{u})\|^2_{C(\bar{Q})^2 \times L^\infty(\Sigma)}, \quad (3.4)$$

$(k = 0, 1, \dots)$.*, holds with is a certain positive constant c_ρ.*

Remark: In [13], r corresponds to $2/3 varepsilon$, ρ to $\varepsilon/3$. If the second order sufficient optimality condition is required for all $(x,t) \in \Sigma$, then the quadratic sub-problems are convex and the restriction to U^r_{ad} is not necessary. In our test examples, we did not use U^r_{ad}.

A different method of Newton type, presented by Kelley and Sachs [22] for the control of ordinary differential equations, is able to avoid this restriction to a neighborhood. However, the authors have to impose some structural assumptions on the active set and conditions on the slope of the switching function at the junction points.

This convergence result remains true for a very general class of parabolic control problems. We refer to [32]. However, it is more or less only of theoretical value. Our convergence analysis is based on an exact solution of the quadratic subproblems. Without aiming to give a rigorous error analysis for an inexact solution of these problems, we briefly sketch a simple estimate.

Let us denote by $\alpha(h)$ the distance between w^h_n and the exact solution w_n of (QP_{n-1}). It is known that $\alpha(h) = O(h)$ can be expected. Denote by h^+ the mesh size for discretizing (QP_n). An easy estimate shows that quadratic convergence is preserved, if h^+ is defined according to the rule

$$\alpha(h^+) \le c\,\|w^h_n - \bar{w}\|^2_W$$

(cf. [13]). Then

$$\|w^{h^+}_{n+1} - \bar{w}\|_W \le 2\,c\,\|w^h_n - \bar{w}\|^2_W$$

is obtained for all steps of the SQP method with the same constant c. This gives a rule to adapt the precision for the solution of the quadratic subproblems: Roughly speaking, the mesh size has to be proportional to the current accuracy of the SQP step.

Clearly, this strategy can rarely be used in practice, as the size of the problems would grow rapidly. The bottleneck is here the precision to solve the forward-backward system (3.2)–(3.3). The requirement to adapt the accuracy for solving these equations will lead in a few steps to an enormous number of state variables. We shall not try to do this. Our aim is to solve the quadratic subproblems with a quite high precision (whatever this means). In this case, the speed of convergence will mainly be limited by $\alpha(h_n)$ after some steps.

4 Approximation of the quadratic subproblems by a direct method

The following direct method works very well for the solution of the quadratic sub–problems, if the requirement of precision is quite low: Let $h > 0$ denote a certain mesh size characterizing the discretization of the parabolic PDE and the discretization of the control u.

Let the boundary domain Σ be subdivided into m parts Σ_j, $j = 1, \ldots, m$, where $m \in I\!N$. We are looking for the control u as a piecewise constant function

$$u_h = \sum_{j=1}^{m} u_j e_j, \tag{4.1}$$

where $u_j \in I\!R$ and e_j is the characteristic function of Σ_j. Let (y_k, u_k, p_k) be given fixed.

In order to solve the linear state equation (3.2) for a given $u = u_h$ we have to solve it first for $y_I = 0$ and every basis function e_j. Let y_j denote the state function associated to the basis control $u = e_j$ and to the initial state $y_I = 0$. Then the state y_h corresponding to u_h is given by the superposition principle as a linear combination of basis states y_j and the solution of (3.2) for $u = 0$.

Denote by $\underline{u}$ the vector $(u_1, \ldots, u_m)^T$. After discretizing the objective functional (3.1) we arrive at a quadratic programming problem for $\underline{u}$:

$$\begin{aligned} \underline{u}^T C \underline{u} + l^T \underline{u} &= \min! \qquad (4.2)\\ u_a \le u_i &\le u_b, \quad i = 1, \ldots, m, \qquad (4.3) \end{aligned}$$

where $C \in I\!R^m \times I\!R^m$, $l \in I\!R^m$. If the resulting dimension m is not too large (say some hundred), this problem can be solved efficiently by standard software packages. For instance we made good experiences with E04NAF (NAG library) developed by Gill and Murray [11]. We refer also to Alt and Mackenroth [4] or Mackenroth [26], who reported on the same positive experience with this technique. For large m, the storage capacity of the computer may soon be exceeded, as C is very large. Moreover, C has to be computed (e.g. the state equation has to be solved for all basic functions e_j, the occuring basic states y_j have to be inserted into the objective functional and the coefficients corresponding to the quadratic parts of $\bar{u}$ are finally the entries of C) and stored in each SQP–step. For that reason, we have decided to choose another approach to solve the quadratic subproblems.

5 A multigrid approach

The essential difficulty for solving the linear-quadratic subproblems is not connected with the presence of the control-constraint $u \in U_{ad}$. It appears also in the unconstrained case, where a large-scale backward-forward system of two coupled parabolic equations has to be solved. A way to solve unconstrained optimal control problems was presented by Hackbusch [14]. Let us give a brief sketch of this idea. We consider for simplicity the unconstrained quadratic optimal control problem to *minimize*

$$\frac{1}{2}\|y(T,\cdot)-y_T\|^2_{L^2(\Omega)}+\frac{\lambda}{2}\|u\|^2_{L^2(\Sigma)} \tag{5.1}$$

subject to

$$\begin{array}{rcll} y_t(x,t) & = & \Delta y(x,t) & \text{in} \quad Q \\ y(0,x) & = & y_I(x) & \text{in} \quad \Omega \\ \partial_n y(x,t) & = & u(x,t)-y(x,t) & \text{on} \quad \Sigma. \end{array} \tag{5.2}$$

Suppose that $\bar{u}$ is the optimal solution of problem (5.1)–(5.2). Then the optimal triplet $(\bar{u},\bar{y},\bar{p})$ must fulfil the state equation (5.2) and the first order necessary optimality conditions including the adjoint equation

$$\begin{array}{rcll} -p_t(x,t) & = & \Delta p(x,t) & \text{in} \quad Q \\ p(x,T) & = & y(x,T)-y_T(x) & \text{in} \quad \Omega \\ \partial_n p(x,t) & = & -p(x,t) & \text{on} \quad \Sigma \end{array} \tag{5.3}$$

and

$$u(x,t)=-\frac{p(x,t)}{\lambda} \quad \text{on} \quad \Sigma. \tag{5.4}$$

Introduce now an operator T mapping the control space $U=L^2(\Sigma)$ into itself by

$$(Tu)(x,t)=-\frac{p(x,t)}{\lambda}.$$

Please note that the chain $u\mapsto y\mapsto p\mapsto Tu$ defined by (5.2), (5.3), and (5.4) is behind this construction.

The operator T is well defined because the state and adjoint equations have unique weak solutions. An optimal solution $\bar{u}$ has to be a fixed point of T.

$$\bar{u}=T\bar{u}. \tag{5.5}$$

It is obvious that T is affine linear

$$Tu=Ku+f,$$

with a compact operator K in $L^2(\Sigma)$ and a fixed $f\in L^2(\Sigma)$. K is of *Fredholm type* with nonnegative kernel. This can be illustrated most easily with the Green's function

$G = G(x,\xi,t)$ for (5.2): Then $G \geq 0$ and

$$
\begin{aligned}
y(x,t) &= \int_\Omega G(x,\xi,t) y_I(\xi)\, d\xi + \int_0^t \int_\Gamma G(x,\xi,t-s) u(\xi,s)\, dS_\xi ds \\
p(x,t) &= \int_\Omega G(x,\xi,T-t)(y(\xi,T) - y_T(\xi))\, d\xi \\
&= \int_0^T \int_\Omega \int_\Gamma G(x,\xi,T-t) G(\xi,\eta,T-s) u(\eta,s)\, dS_\eta d\xi ds \\
&+ \left\{ \int_\Omega \int_\Omega G(x,\xi,T-t) G(\xi,\eta,T) y_I(\eta)\, d\eta d\xi - \int_\Omega G(x,\xi,T-t) y_T(\xi)\, d\xi \right\} \\
&=: -\lambda(Ku + f).
\end{aligned}
\tag{5.6}
$$

This property can be shown also in the framework of weak solutions without relying on Green's functions. However, the representation (5.6) shows best the nature of K as integral operator. It stands behind the multigrid strategy to determine $u(x,t)$. Although this cannot be realized in real computations we tacitly assume that $G(x,\xi,t)$ is exactly known. In other words, we assume to solve the PDE's (5.2)–(5.3) exactly. The multigrid strategy refers to a discretization of u and to an associated collocation method applied to the equation $u = Ku + f$. This means for a fixed grid that Σ is subdivided into $\bar\Sigma = \bar\Sigma_1 \cup \bar\Sigma_2 \cup \ldots \cup \bar\Sigma_m$, u_h is taken constant u_j on Σ_j, and $y(x,t)$, $p(x,t)$ are evaluated at prescribed points $(t_j, x_j) \in \Sigma_j$, $j = 1,\ldots,m$ (for instance, at certain "midpoints" of Σ_j) by solving their PDE's exactly (in practice this means solving the PDE by a sufficiently high precision).

The main steps of such a multigrid algorithm, described here for two grids, are well known. We refer, for instance, to Hackbusch [15].

Let $\bar\Sigma = \bar\Sigma_1^h \cup \ldots \cup \bar\Sigma_{m(h)}^h$ denote the partition of Σ associated to the mesh size paramter h. Then one multigrid–step (**MG**) is performed as follows.

1. Let an iterate u_h^k be given on the fine grid associated to parameter h .
2. Determine the residual $r_h = u_h^k - Tu_h^k$.
3. Reduce the residual to the coarse grid associated to l: $r_l = Rr_h$.
4. Compute a correction by solving $(I - K)d_l = r_l$, where I is the identity.
5. Prolongate the correction to the fine grid, $d_h = Pd_l$.
6. Determine the new iterate $u_h^{k+1} = u_h^k + d_h$.
7. If $\|u_h^k - u_h^{k+1}\|$ is not small enough, then set $u_h^k := u_h^{k+1}$, $k := k+1$, go to 2. Otherwise stop the algorithm.

Here, P and R are some prolongation and restriction operators respectively. $\|u_h^k - u_h^{k+1}\|$ is considered in an appropriate norm, l denotes the mesh size of the coarse grid.

At step 2 we have to apply the operator T. In theory, we get Tu by (exactly) solving the parabolic equation for y, inserting y into the adjoint equation, determining p and using finally the representation (5.4). In real computations, the parabolic equations have to be somehow discretized. Any sufficiently accurate solver can be used for this purpose. We do not consider the difficult error analysis connected with the approximate solution of the partial differential equations. In our further presentation we continue to regard the mapping $u \mapsto Tu$ as given exactly.

The bottleneck of the multigrid method is hidden in establishing the coarse–grid system

$$(I - K)d_l = r_l$$

or its discretized version

$$(I - K_l)\underline{d}_l = \underline{r}_l$$

in step 4. The vectors $\underline{d}_l$, $\underline{r}_l$ representing d_l, r_l belong to the finite–dimensional space $\mathbb{R}^{m_l}$. We need the associated matrix representation K_l of K. Let $\{e_1, \ldots, e_{m_l}\}$ be a basis for this space. This can be done by m_l–times applying the operator T

$$Ke_i = Te_i - f, \qquad i = 1, \ldots, m_l$$

on the coarse grid. This is computationally expensive, as the coarse grid system has to be solved very often during an SQP method.

An iterative method for solving systems of linear equations does not require the computation of K_l. Here, we need only $K_l\underline{u}_l$ for certain vectors $\underline{u}_l$ occuring in the iteration process. If the iteration stops after a number of steps less than the dimenson of $\underline{u}_l$, then the effort occuring in step 4 of the multigrid algorithm decreases.

Another promising approach for handling the optimality system (5.2)–(5.3) is to solve directly the occuring coupled forward–backward system. However, this leads to a system of much higher dimension than in the multigrid method proposed before. Moreover, an effective numerical technique for parabolic backward–forward systems has still to be developed. Therefore, we decided to use the multigrid approach, where the state and adjoint state equations are decoupled.

Remark: In principle, the multigrid strategy is able to handle any mesh-size, which is useful to solve the partial differential equations on the available computer. In this way, our aim of high-precision computations can be achieved. However, the method is quite slow.

6 Control constraints

If the pointwise constraints $u \in U_{ad}$ are not imposed on u in (3.2), then the SQP method is nothing else than the known Newton method for solving the equations of the optimality system (consisting of (1.2), (2.1), and $\bar{u} = -\lambda^{-1}\bar{p}$). One step of the method can be performed by the technique described in the preceding section.

Let us now take into account the restrictions (1.3) on the control u. Once again we explain the technique for the simplified linear–quadratic problem (5.1)–(5.2), now with the additional constraint (1.3). We use an active set strategy due to Bertsekas [7] which was already succesfully applied by other authors, for instance by Heinkenschloss and Sachs [18]. Let us first formulate this algorithm (**B**) and then discuss its steps.

1. Denote by $\underline{u}^k = (u_1^k, \ldots, u_m^k)^T$ be the vector representing the iterate u_h^k (h fixed), fix positive numbers ε and σ and let $I = \{1, \ldots, m\}$ be the index set associated to u_h^k.
2. Solve (5.2), (5.3) and get p_h^k with the same discretization as u_h^k. Let $\underline{p}^k = (p_1^k, \ldots, p_m^k)^T$ be the representation of p_h^k.
3. Define the sets of strongly active inequalities by $I_a^\varepsilon = \{i \in I : u_i^k = u_a \text{ and } \lambda u_i^k + p_i^k > \varepsilon\}$ and $I_b^\varepsilon = \{i \in I : u_i^k = u_b \text{ and } \lambda u_i^k + p_i^k < -\varepsilon\}$.
4. Set $\hat{u}_i = u_i^k$ for all $i \in I_a^\varepsilon \cup I_b^\varepsilon$.
5. Solve the unconstrained problem (5.1)–(5.2) for u_i^k, $i \in I \setminus (I_a^\varepsilon \cup I_b^\varepsilon)$, while the remaining components of $\underline{u}^k$ are fixed due to 4, denote the solution by v_h^k with vector representation $\underline{v}^k$.
6. Set $u_h^{k+1} = \Pi v_h^k$, where Π is the projection onto $[u_a, u_b]^m$.
7. If $\|u_h^{k+1} - u_h^k\| \geq \sigma$ then put $u_h^k := u_h^{k+1}$, $k := k + 1$ and go to 2. Otherwise stop the iteration.

To illustrate the idea of this technique we consider the first order necessary optimality conditions for problem (5.1), (5.2), (1.3). The optimal triplet $(\bar{u}, \bar{y}, \bar{p})$ has to satisfy the adjoint equation (5.3) together with the variational inequality

$$\int_\Sigma (\bar{u} + \lambda\bar{p})(u - \bar{u})\, dSdt \geq 0 \quad \forall u \in U_{ad}. \tag{6.1}$$

A standard discussion of this inequality shows that

$$\bar{u} = \begin{cases} u_a & , \text{ if } \bar{u} + \lambda\bar{p} > 0 \\ u_b & , \text{ if } \bar{u} + \lambda\bar{p} < 0 \\ -\dfrac{\bar{p}}{\lambda} & , \text{ if } \bar{u} + \lambda\bar{p} = 0 \end{cases} \tag{6.2}$$

(see, for instance, [9]) . These three possible cases for $\bar{u}$ are reflected by step 3 and 4 of the algorithm. If $|u_i^k + \lambda p_i^k| > \varepsilon$, then we can expect that this index i belongs to an active inequality. Therefore, we keep this value u_i^k fixed at the boundary in the next step.

7 Numerical tests

7.1 The one–dimensional case

We have reported on our 1D–computational experience for $\Omega = (0,1)$ in the paper [13]. Let us recall the results for comparison. In our test examples, the control $u = u(t)$ is acting on the right end of $\Omega = (0,1)$. $\Sigma = \{0,1\} \times (0,T)$ splits into 2 parts and $b = 0$, $u = 0$ is kept fixed on the left part $x = 0$. However, we do not need this formal expression of the setting to make the problem comparable to the general problem of section 2. We just formulate the state equation as

$$\begin{array}{rcll} y_t(x,t) & = & y_{xx}(x,t) & \text{in} \quad Q \\ y(0,x) & = & y_I(x) & \text{in} \quad \Omega \\ y_x(t,0) & = & 0 & \text{on} \quad (0,T] \\ y_x(t,1) & = & b(y(t,1)) + u(t) & \text{on} \quad (0,T]. \end{array} \tag{7.1}$$

Let the interval $[0,T]$ be divided by the equidistant grid

$$0 \;=\; t_0 < t_1 < \ldots < t_{n_t} = T,$$

where n_t is a given integer. Thus, the subdomains Σ_k are given here by the intervals (t_{k-1}, t_k), $k = 1, \ldots, n_t$. The discretization of u is performed as described in (4.1). We considered the following test example (going back to Schittkowski [29]).

Example 1 This example is a linear–quadratic control problem of the type discussed in section 6. It is included here to stress that a very fine discretization of the control u and the PDE is needed to compute a sufficiently precise optimal control. "Sufficiently precise" means in this test example that a further refinement of the underlying grid did not change the graphical plot of the control.

We took $T = 1.58$, $\lambda = 0.001$, $y_T = 0.5(1-x^2)$, $y_I \equiv 0$, $u_a = -1$, $u_b = 1$ and $b(y) = -y$. The state and adjoint equations are solved on $Q = (0,T) \times (0,1)$ by a Crank–Nicholson type finite difference method. Denote by n_t the discretization parameter of u and by n_{y_t}, n_{y_x} the parameters of y (i.e., the discretization with respect to time and space used in the PDE). We determined optimal controls for the following triplets of (n_t, n_{y_t}, n_{y_x}): $(50,100,100)$, $(200,400,400)$, $(400,800,800)$. The results showed that the mesh-size $(400,800,800)$ was necessary to obtain graphically exact controls. A further refinement did not change the computer plot of the optimal control.

5 iterations (**B**) were needed to get the result for the finest discretization. These steps required 56, 34, 30, 14 and 1 multigrid iterations (**MG**), respectively.

Example 2

This is a nonlinear problem with almost the same data as above, but with the nonlinear boundary condition

$$y_x(t,1) = u(t) - y(t,1)^2$$

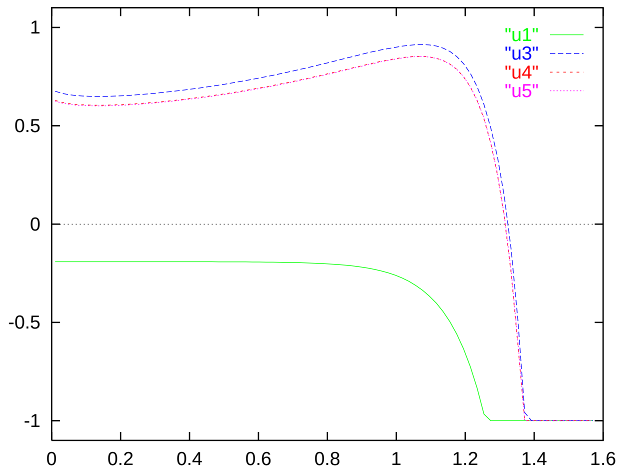

Figure 1: SQP iteration for the nonlinear problem

solved by the SQP method. The iteration was started at $(y_0, u_0, p_0) = (0, 0, 0)$, as mesh size we took $(n_t, n_{y_t}, n_{y_x}) = (400, 800, 800)$, still quite moderate in view of the experience with example 1. We avoid any table of numbers for the progress of computation. It would pretend a high accuracy, which cannot be backed by the precision of the discretization. Once the SQP method is in the region of quadratic convergence, after a few iterations the reached precision is not compatible with the solution of the PDE. Please note that the mesh size 400 for the control will at most ensure a precision of the order 10^{-2} for the linear–quadratic sub–problems. This is just "graphical precision". In Figure 1 some iterates are represented.

7.2 The two–dimensional case

Next, we consider our problem (5.1)–(5.2) in $\Omega = (0, 1) \times (0, 1) \subset I\!R^2$. The control is acting on $\Gamma_1 = \{(x_1, x_2) \in \bar{\Omega} : x_2 = 1\}$, define $\Gamma_2 = \Gamma \setminus \Gamma_1$. The boundary condition is slightly changed,

$$\begin{aligned} \partial_\nu y &= b(y) + u \quad \text{on} \quad \Gamma_1 \\ \partial_\nu y &= -y \quad \text{on} \quad \Gamma_2. \end{aligned} \tag{7.2}$$

As before, the invervals $[0, T]$ and $[0, 1]$ are splitted into equidistant subintervals,

$$\begin{aligned} 0 &= t_0 < t_1 < \ldots < t_{n_t} = T \\ 0 &= x_0 < x_1 < \ldots < x_{n_x} = 1, \end{aligned}$$

where n_t and n_x are positive integers.

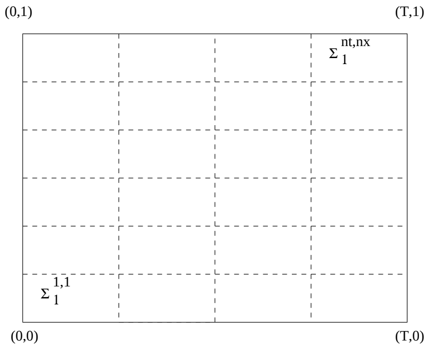

Figure 2: Domain of $\boldsymbol{u_h}$

We split the control domain $\Sigma_1 = \Gamma_1 \times (0,T)$ for the control u into subdomains $\Sigma_1^{ki} = (t_{k-1}, t_k) \times (x_{i-1}, x_i)$, $k = 0, \ldots, n_t$, $i = 0, \ldots, n_x$. The partition of Σ is shown in Figure 2.

To solve the underlying parabolic differential equations, the domain Ω was divided into equidistant subdomains $\Omega_{ij} = \{(x_{i-1}^1, x_i^1) \times (x_{j-1}^2, x_j^2) : i = 0, \ldots, n_{x^1}, j = 0, \ldots, n_{x^2}\}$, where n_{x^i} is the number of equidistant subintervals of $[0,1]$ in direction x_i, $i = 1,2$. The time interval $[0,T]$ is splitted into n_τ equidistant subparts. We put $h_\tau = \frac{T}{n_\tau}$, $h_{x^1} = \frac{1}{n_{x^1}}$ and $h_{x^2} = \frac{1}{n_{x^2}}$.

Remark 7.1 *The number of subintervals of* $[0,1]$ *in* x_i*–direction was related to the number of subintervals of* $[0,1]$ *in* x*–direction for* u_h*:* $n_{x^i} = Mn_x$*,* $i = 1,2$*, where* $M \in \mathbb{N}$*. Analogously,* $n_\tau = Mn_t$*. We used* $M = 2$ *in the computations.*

Owing to the simple geometry of Ω, a finite difference splitting-up method was selected, since it was faster than available finite element codes.

The examples below are computed with the Bertsekas active set strategy and the multigrid method for the unconstrained quadratic subproblems. In the multigrid algorithm, the coarse–grid systems of linear equations were solved by a *conjugate gradient* method.

Example 3

This is a convex linear–quadratic control problem, used to compare the precision of our computations with known results. We take $T = 1$, $\lambda = 0.001$, $y_I \equiv 0$, $y_T = 0.5x_1x_2 + 0.25$, $b(y) = -y$, $u_a = 0$, $u_b = 1$. This problem was considered by Mackenroth [26] with $\lambda = 0$. Therefore, the result was a control of bang–bang type. Our control is close to that of Mackenroth, but continuous, as $\lambda > 0$. Further parameters are $n_t = 10$, $n_x = 10$, $M = 2$ and $n_{x^2} = n_{x^1}$, thus we have

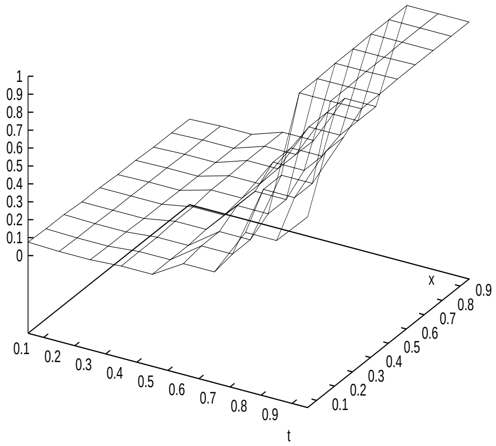

Figure 3: Optimal control of example 3

mesh size (n_x, n_t) for the control on the coarsest grid:	$(5, 5)$,
mesh size (n_x, n_t) on the finest grid:	$(10, 10)$,
mesh size (n_{x^1}, n_{x^2}, n_t) for the state:	$(20, 20, 20)$,
maximal number of state variables:	8000.

The result is shown in Figure 3. The approach considered here allows to solve the problem with a finer discretization.

Example 4

We regard the same problem as above with fine discretization: $T = 1$, $\lambda = 0.001$, $y_I \equiv 0$, $y_T = 0.5x_1x_2 + 0.25$, $b(y) = -y$, $u_a = 0$, $u_b = 1$, $n_t = 80$, $n_x = 80$, $M = 2$ and $n_{x^2} = n_{x^1} = 160$. Thus we have

mesh size (n_x, n_t) for the control on the coarsest grid:	$(5, 5)$,
mesh size (n_x, n_t) on the finest grid:	$(80, 80)$,
mesh size (n_{x^1}, n_{x^2}, n_t) for the state:	$(160, 160, 160)$,
maximal number of state variables:	4.096.000.

7 iterations (**B**) were needed to get the result for the finest discretization. These iterations required 125, 67, 45, 32, 23, 13, and 1 multigrid steps (**MG**), respectively. The result is shown in Figure 4.

Finally, we discuss a nonlinear test example solved by the SQP–method.

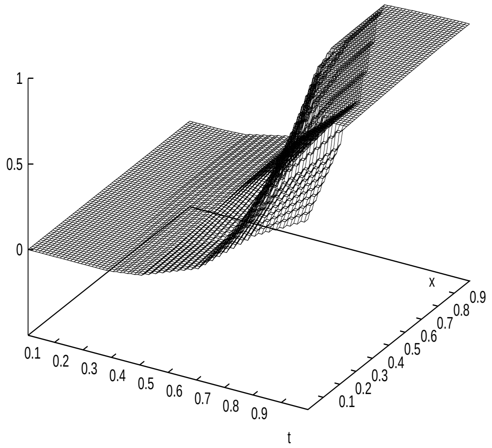

Figure 4: Optimal control of example 4

Example 5

Here, the boundary condition is replaced by a non–linear one. We put $T = 1$, $\lambda = 0.001$, $y_I \equiv 0$, $y_T = 0.5x_1x_2 + 0.25$, $b(y) = -y^2$, $u_a = 0$, $u_b = 0.2$, $n_t = 80$, $n_x = 80$, $M = 2$, and $n_{x^2} = n_{x^1} = 160$. The initial iterate of the SQP algorithm was the triplet $(0, 0, 0)$.

The progress of iteration is shown in Figures 5–9. Our initial iterate was outside the convergence region of the SQP method. This region was hit by chance in the third step by u_3. A usual globalization technique would avoid this behaviour.

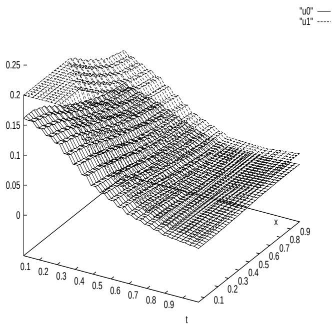

Figure 5: Example 5, $\boldsymbol{u_0}$ and $\boldsymbol{u_1}$

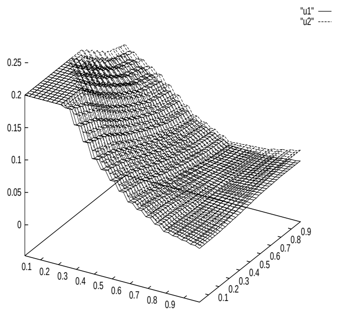

Figure 6: Example 5, $\boldsymbol{u_1}$ and $\boldsymbol{u_2}$

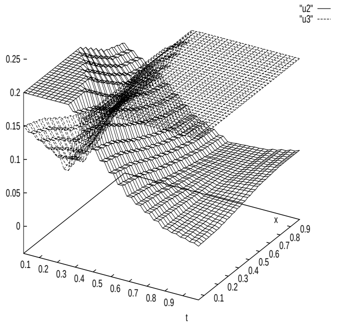

Figure 7: Example 5, $\boldsymbol{u_2}$ and $\boldsymbol{u_3}$

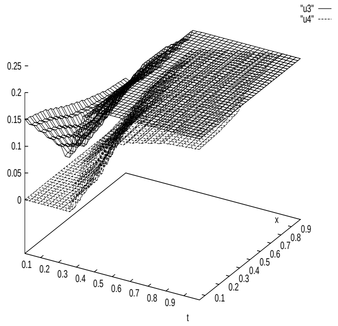

Figure 8: Example 5, $\boldsymbol{u_3}$ and $\boldsymbol{u_4}$

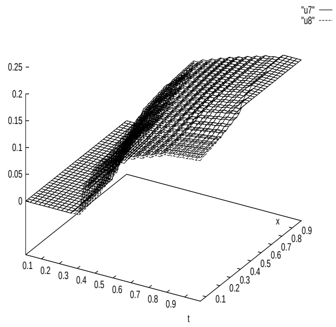

Figure 9: Example 5, $\boldsymbol{u_7}$ and $\boldsymbol{u_8}$

8 Final Comment

The method presented in this paper is not yet effective. SQP–method, Bertsekas active set strategy and multigrid technique form a chain of 3 nested iteration schemes. Although the SQP method itself exhibits the expected fast convergence, the other inherent iteration procedures are slower and lead to long running times. Certainly, this long time is mainly connected with the high precision of the computation. The same effect was mentioned by Gill and others [10] who applied the software system DASOPT to solving a parabolic optimal control problem for a quasilinear heat equation in a two-dimensional rectangular domain Ω with a moderate discretization. We confirmed their estimate $(n_{x_1} \cdot n_{x_2})^2$ for the order of the computational time in our own tests.

Nevertheless, it is obvious that our procedure is not optimal and can be improved in several ways. This was not our primary intention. We aimed to complete the theory of the standard (continuous) SQP method for parabolic control problems by associated numerical test examples. The method presented here is very close to the one discussed in our convergence analysis. Owing to its fast convergence, there was a need to solve the quadratic subproblems with comparably high precision. This paper shows one way leading to acceptable results.

References

[1] W. Alt. The Lagrange–Newton method for infinite–dimensional optimization problems. *Numer. Funct. Anal. and Optimiz.*, 11:201–224, 1990.

[2] W. Alt. The Lagrange Newton method for infinite-dimensional optimization problems. *Control and Cybernetics*, 23:87–106, 1994.

[3] W. Alt. Discretization and mesh independence of Newton's method for generalized equations. To appear.

[4] W. Alt and U. Mackenroth. On the numerical solution of state-constrained coercive parabolic optimal control problems. In *Optimal Control of Partial Differential Equations*, volume 68 of *Lecture Notes ISNM*, pages 44–62, Birkhäuser Verlag, 1992. Hoffmann, K.H. and Krabs, W (eds.).

[5] W. Alt and K. Malanowski. The Lagrange–Newton method for nonlinear optimal control problems. *Comp. Opt. and Applic.*, 2:77–100, 1993.

[6] W. Alt, R. Sontag, and F. Tröltzsch. An SQP Method for Optimal Control of a Weakly Singular Hammerstein Integral Equation. *Appl. Math. Opt.*, 33:227–252, 1996.

[7] D. B. Bertsekas. Projected Newton methods for optimization problems with simple constraints. *SIAM J. Control and Optimization*, 20:221–246, 1982.

[8] E. Casas, F. Tröltzsch and A. Unger. Second order sufficient optimality conditions for a nonlinear elliptic control problem. *J. of Analysis and Applications (ZAA)*, 15:687 - 707, 1996.

[9] K. Eppler and F. Tröltzsch. On switching points of optimal controls for coercive parabolic boundary control problems. *Optimization*, 17:93–101, 1986.

[10] P. E. Gill, K. Park, L. Petzold, J. B. Rosen, L. O. Jay. Numerical optimal control of parabolic PDEs using DASOPT. Report NA 96-1, Dep. of Math., University of California, San Diego.

[11] P. E. Gill, W. Murray, M. A. Saunders, and M. H. Wright. Users's guide for SOL/QPSOL: A Fortran package for quadratic programming. Technical Report No. 82–7, Department of Operations Research, Stanford University, Stanford, CA 94305, 1982.

[12] H. Goldberg and F. Tröltzsch. Second order sufficient optimality conditions for a class of nonlinear parabolic boundary control problems. *SIAM J. Control and Optimization*, 31, 1007–1025, 1993.

[13] H. Goldberg and F. Tröltzsch. On a Lagrange–Newton method for a nonlinear parabolic boundary control problem. Preprint TU-Chemnitz, Fak. f. Mathematik, Report 96–8, to appear.

[14] W. Hackbusch. On the fast solving of parabolic boundary control problems. *SIAM J. Control and Optimization*, 17, No. 2:231–244, 1979.

[15] W. Hackbusch. *Integralgleichungen.* B. G. Teubner, Stuttgart, 1989.

[16] W. Hackbusch and Th. Will. A numerical method for a parabolic bang–bang problem. *Control and Cybernetics*, 12, No. 3–4:100–116, 1983.

[17] M. Heinkenschloss. The numerical solution of a control problem governed by a phase field model. *Optimization Methods and Software*, 7, 1997. To appear.

[18] M. Heinkenschloss and E. W. Sachs. Numerical solution of a constrained control problem for a phase field model. *Control and Estimation of Distributed Parameter Systems, Int. Ser. Num. Math.*, 118:171–188, 1994.

[19] M. Heinkenschloss and F. Tröltzsch. Analysis of the Lagrange-SQP-Newton method for the control of a Phase field equation. *Virginia Polytechnic Institute and State Unversity*, ICAM Report 95-03-01.

[20] C.T. Kelley and E.W. Sachs. Fast algorithms for compact fixed point problems with inexact function evaluations. *SIAM J. Scientific and Stat. Computing* 12:725–742, 1991.

[21] C.T. Kelley and E.W. Sachs. Multilevel algorithms for constrained compact fixed point problems. *SIAM J. Scientific and Stat. Computing* 15:645–667, 1994.

[22] C.T. Kelley and E.W. Sachs. Solution of optimal control problems by a pointwise projected Newton method. *SIAM J. Contr. Optimization* 33:1731–1757, 1995.

[23] S.F. Kupfer and E.W. Sachs. Numerical solution of a nonlinear parabolic control problem by a reduced SQP method. *Computational Optimization and Applications* 1:113–135, 1992.

[24] J.L. Lions and E. Magenes. *Problèmes aux limites non homogènes et applications*, volume 1–3. Dunod, Paris, 1968.

[25] K. Machielsen. Numerical solution of optimal control problems with state constraints by sequential quadratic programming in function space. CWI Tract, 53, Amsterdam, 1987.

[26] U. Mackenroth. Numerical solution of some parbolic boundary control problems by finite elements. In *Control Problems for Systems Described by Partial Differential Equations and Applications*, volume 97 of *Lecture Notes Contr. Inf. Sci.*, pages 325–335, Berlin, 1987. Springer–Verlag.

[27] H. Maurer. First and second order sufficient optimality conditions in mathematical programming and optimal control. *Math.Programming Study*, 14:163–177, 1981.

[28] J.P. Raymond and H. Zidani. Hamiltonian Pontryagin's principles for control problems governed by semilinear parabolic equations. Preprint 1995, to appear.

[29] K. Schittkowski. Numerical solution of a time–optimal parabolic boundary-value control problem. *JOTA*, 27:271–290, 1979.

[30] F. Tröltzsch. Convergence of an SQP–Method for a class of nonlinear parabolic boundary control problems. In W. Desch, F. Kappel, K. Kunisch, eds., Control and Estimation of Distributed Parameter Systems. Nonlinear Phenomena. Int. Series of Num. Mathematics, Vol. 118, Birkhäuser Verlag, Basel 1994, pp. 343–358.

[31] F. Tröltzsch. An SQP method for the optimal control of a nonlinear heat equation. *Control and Cybernetics*, 23(1/2):267–288, 1994.

[32] F. Tröltzsch. On the Lagrange–Newton method for the optimal control of semilinear parabolic equations, 1997, to appear.