ON NUMERICAL PROBLEMS CAUSED BY DISCONTINUITIES IN CONTROLS

Christian Großmann, Antje Noack, Reiner Vanselow *
Dresden University of Technology
Institute of Numerical Mathematics
D - 01062 Dresden, Germany
{grossm, noack, vanselow}@math.tu-dresden.de

Abstract The regularity of solutions of parabolic initial-boundary value problems directly depends upon the regularity of the boundary data. Reduced regularity of boundary data arise e.g. in optimal boundary control problems governed by evolution equations by a discretization of the control by piecewise constant functions and results in refined grids if automatic step size procedures in time are applied. In the present study effects to numerical methods for solving the state equations are illustrated. Moreover, an appropriate splitting of the solution is used to improve the numerical behavior of the discretization technique as well as of the optimization method applied to the control problem itself.

Keywords: Boundary control, parabolic equation, discretization.

1. Introduction

Smoothness properties of solutions of parabolic initial-boundary value problems directly depend upon the smoothness of initial and boundary data. As a consequence, discretizing the boundary control by piecewise given functions generically results in a reduced smoothness of solutions of the related state equations. However, the efficiency of numerical methods for partial differential equations depends on the regularity of the desired solution. This yields specific effects like severe local grid refinements in time when standard discretization techniques are applied. In the present paper we investigate such effects and in case of piecewise constant Dirichlet controls we use a splitting of the solution to improve

*Partial funding of this research provided by DFG grant GR1777/2-2.

The original version of this chapter was revised: The copyright line was incorrect. This has been corrected. The Erratum to this chapter is available at DOI: 10.1007/978-0-387-35699-0_19
E. W. Sachs et al. (eds.), *System Modeling and Optimization XX*

the numerical behavior of the discretization technique as well as of the optimization method applied to the control problem.

Throughout the paper we consider spatial one-dimensional boundary heat control problems

$$J(u) := \frac{1}{2}\int_0^1 [w(x,T;u) - q(x)]^2\,dx + \frac{\alpha}{2}\int_0^T [u(t) - p(t)]^2\,dt \to \min_{u\in U}\,! \quad (1.1)$$

subject to the state equations

$$\begin{aligned} \frac{\partial w}{\partial t} - \sigma^2\,\frac{\partial^2 w}{\partial x^2} &= f \quad \text{in} \quad Q := (0,1)\times(0,T], \\ w &= 0 \quad \text{on} \quad \Gamma_l := \{0\}\times(0,T], \\ \gamma_D\, w + \gamma_N\,\frac{\partial w}{\partial x} &= u \quad \text{on} \quad \Gamma_r := \{1\}\times(0,T], \\ w &= g \quad \text{on} \quad Q_0 := [0,1]\times\{0\}, \end{aligned} \quad (1.2)$$

with control u and the state $w(\cdot,\cdot;u)$ in the weak sense (cf. [11], [12]). Here γ_D, $\gamma_N \geq 0$ are given coefficients satisfying $\gamma_D + \gamma_N > 0$ and $\alpha > 0$ is a fixed regularization parameter. Further, $U \neq \emptyset$, $U \subset L_\infty(0,T)$ denotes a set of admissible controls, $q \in L_2(0,1)$ is the given target temperature and $p \in U$ denotes some fixed reference control.

For controls $u \in U$ here we restrict ourselves to discretizations $u^\tau \in U^\tau$ defined over a given time grid

$$0 = t^0 < t^1 < \cdots < t^{M^c-1} < t^{M^c} = T \quad (1.3)$$

by piecewise constant functions, i.e.

$$u^\tau \in U^\tau \quad \Longleftrightarrow \quad u^\tau(t) = u^k \in \mathbb{R},\ \forall t \in (t^{k-1}, t^k],\ k = 1,\ldots,M^c. \quad (1.4)$$

Here M^C denotes the number of time intervals for control discretization. To distinguish between discretizations of control and states we indicate the first ones by upper scripts (as above) and the second ones by lower scripts.

In case of Dirichlet controls, i.e. $\gamma_D = 1$, $\gamma_N = 0$, jumps of u^τ at inner grid points t^k, $k = 1,\ldots,M^c - 1$, cause discontinuities of the related solution $w(\cdot,\cdot;u^\tau)$. In literature, there are several results on the numerical treatment of the heat equation with irregular solutions, where the irregularities result from the functions f or g (cf. [2], [6], [9], [10]).

In literature one can find three approaches to overcome such difficulties. In the first one fitted methods are constructed with coefficients which are adapted to the singularities (cf.[6]). In the second approach a standard method is chosen but with specifically refined meshes in the

neighborhood of singularities (cf. [2] , [9], [10]). The third approach splits off the singularities. In the present paper we apply this splitting. Related details are discussed in the following section.

2. Numerical treatment of the state equations

2.1 Splitting in case of Dirichlet conditions

In the present subsection we consider the case $\gamma_D = 1$, $\gamma_N = 0$. At first, we assume compatibility at $x = 0$, i.e. $g(0) = 0$, and (for the sake of a uniform description) we extend the piecewise constant function u^τ to $t = 0$ by $u^\tau(0) := u^0 := g(1)$.

Let us now introduce functions $w^k : Q \to \mathbb{R}$, $k = 1, \ldots, M^c$, by

$$w^k(x,t) := \begin{cases} 1 - \operatorname{erf}\left(\dfrac{1-x}{2\sigma\sqrt{t-t^{k-1}}}\right) & , \text{ if } x \in [0,1],\ t > t^{k-1} \\ 0 & , \text{ if } x \in [0,1],\ t \le t^{k-1} \end{cases} \tag{2.1}$$

with the error function

$$\operatorname{erf}(\xi) := \frac{2}{\sqrt{\pi}} \int_0^\xi e^{-s^2} ds, \quad \xi \in \mathbb{R}. \tag{2.2}$$

Definition (2.1), (2.2) yields $w^k \in C^\infty(\bar{Q} \backslash \{(1, t^{k-1})\})$ and

$$\frac{\partial w^k}{\partial t} - \sigma^2 \frac{\partial^2 w^k}{\partial x^2} = 0 \qquad \text{in } Q.$$

Further, w^k has a jump w.r.t. t at $(1, t^{k-1})$. Hence, occurring discontinuities of the solution of (1.2) at the points $(1, t^k)$, $k = 0, \ldots, M^c - 1$, originated by jumps in u, can be captured by the functions w^k. Namely, using superposition, the solution $w(\cdot, \cdot\,; u^\tau)$ of (1.2) can be written as

$$w(x,t;u^\tau) = \hat{w}(x,t;u^\tau) + v(x,t;u^\tau), \qquad (x,t) \in \bar{Q} \tag{2.3}$$

for any given $u^\tau \in U^\tau$, where $\hat{w}(\cdot, \cdot\,; u^\tau)$ is defined by

$$\hat{w}(x,t;u^\tau) := \sum_{k=1}^{M^c} \left(u^k - u^{k-1}\right) w^k(x,t), \qquad (x,t) \in \bar{Q} \tag{2.4}$$

and $v(\cdot, \cdot\,; u)$ denotes the solution of the related parabolic problem

$$\begin{aligned} \frac{\partial v}{\partial t} - \sigma^2 \frac{\partial^2 v}{\partial x^2} &= f && \text{in } Q, \\ v = -\hat{w} \quad \text{on } \Gamma_l, \qquad v &= 0 && \text{on } \Gamma_r, \\ v &= g && \text{on } Q_0. \end{aligned} \tag{2.5}$$

Due to $\hat{w}(0,0;u^\tau) = g(0) = 0$ for any $u^\tau \in U^\tau$, the smoothness of $\hat{w}$ at $x = 0$ and sufficiently smooth functions f and g, the discontinuities of w are completely captured by $\hat{w}$. Hence, problem (2.5) allows a better numerical treatment than the original PDE.

2.2 Discretization of the state equations

In the preceeding section we described the principle impact of piecewise discretizations of controls to the smoothness of the solutions of the state equations. Now, we sketch consequences of reduced regularity to numerical methods applied to (1.2) with discretized boundary data.

Among the variety of methods let us consider semi-discretization by standard method of lines (MOL) as well as full discretization schemes. The major difference of both approaches is that in the first one standard ODE solvers with efficient step size control can be applied while the full scheme provides a direct access to the time grid which will later be advantageous in evaluating adjoint states for the optimal control problem.

Consider some spatial grid $\{x_i\}_{i=0}^N$ over the interval $[0,1]$, i.e.

$$0 = x_0 < x_1 < \cdots < x_{N-1} < x_N = 1. \tag{2.6}$$

Using simple finite differences we obtain a spatial semi-discretization of the PDE by

$$\begin{aligned} h_{i+1/2}\,\frac{dw_i}{dt}(t) - \sigma^2 \left[\frac{w_{i+1}(t) - w_i(t)}{h_{i+1}} - \frac{w_i(t) - w_{i-1}(t)}{h_i}\right] \\ = h_{i+1/2}\, f(x_i,t), \qquad i = 1,\ldots,N-1 \end{aligned} \tag{2.7}$$

with $h_i := x_i - x_{i-1}$, $i = 1,\ldots N$ and $h_{i+1/2} := (h_i + h_{i+1})/2$. Here and in the sequel w_i denote functions which approximate $w(x_i,\cdot;u^\tau)$. In addition to (2.7) the boundary conditions from (1.2) at $x = 1$ are taken into account by

$$\gamma_D\, w_N(t) = u^\tau(t), \quad t \in (0,T] \tag{2.8a}$$

and

$$\begin{aligned} \frac{h_N}{2}\,\frac{dw_N}{dt}(t) &= \frac{\sigma^2}{\gamma_N}\,[u^\tau(t) - \gamma_D w_N(t)] \\ &\quad -\sigma^2\,\frac{w_N(t) - w_{N-1}(t)}{h_N} + \frac{h_N}{2} f(x_N,t), \quad t \in (0,T] \end{aligned} \tag{2.8b}$$

for $\gamma_N = 0$ and $\gamma_N \neq 0$, respectively, while at $x = 0$ we have in both cases $w_0(t) = 0$. If we consider splitting then instead of (1.2)

we apply semi-discretization to problem (2.5) and we have $v_0(t) = -\hat{w}(0,t;u^\tau)$, $v_N(t) = 0$.

Together with the initial conditions

$$w_i(0) = g(x_i), \quad i = 1,\ldots,N \tag{2.9}$$

we obtain an IVP system for the functions w_i. Notice that in case of Dirichlet control the number of unknowns is $N-1$ otherwise N. We will not explicitly distinguish these cases and write for simplicity in the sequel just N.

In our first approach we treat the IVP (2.7)-(2.9) by standard ODE codes for stiff IVPs. In particular, in our study we applied BDF-codes and trapezoidal rule with automatic step size control.

Alternatively to semi-discretization and standard ODE codes, to which in the sequel we refer shortly as semi-discretization, in a second approach we apply implicit Euler method with a fixed time step T/M to (2.7) - (2.9), which we denote in the sequel as full discretization.

In both approaches discrete states are denoted by $w_{i,j}$, $i = 0,1,\ldots,N$, $j = 0,1,\ldots,M$, where M is the number of time steps.

3. Numerical treatment of the control problem

3.1 Gradient evaluation

Discretization of the controls and the state equations leads to an approximation of the original optimal control problem (1.1), (1.2) by a finite dimensional quadratic programming problem. Let us consider the case that no constraints are imposed upon the controls.

The state equations result in an affine mapping transferring discrete controls $u^\tau \in U^\tau$ into discrete terminal states $w_{\cdot,M}$, i.e. we have

$$w_{\cdot,M} = A_{h,\tau}\, u^\tau + a_{h,\tau} \tag{3.1}$$

with some matrix $A_{h,\tau} \in \mathcal{L}(\mathbb{R}^{M^c}, \mathbb{R}^N)$ and some vector $a_{h,\tau} \in \mathbb{R}^N$. With discrete scalar products $\langle\cdot,\cdot\rangle$ in $\mathbb{R}^N$ and $\mathbb{R}^{M^c}$, respectively, we obtain problems of the type

$$J_{h,\tau}(u^\tau) \to \min\,! \qquad \text{s.t.} \quad u^\tau \in U^\tau \tag{3.2}$$

with

$$J_{h,\tau}(u^\tau) := \frac{1}{2}\langle A_{h,\tau}\, u^\tau - \overline{q}_{h,\tau}, A_{h,\tau}\, u^\tau - \overline{q}_{h,\tau}\rangle + \frac{\alpha}{2}\langle u^\tau - p^\tau, u^\tau - p^\tau\rangle. \tag{3.3}$$

Here $\overline{q}_{h,\tau} := q_h - a_{h,\tau}$, and $p^\tau \in U^\tau$ denotes some approximation of p. Further, the necessary optimality conditions are given by

$$J'_{h,\tau}(u^\tau) = A^*_{h,\tau}\,(A_{h,\tau}u^\tau - \overline{q}_{h,\tau}) + \alpha\,(u^\tau - p^\tau) = 0.$$

It should be noticed that in case of full discretization $A_{h,\tau}$, $a_{h,\tau}$ are known, but will not be constructed explicitly because of the dynamic nature of the discrete state equations. However, in case of semi-discretization where some ODE software code is applied to (2.7)-(2.9) then $A_{h,\tau}$, $a_{h,\tau}$ depend on various additional features, like built-in automatic step size controls. In case of semi-discretization as well as full discretization is applied the image $A_{h,\tau}u^\tau$ can be determined for any $u^\tau \in U^\tau$ by discrete time integration. Moreover, adjoint equations provide an efficient tool for gradient evaluations replacing the calculation of $A^*_{h,\tau}(A_{h,\tau}u^\tau - \overline{q}_{h,\tau})$. For the optimal control problem (1.1), (1.2) the corresponding adjoint problem is defined by (cf. [1], [4], [7], [11])

$$\begin{aligned} \frac{\partial z}{\partial t} + \sigma^2 \frac{\partial^2 z}{\partial x^2} &= 0 \quad \text{in} \quad Q, \\ z = 0 \quad \text{on} \quad \Gamma_l, \quad \gamma_D\, z + \gamma_N \frac{\partial z}{\partial x} &= 0 \quad \text{on} \quad \Gamma_r, \\ z &= w - q \quad \text{on} \quad [0,1] \times \{T\} \end{aligned} \tag{3.4}$$

and the reduced gradient of the objective at $u \in U$ in direction $s \in L_\infty(0,T)$ is given by

$$J'(u)\, s = \frac{\sigma^2}{\gamma_D + \gamma_N} \left\langle z(1,\cdot\,;u) - \frac{\partial z}{\partial x}(1,\cdot\,;u) + \alpha(u-p), s \right\rangle_{(0,T)}. \tag{3.5}$$

Notice that after reversing the time orientation the adjoint problem (3.4) is of parabolic type as the state equation (1.2). However, unlike in the state equation in the adjoint equation we meet incompatibility only at one time level, namely $t = T$.

For the remaining part of this section we restrict ourselves to the case $\gamma_D = 1$, $\gamma_N = 0$. Further, for simplicity in the sequel we consider equidistant spatial grids and denote its step size by $h > 0$.

When applying standard ODE solvers to the related semi-discrete IVP (2.7)-(2.9) and an appropriate discretization to the scalar product in (3.5) we obtain the following approximation of the discrete directional derivative

$$J'(u^\tau)s^\tau \approx \sum_{j=1}^{M^c} \left[\frac{\sigma^2}{h} \sum_{k \in K_j} (\vartheta_k - \vartheta_{k-1}) z_{N-1,k-1} + \alpha\, \tau^j\, (u^j - p^j) \right] s^j, \tag{3.6}$$

where $u^j, p^j, s^j \in \mathbb{R}$, $j = 1,\ldots,M^c$ are the coefficients of $u^\tau, p^\tau, s^\tau \in U^\tau$, $(z_{i,j})$ is the discrete solution of the adjoint problem (3.4), $\{\vartheta_k\}_{k=0}^M$ denotes the time grid generated by the applied ODE solver and

$$K_j := \{k \in \{1,\ldots,M\}\colon \vartheta_k \in (t^{j-1}, t^j]\},\ \tau^j := t^j - t^{j-1},\ j = 1,\ldots,M^c.$$

To obtain (3.6) from (3.5) besides simple integration, the derivative of $\frac{\partial z}{\partial x}$ at $x = 1$ is approximated by one sided finite differences where we take into account the boundary condition $z(1, \cdot) = 0$. Equation (3.6) provides the representation of the discrete gradient $J'(u^\tau)$ via the adjoints.

In case of full discretization, the discrete gradient can be evaluated directly via the corresponding discrete adjoint system. Similarly to the continuous adjoint system after time reversal it turns out to be an implicit Euler scheme again. The obtained formula for the discrete gradient (3.6) can be also interpreted as an approximation of the continuous one.

Our numerical experiments confirmed the fact that the discrete adjoints of the full discretization lead to exact gradients as generated in automatic differentiation tools (see [3]). However, if software tools are applied to semi-discretization of (2.5) and of the adjoint equations (3.4) then only an approximation of the gradients is obtained. One reason for that deviation is that applications of ODE solvers with time step control lead to discretizations of the states and adjoint states with different time grids. Thus the discretization of the adjoint states is not adjoint to the discrete states in the sense of the discrete L_2-norm but only an approximation. Moreover, the summation in the formula for the discrete gradient (3.6) causes a further amplification of the error. Hence, to guarantee convergence of optimization techniques based on this approach a sufficiently high order of accuracy in the applications of ODE software is required which becomes rather expensive for fine discretizations.

3.2 Selected minimization techniques

Since the gradient can be obtained quite easily via adjoint states conjugate gradient methods as well as quasi-Newton techniques (e.g. Broyden's symmetric update, DFP-method, ...) are appropriate for solving the discrete quadratic minimization problem (3.2).

To make the paper self contained we describe briefly the major steps of methods used in our tests for solving (3.2). Let us denote the elements of a sequence $\{\mathbf{u}^l\}$ of discrete controls by $\mathbf{u}^l := u^{\tau,l} \in U^\tau$. In the considered piecewise constant approximation we can represent $\mathbf{u}^l$ by its coefficients $u^{k,l} \in \mathbb{R}$, $k = 0, 1, \ldots, M^c$, $l = 0, 1, \ldots$.

As one of the methods of choice we applied conjugate gradient methods. Starting with some $\mathbf{u}^0 \in U^\tau$ and $\beta_0 := 0$, these methods generate a minimizing sequence $\{\mathbf{u}^l\} \subset U^\tau$ recursively by

$$\begin{aligned} \mathbf{s}^l &:= -J'(\mathbf{u}^l) + \beta_l \, \mathbf{s}^{l-1}, \qquad \beta_{l+1} := \frac{\|J'(\mathbf{u}^l)\|_2^2}{\|J'(\mathbf{u}^{l-1})\|_2^2} \\ \mathbf{u}^{l+1} &:= \mathbf{u}^l + \alpha_l \mathbf{s}^l, \quad \text{with Cauchy step size} \quad \alpha_l > 0. \end{aligned} \tag{3.7}$$

CG methods terminate with the optimal control given by the final minimizer in a finite number of steps provided exact function and gradient evaluations are applied and no rounding errors occur. This is, however, unrealistic in the problems under consideration but the convergence can be accelerated by appropriate preconditioning (cf. [5], [8]). We detected that in the case of Dirichlet control the analytic solution (2.4) which captures jumps in boundary data serves for preconditioning.

In case of unconstrained controls the Cauchy (i.e. minimizing) step size α is easily obtained. However, penalty methods for the treatment of constraints require additional step size procedures.

As other methods of choice we included quasi-Newton methods into our study. Their basic idea is to define the search direction $\mathbf{s}^l$ at $\mathbf{u}^l$ by

$$H_l \mathbf{s}^l = -J_l', \tag{3.8}$$

where $J_l' := J'(\mathbf{u}^l)$ and H_l, $l = 0, 1, \ldots$, denote matrices satisfying the related quasi-Newton equation

$$H_l (\mathbf{u}^l - \mathbf{u}^{l-1}) = J_l' - J_{l-1}', \quad l = 1, 2, \ldots. \tag{3.9}$$

Starting with the identity $H_0 := I$ the matrices H_l are updated by appropriate formulas. In particular, we considered Broyden's symmetric update. Let

$$\mathbf{r}^{l+1} := J_{l+1}' - J_l' - H_l (\mathbf{u}^{l+1} - \mathbf{u}^l).$$

Then the new matrix H_{l+1} is defined by

$$H_{l+1} := H_l + \frac{\mathbf{r}^{l+1} (\mathbf{r}^{l+1})^T}{(\mathbf{r}^{l+1})^T (\mathbf{u}^{l+1} - \mathbf{u}^l)}. \tag{3.10}$$

In the evaluation $\mathbf{u}^{l+1} := \mathbf{u}^l + \alpha_l \mathbf{s}^l$ the step size $\alpha_l > 0$ has been selected according to a simplified Armijo rule. For a detailed description of CG-methods and quasi-Newton methods we refer e.g. to [5], [8].

Occurring constraints

$$|u^j| \leq 1, \quad j = 1, \ldots, M^c$$

on controls have been included by the penalty term ($\rho > 0$)

$$P_\rho(u^\tau) := \frac{c}{2} \sum_{j=1}^{M^c} \tau^j \left[\sqrt{(u^j + 1)^2 + \rho} + \sqrt{(u^j - 1)^2 + \rho} - 2 \right]. \tag{3.11}$$

For $\rho \to 0+$ this tends uniformly to the well-known non-smooth penalty

$$P_0(u^\tau) := c \sum_{j=1}^{M^c} \tau^j \left[\max\{0, -u^j - 1\} + \max\{0, u^j - 1\} \right],$$

which is exact for sufficiently large constant $c > 0$. For $\rho > 0$ the penalty P_ρ is infinitely often differentiable. This forms an advantage in comparison with loss functions. Further, unlike for barriers the values $P_\rho(u^\tau)$ are finite for any discrete control u^τ. For the first derivative and the Hessian we have

$$P'_\rho(u^\tau)\, s^\tau = \frac{c}{2} \sum_{j=1}^{M^c} \tau^j \left[\frac{u^j + 1}{\sqrt{(u^j+1)^2 + \rho}} + \frac{u^j - 1}{\sqrt{(u^j-1)^2 + \rho}} \right] s^j$$

and

$$P''_\rho(u^\tau) = \frac{c\rho}{2} \operatorname{diag} \left(\tau^j \left[\frac{1}{((u^j+1)^2 + \rho)^{3/2}} + \frac{1}{((u^j-1)^2 + \rho)^{3/2}} \right] \right),$$

respectively. These derivatives have been used directly in the quasi-Newton methods, i.e. only components related to $J(\cdot)$ are taken into consideration by the quasi-Newton update. On the other hand, in Armijo's step size rule only penalty terms have to be repeatedly evaluated due to the quadratic nature of $J(\cdot)$. This accelerates the code compared to an application of an all-purpose minimization routine.

4. Numerical experiments

4.1 Preliminaries

In our numerical experiments we tested the performance of different techniques applied to IBVPs (1.2) with discontinuous boundary data as well as studied effects in connection with boundary control problems of tracking type. All experiments are implemented in MATLAB. The focus in Examples 1, 2 was directed towards the behavior of automatic step size procedures in ODE codes and to an improvement of the efficiency of such codes by using the splitting described in subsection 2.1. In connection with optimal control in Examples 3, 4 we studied the influence of discontinuities in boundary data on the convergence of minimization techniques.

In all examples we choose equidistant grids $x_i = i/N$ and $t^k = (Tk)/M^c$. Further, in the first two examples we choose $\sigma = 1/2$, $T = 1$, but in the last two $\sigma = 1$, $T = 0.1$.

The following tables and figures report on numerical results obtained by the BDF-code **ode15s** (option BDF=on) using several maximal orders of consistency (option MaxOrder) and trapezoidal rule **ode23t**, respectively. If not written otherwise, the default values of the relative and absolute error tolerance RelTol=1e−3 and AbsTol=1e−6, respectively, are used. Further, in the Dirichlet case ($\gamma_D = 1$, $\gamma_N = 0$) we split

the experiments in direct solving problem (1.2) by the method of line (named 'direct' in the following tables) and in applying superposition (2.3) to treat occurring jumps in boundary data. In the latter case we solve numerically the remaining smooth problem (2.5). All described effects depend on M^c and the height of the jumps.

4.2 Example 1 (state equations)

For the first example we choose $f \equiv 0, ; g \equiv 0, \quad M^c = 3, \quad N = 50$ with boundary data u^τ in (1.4) according to $\{u^k\} = (1, -2, 3)^T$. Fig. 1 shows the obtained solution $w(x, t; u^\tau)$ for Dirichlet and Neumann boundary conditions, respectively. The number of required time steps is

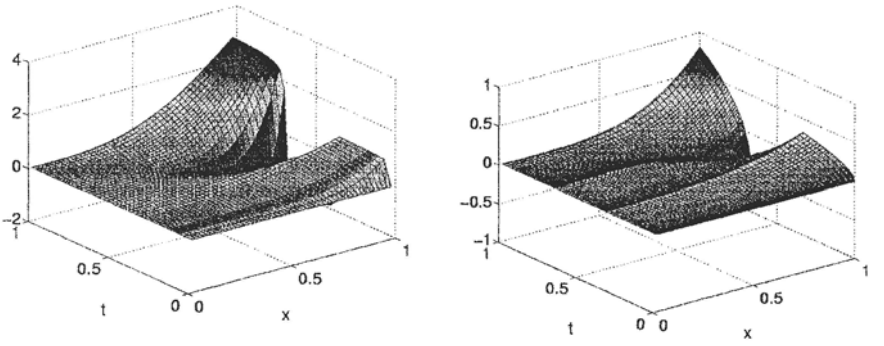


Figure 1. $\gamma_D = 1, \ \gamma_N = 0$ and $\gamma_D = 0, \ \gamma_N = 1$

reported in Tab. 1. For the trapezoidal rule code the related results are marked with T instead of the order as done for BDF code.

treatment	direct				superposition			
maximal order	1	2	5	T	1	2	5	T
obtained time steps	2331	518	322	377	364	111	58	81

Table 1. Comparison of different approaches

The left two graphs in Fig. 2 illustrate the behavior of the automatic step size control when applied directly or after splitting in case of Dirichlet boundary conditions. Further, in the right graph step size results in case of Neumann boundary conditions are reported.

The numerical experiments show (see Fig. 2) that each jump in the control u^τ reduces the time step size drastically. On the other hand, splitting-off the discontinuities (in case of Dirichlet-boundary conditions) in advance avoids these time step size reductions and, hence, yields a more effective numerical procedure.

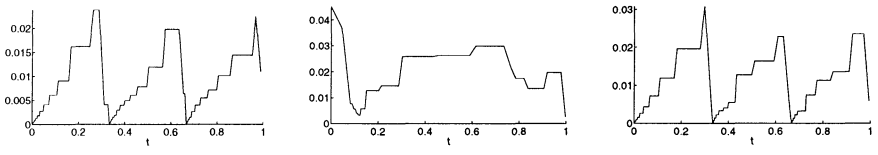


Figure 2. Time step sizes for order 5

4.3 Example 2 (state equation, known exact solution)

In this example we consider a problem with Dirichlet boundary conditions where the exact solution is known. The required discontinuous boundary data are generated by means of the function $\hat{w}$ introduced in Section 2. Unlike in the previous tests here we concentrate on the error behavior. Let the exact solution be given by

$$w(x,t;u^\tau) = g(x) - \left[\hat{w}(x,t;u^\tau) - (1-x)\,\hat{w}(0,t;u^\tau)\right]$$

with $g(x) = 10x^2(1-x)^2$. To study one internal jump only we choose $M^c = 2$ and u^τ in (1.4) according to $\{u^k\} = (1,-1)^T$.

Fig. 3 shows the obtained solution $w(x,t;u^\tau)$ and together with Fig. 4 the error of the BDF-code with MaxOrder=5 for superposition and the direct approach, respectively. In the right picture of Fig. 4 the neighborhood of the point $(x,t) = (1,0)$, where a jump is located, is cut off.

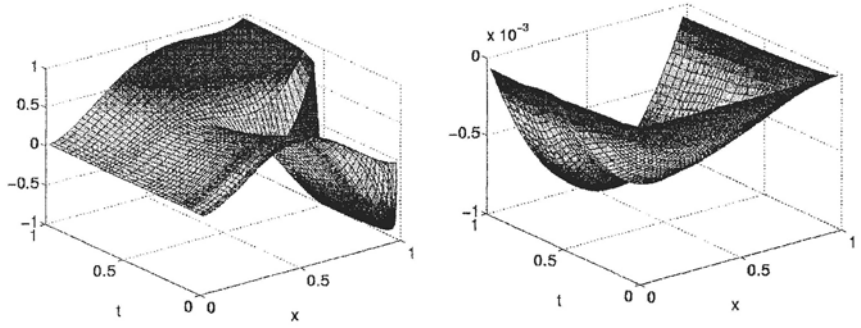


Figure 3. Solution $w(x,t;u^\tau)$ and Error for superposition

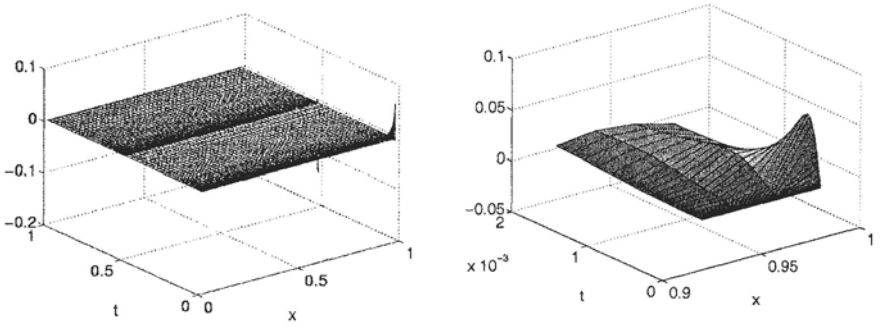


Figure 4. Error in case of the direct solution

Choosing different numbers N of spatial grid points with fixed accuracy RelTol=AbsTol=1e−8 we obtain

N	1600	800	200	50
obtained time steps	1274/133	1138/135	894/140	666/150
error at $t = 1$	9e-07	3e-06	6e-05	9e-04

Table 2. Comparison of required time steps for different N

where in the second row of Tab. 2 the first number is related to direct treatment, the second to superposition.

The numerical experiments reflect (see Tab. 2 and Fig. 3,4), that the step size reduction is the more severe the larger N is.

Finally, we notice that the numerical solution converges at $t = T$, although there is no convergence locally near jumps.

4.4 Example 3 (unconstrained control problem)

We consider the optimal control problem (1.1), (1.2) with

$$p \equiv 0,\ f \equiv 0,\ g \equiv 0 \quad \text{and} \quad q(x) = 0.05\, x^2 \sin(4\pi x).$$

The convergence behavior of a CG-algorithm as well as a quasi-Newton method with Broyden's update is compared for both the approaches discussed in Subsection 3.1, i.e. that the calculation of the discrete gradient (3.5) is based on semi-discretization with discretized continuous adjoints and full discretization with discrete adjoints, respectively. In case of semi-discretization superposition is used for the solution of state as well as for the adjoint state equations. The remaining regular problems were treated by the BDF-code of MATLAB with MaxOrder=5. In Fig. 5 we

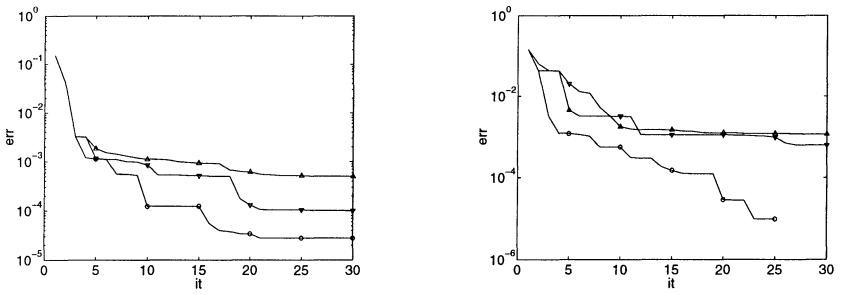
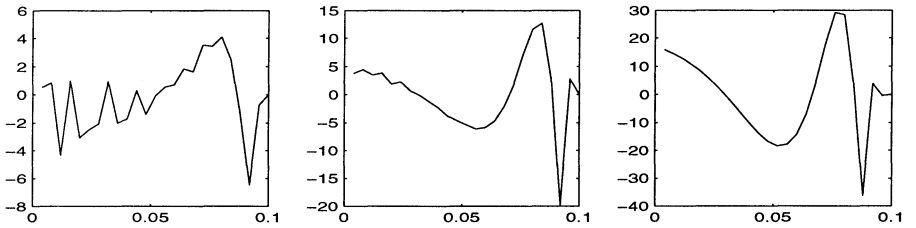


Figure 5. CG-algorithm and Broyden's method

included results from semi-discretization, RelTol=AbsTol=1e−5, semi-discretization, RelTol=AbsTol=1e−12 and full discretization, $M = 500$. The related curves are marked by $\Delta-$, $\nabla-$ and $\circ-$, respectively.

Further, in Fig. 6 the corresponding optimal controls received by the CG-algorithm are reported. We indicate that further slow improvements were obtained beyond the iteration steps plotted in Fig. 6.



a) semi-discretization, 1e−5 b) semi-discretization, 1e−12 c) full discretization

Figure 6. Optimal control obtained by CG-algorithm

In Tab. 3 the influence of the control grid is given for full discretization. Semi-discretizations with sufficiently high accuracy in the ODE solvers show a similar behavior. In general we remark that additionally

M^c	CG method	Broyden's update
10	4.38e-03	4.84e-03
25	2.59e-03	2.78e-03
50	1.42e-03	1.87e-03
100	1.40e-03	1.73e-03

Table 3. Comparison of convergence behavior for different control grids

to slower convergence semi-discretization in both cases of accuracy is

more expensive, i.e. consumes significantly more computer time, than the full discretization.

4.5 Example 4 (constrained control problem)

We choose $f \equiv 0$, $g \equiv 0$. Further, we start with a control problem (1.1), (1.2) which possesses the optimal solution

$$u_{ref}(t) = 1.5 \sin\left(\frac{4\pi t}{T}\right), \quad t \in [0,T]$$

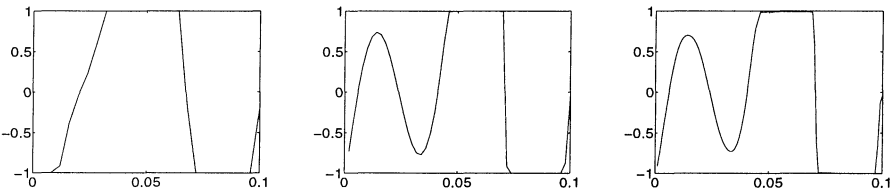
if no constraints are given for the controls. Using this the functions q and p are defined by $q(x) := w(x,T;u_{ref})$ and $p := u_{ref}$, respectively, with the solution $w(\cdot,\cdot\,;u_{ref})$ of the state equation (1.2) for $u = u_{ref}$.

M^c	semi-discretization	full discretization	clipping
10	1.41e-03	1.36e-03	6.93e-03
25	8.83e-04	2.71e-04	2.30e-0.3
50	3.23e-04	1.06e-05	1.10e-03
100	7.28e-04	9.84e-06	6.60e-04

Table 4. Obtained objective values for different control grids

In Tab. 4 the achieved optimal values are reported for the two approaches. In addition, we show in the last column the objective value for the discrete control which is obtained from the unconstrained optimal one by simple clipping along the constraints.

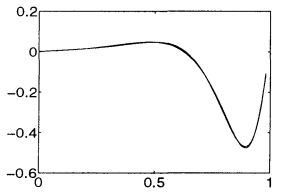
The following Fig. 7 shows discrete optimal controls obtained by Broyden's update (3.10) to the quadratic part (from the state equations) and by direct use of up to second order derivatives of the penalties as given in Section 3. Further, in Fig. 8 the approximation of the tracked target and a comparison between the constrained and the unconstrained optimal controls are given.



$M^c = 25$ $M^c = 50$ $M^c = 100$

Figure 7. Discrete optimal controls, full discretization

The computational experiments showed a very similar behavior as in the unconstrained case. In semi-discretization the state as well as the



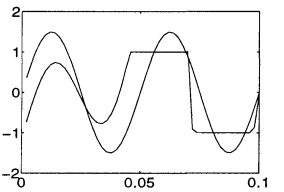


Figure 8. Approximation of the target Constrained, unconstrained control

adjoint system have to be solved with a sufficiently high accuracy to ensure a good approximation of the gradient. This, however, results in high time consumption in the applied ODE solver. On the other hand, full discretization in more general cases (in particular in higher spatial dimensions) requires additional preparatory work compared with the use of available software codes.

5. Conclusions

Piecewise constant discretization of boundary controls yields a reduced smoothness of the solutions of state equations. In all our considered examples this resulted in locally small step sizes if ODE solvers were applied to a semi-discretization of the state equations. These problems could be avoided by considering in advance a specific splitting of the state equations.

In the examples of optimal control problems semi-discretization was only used in connection with a separation of the discontinuities. Hence, the ODE solvers were, in fact, applied to the regular subproblem. Nevertheless, this approach turned out to be more time consuming then full discretization combined with discrete adjoints. In addition, full discretization often yielded better values of the objectives and proved to be faster for comparable accuracy. Further, if lower accuracies were applied to speed up the ODE codes in semi-discretization then the optimization became slow due to the fact that discretizations of continuous adjoint problems lead to only rough approximations of gradients.

References

[1] Casas, E. (1997). Pontryagin's principle for state-constraint boundary control problems of semilinear parabolic equations. *SIAM J. Control Optim.* 35:1297-1327.

[2] Crouzeix, M. and Thomee, V. (1987). On the discretization in time of semilinear equations with nonsmooth initial data. *Math. Comput.* 49:359-377.

[3] Griewank, A. (2000). *Evaluating derivatives: Principles and techniques of algorithmic differentiation.* SIAM Publ., Philadelphia.

[4] Grossmann, C. and Noack, A. (2001). Linearizations and adjoints of operator equations – constructions and selected applications. *TU-Preprint MATH-NM-08-01.*

[5] Grossmann, C. and Terno, J. (1993). *Numerik der Optimierung.* Teubner, Stuttgart.

[6] Hemker, P.W. and Shishkin, G.I. (1993). Approximation of parabolic PDEs with a discontinuous initial condition. *East-West J. Numer. Math.* 1:287-302.

[7] Kelley, C.T. and Sachs, E.W. (1999). A trust region method fo parabolic boundary control problems. *SIAM J. Optim.* 9:1064-1081.

[8] Nocedal, J. and Wright, S.J. (1999). *Numerical optimization.* Springer, New York.

[9] Rannacher, R. (1984). Finite element solution of diffusion problems with irregular data. *Numer. Math.* 43:309-327.

[10] Sammon, P. (1983). Fully discrete approximation methods for parabolic problems with nonsmooth initial data. *SIAM J. Numer. Anal.* 20:437-470.

[11] Tröltzsch, F. (1984). *Optimality conditions for parabolic control problems and applications.* Teubner, Leipzig.

[12] Tröltzsch, F. (1994). Semidiscrete Ritz-Galerkin approximation of non-linear parabolic boundary control problems - strong convergence of optimal controls. *Appl. Math. Optim.* 29:309-329.

[13] Tychonoff, A.N. and Samarsky, A.A. (1959). *Differentialgleichungen der Mathematischen Physik.* Verlag d. Wissenschaft, Berlin.