M a t h e m a t i c a
B a l k a n i c a

New Series Vol. 24, 2010, Fasc.3-4

Semi-Smooth Newton Methods for the Time Optimal Control of Nonautonomous Ordinary Differential Equations [1]

Jelena Rubeša and Karl Kunisch

The control problem of minimal time transition between two stationary points are formulated in a framework of an indirect numerical method. The problem is regularized and the monotone behavior of the regularisation procedure is investigated. Semi-smooth Newton method applied on the regularized problems converge superlinearly and usually produce a very accurate solution. Differently from other methods, this one does not need a-priory knowledge of the control switching structure. A code was developed in the C++ language and the NVIDIA CUDA technology made it even faster.

AMS Subj. Classification: 49J15, 49M15

Key Words: time optimal control, switching structure, regularization, semi-smooth Newton method, CUDA technology

Introduction

This paper addresses time optimal control problems for a class of linear non-autonomous multi-input controls systems for ordinary differential equations. Due to structural difficulties, time optimal control has been receiving a considerable amount of attention for decades. Much of the literature up to the late sixties is covered in [5]. Many recent results can be found or are referenced in [2, 8, 10]. Time optimal control for infinite dimensional systems is considered in [3], for example.

The optimality system associated to time optimal control problems with pointwise constraints on the controls is complicated due to lack of smoothness of the optimal controls. In fact, the first order optimality system for time optimal control problems contains a multi-valued operation which complicates the

[1]This work was completed with the support of project NAWI Graz.

use of fast numerical methods. For this reason we introduce a regularization to the time optimal problem. In section 1 the behavior of the solutions of the regularized problems as the regularization parameter ε tends to zero is investigated. In particular monotonic structure of the solutions with respect to ε is shown. An optimality system for the regularized problems is derived under a normality condition. The optimal controls of the regularized problems are $W^{1,\infty}$ regular and converge to a minimum norm solution of the original problem as the regularization parameter tends to zero.

The optimality system of the regularized problems is still not C^1 but the optimal controls are now Lipschitz functions. In finite dimensions, locally Lipschitz continuous functions are almost everywhere differentiable, by Rademacher's theorem. This concept is not available in infinite dimensions so second order methods with local quadratic convergence order are not directly applicable. However, sufficient conditions will be obtained in section 2 which imply that semi-smooth Newton methods [6] are well-posed and locally superlinearly convergent.

Section 3 contains a brief description of numerical results. We compare the chosen regularization to an alternative one, which has stronger regularization properties. Since the optimal controls of the original time optimal problems are typically not continuous, it appears that our choice of regularization which leads to $W^{1,\infty}$ regularized controls is preferable over others which provide smoother ones.

Let us note that the approach that we propose for solving time optimal problems deviates from traditional approaches, which are frequently grouped into direct and indirect methods. The basic idea of direct methods is to discretize the control problem in order to obtain nonlinear programming (NLP) problem, which may be solved by NLP techniques such as Initial Value Solver (IVS) and Sequential Quadratic Programming (SQP). These methods use only control and state variables. In this sense direct methods are easier to implement but less accurate than the indirect methods; cf., e.g. Reference [1], [8].

The second group consists of *indirect optimization methods*, which are based on the solution of the first order necessary conditions of optimality, the Pontryagin minimum principle (PMP). We shall use this approach. The two point boundary value problem (TPBVP) arising from the PMP is solved by the *Semi-smooth Newton's method*, and quite contrary to the multiple shooting method, does not need a priori knowledge of switching structure. In practice, it is usually rather difficult to determinate the optimal switching structure in

advance, especially for the multi switch problems. The Newton's method will converge superlinearly and usually produces a very accurate solution.

1. The Time-optimal problem and its Regularization

Consider the time-optimal control problem for the non-autonomous linear multi-input system

$$\text{(P)} \qquad \begin{cases} \min \int_{t_0}^{t_1} dt \\ \text{subject to} \\ \quad \frac{d}{dt}x(t) = A(t)x(t) + B(t)u(t),\ |u(t)|_{\ell^\infty} \le 1, \\ \quad x(t_0) = x_0,\ x(t_1) = x_1, \end{cases}$$

where t denotes a scalar time variable on some nonempty closed interval $J \doteq [t_0, t_1]$ of the positive real line. Here $A \in L^\infty(t_0, t_1; \mathbb{R}^{n\times n})$ and $B \in L^\infty(t_0, t_1; \mathbb{R}^{n\times m})$ are two given coefficient matrices. The columns of $B(t)$ are denoted by $b_i(t)$ for $i = 1, 2, \ldots, m$. The vector function $x(\cdot)$ satisfying problem (P) is a system trajectory and a vector x is a state of the system. We assume that the initial state x_0 and terminal state x_1 are two given n-dimensional column vectors. The vector $u(\cdot)$ is control used to modify the system. We will assume that the control u is a bounded measurable m-vector, and $|\cdot|_{\ell^\infty}$ denotes the infinity-norm on $\mathbb{R}^m$. Also we assume that x_1 can be reached in finite time by an admissible control. Then (P) admits a solution with optimal (τ^*, x^*, u^*), where τ^* is the optimal time, x^* is the optimal state and u^* is the optimal control. For the proof we refers to [3].

The first order optimality system for (P) can be expressed in terms of the adjoint p and the Hamiltonian

$$H(x, u, p_0, p) = p_0 + p^T(Ax + Bu),$$

as

$$\text{(1.1)} \qquad \begin{cases} \dot{x} = Ax + Bu,\ x(t_0) = x_0,\ x(t_1) = x_1, \\ -\dot{p} = A^T p, \\ u = argmin_{|u|_{\ell^\infty}\le 1} H(x, u, p_0, p), \\ p_0 + p(t_1)^T(A(t_1)x(t_1) + B(t_1)u(t_1)) = 0,\ p_0 \ge 0, \end{cases}$$

where the superscript T denotes transposition. For details, see page 27 of [4]. Due to the special structure of H the optimal control can be expressed as

$$u_i = -\sigma(s_i), \tag{1.2}$$

where $s_i = p^T b_i$ denotes the switching function for the ith control variable and σ denotes the coordinate-wise operation

$$\sigma(s_i) = \begin{cases} -1 & \text{if } s_i < 0 \\ [-1,\, 1] & \text{if } s_i = 0 \\ 1 & \text{if } s_i > 0. \end{cases} \tag{1.3}$$

Introducing the transformation $\hat{t} = \frac{t}{\tau}$ and setting $t_0 = 0$, $t_1 = \tau$,

$$\hat{x}(\hat{t}) \doteq x(\tau\hat{t}) = x(t),\ \hat{p}(\hat{t}) \doteq p(\tau\hat{t}) = p(t),\ \hat{u}(\hat{t}) \doteq u(\tau\hat{t}) = u(t),$$
$$\hat{A}(\hat{t}) \doteq A(\tau\hat{t}) = A(t),\ \hat{B}(\hat{t}) \doteq B(\tau\hat{t}) = B(t),$$

we obtain the following equivalent system to (1.1):

$$\begin{cases} \dot{\hat{x}} = \tau\hat{A}\hat{x} + \tau\hat{B}\hat{u},\ \hat{x}(0) = x_0,\ \hat{x}(1) = x_1 \\ -\dot{\hat{p}} = \tau\hat{A}^T\hat{p} \\ \hat{u} = argmin_{|\hat{u}|_{\ell^\infty} \le 1} H(\hat{x}, \hat{u}, \hat{p}) \\ 1 + \hat{p}(1)^T(\hat{A}(1)\hat{x}(1) + \hat{B}(1)\hat{u}(1)) = 0 \end{cases} \tag{1.4}$$

where $\hat{t} \in [0, 1]$. The non-differentiable operation involved in characterizing the optimal control,

$$u = -\sigma(B^T p),$$

compare (1.2), prohibits the use of Newton-type methods for solving (1.4) numerically.

Therefore a family of regularized problems

$$(P_\varepsilon) \qquad \begin{cases} \min_{\tau \ge 0} \int_0^\tau (1 + \frac{\varepsilon}{2}\,|u(s)|^2)\, ds \\ \text{subject to} \\ \qquad \frac{d}{dt}x(t) = A(t)x(t) + B(t)u(t),\ |u(t)|_{\ell^\infty} \le 1, \\ \qquad x(0) = x_0,\ x(\tau) = x_1, \end{cases}$$

with $\varepsilon > 0$ is considered. The norm $|\cdot|$ used in the cost-functional denotes the Euclidean norm. It is straightforward to argue the existence of a solution $(u_\varepsilon, x_\varepsilon, \tau_\varepsilon)$.

The convergence of the solutions $(x_\varepsilon, p_\varepsilon, u_\varepsilon, \tau_\varepsilon)$ of (P_ε) to a solution (x^*, p^*, u^*, τ^*) of (P) is considered next. Note that (x^*, p^*, u^*, τ^*) is unique if the normality condition is assumed.

Theorem 1.1. *For $\varepsilon \to 0^+$ we have $\tau_\varepsilon \to \tau^*$ and every convergent subsequence of solutions $\{(u_\varepsilon, x_\varepsilon)\}_{\varepsilon>0}$ to (P_ε) converges in $L^2(0, \tau_\epsilon; \mathbb{R}^m) \times W^{1,2}(0, \tau_\epsilon; \mathbb{R}^n)$ to a solution (u^*, x^*) of (P), where u^* is a minimum norm solution.*

Corollary 1.2. *If normality holds, then the solution u^* to (P) is unique, it is bang-bang, and $u_\varepsilon \to u^*$ in L^2 as $\varepsilon \to 0^+$.*

We turn to the optimality condition for (P_ε). Let

$$\sigma_\varepsilon(s) = \begin{cases} -1 & \text{if } s \le -\varepsilon \\ \frac{s}{\varepsilon} & \text{if } |s| < \varepsilon \\ 1 & \text{if } s \ge \varepsilon. \end{cases} \tag{1.5}$$

If σ_ε is applied to a vector, then it acts coordinate-wise.

Theorem 1.3. *Assume that normality of $(\hat{A}, \hat{B})$ holds on $[\alpha, 1]$ for every $\alpha \ge 0$, that $\hat{A} \in W^{1,\infty}(0,1; \mathbb{R}^{n\times n})$, $\hat{B} \in W^{1,\infty}(0,1; \mathbb{R}^{n\times m})$ and let $(x_\varepsilon, u_\varepsilon, \tau_\varepsilon)$ be a solution of (P_ε). If there exist an interval $J_1 \doteq (\alpha, \alpha+\delta)$, $\delta > 0$, $\eta > 0$ such that*

$$|\,\hat{u}_\varepsilon(t)|_{\ell^\infty} \le 1 - \eta \quad \text{for a.e. } t \in J_1, \tag{1.6}$$

then there exists an adjoint state p_ε such that

$$\begin{cases} \dot{x}_\varepsilon = A\,x_\varepsilon + B\,u_\varepsilon,\ x_\varepsilon(0) = x_0,\ x_\varepsilon(\tau_\varepsilon) = x_1 \\ -\dot{p}_\varepsilon = A^T\,p_\varepsilon \\ u_\varepsilon = -\sigma_\varepsilon(B^T\,p_\varepsilon) \\ 1 + \frac{\epsilon}{2}|u_\varepsilon(\tau_\varepsilon)|^2_{\mathbb{R}^m} + p_\varepsilon(\tau_\varepsilon)^T\big(A(\tau_\varepsilon)\,x_\varepsilon(\tau_\varepsilon) + B(\tau_\varepsilon)\,u_\varepsilon(\tau_\varepsilon)\big) = 0. \end{cases} \tag{1.7}$$

The third claim in (1.7) reveals the extra regularity of u_ε:

Corollary 1.4. *Under the assumptions of Theorem 1.3 we have*

$$u_\varepsilon \in W^{1,\infty}(0, \tau_\varepsilon; \mathbb{R}^m)\,.$$

2. Semi-smooth Newton methods

In this section the semi-smooth Newton method for solving the regularized optimality system (1.7) is described and analyzed. It will allow that (1.7) can be solved efficiently in spite of the fact that σ_ε is not Fréchet differentiable.

Throughout we fix $\varepsilon > 0$ and denote by $(x_\varepsilon, u_\varepsilon, \tau_\varepsilon) \in W^{1,2}(0,1) \times L^2(0,1) \times \mathbb{R}$ a solution to (P_ε) with associated adjoint $p_\varepsilon \in W^{1,2}(0,1)$. It is assumed that normality of (A, B) holds on $[\alpha, 1]$ for every $\alpha \geq 0$, $A \in W^{1,\infty}(0,1; \mathbb{R}^{n\times n})$ and $B \in W^{1,\infty}(0,1; \mathbb{R}^{n\times m})$, that

(H1) there exists $\bar{s} \in (0,1)$ such that $|\frac{1}{\varepsilon} B(\bar{s})^T p_\varepsilon(\bar{s})|_{\ell^\infty} = |u_\varepsilon(\bar{s})|_{\ell^\infty} < 1$

and

(H2) $$|b_i(1)^T p_\varepsilon(1)| \neq \varepsilon, \text{ for all columns } i = 1, \dots, m \text{ of } B.$$

We now use the fact that u_ε is continuous and assumption (H1) implies (1.6) in some neighborhood $(\alpha, \alpha + \delta)$ containing $\bar{s}$. The existence of $p_\varepsilon \in W^{1,2}(0,1)$ follows, by Theorem 1.3, such that (1.7) holds. With (H1) and (H2) holding there exists a neighborhood $\mathcal{U}_{p_\varepsilon}$ of p_ε in $W^{1,2}(0,1;\mathbb{R}^n)$, $\bar{t} \in (0,1)$ such that for the interval $(\alpha, \alpha + \delta) \subset (0,1)$ we have for every $p \in \mathcal{U}_{p_\varepsilon}$

$$|B(t)^T p(t)|_{\ell^\infty} < \varepsilon \text{ for } t \in (\alpha, \alpha + \delta) \tag{2.1}$$

and

$$|b_i(t)^T p(t)| \neq \varepsilon \text{ for all } t \in [\bar{t}, 1], \text{ and } i = 1, \dots, m.$$

We set $U = \{u \in L^2(0, 1; \mathbb{R}^m) : u|_{[\bar{t}, 1]} \in W^{1,2}(\bar{t}, 1; \mathbb{R}^m)\}$ endowed with the norm

$$|u|_U = (|u|^2_{L^2(0,1)} + |\dot{u}|^2_{L^2(\bar{t},1)})^{\frac{1}{2}},$$

and introduce

$$F : D_F \subset X \to L^2(0,1;\mathbb{R}^n) \times L^2(0,1;\mathbb{R}^n) \times U \times \mathbb{R}^n \times \mathbb{R},$$

where

$$D_F = W^{1,2}(0,1) \times \mathcal{U}_{p_\varepsilon} \times U \times \mathbb{R},$$

$$X = W^{1,2}(0,1;\mathbb{R}^n) \times W^{1,2}(0,1;\mathbb{R}^n) \times U \times \mathbb{R},$$

and

$$F(x,\,p,\,u,\,\tau) = \begin{pmatrix} \dot{x} - \tau A x - \tau B u \\ -\dot{p} - \tau A^T p \\ u + \sigma_\varepsilon(B^T p) \\ x(1) - x_1 \\ 1 + \frac{\varepsilon}{2}|u(1)|^2 + p(1)^T\,(A(1)x(1) + B(1)u(1)) \end{pmatrix}. \tag{2.2}$$

Note that $F = (F_1, \ldots, F_5)$ is well-defined. This is obvious for F_1, F_2 and F_3. For F_4, F_5 it follows from the fact that $W^{1,2}(0,1)$ embeds continuously into $C(0,1)$. Moreover $F(x_\varepsilon,\, p_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon) = 0$. We shall keep $x(0) = x_0$ as an explicit constraint.

Applying Newton's method to $F = 0$ is impeded by the non-differentiability of σ_ε. We use

$$G\sigma_\varepsilon(s) := \begin{cases} \frac{1}{\varepsilon} & \text{if } |s| < \varepsilon \\ 0 & \text{if } |s| \geq \varepsilon \end{cases} \tag{2.3}$$

as a generalized derivative. The Newton iteration step is given by

$$DF(x,\,p,\,u,\,\tau)(\delta x,\,\delta p,\,\delta u,\,\delta\tau) = -F(x,\,p,\,u,\,\tau) \tag{2.4}$$

where $\delta x(0) = 0$ and DF denotes the Fréchet derivative in all terms of F except for $p \to \sigma_\varepsilon(B^T p)$, for which the generalized derivative is taken according to (2.3). For further reference we give the detailed form of (2.4):

$$\begin{cases} \frac{d}{dt}\delta x - \tau A\,\delta x - \tau B\,\delta u - \delta\tau\,(A\,x + B\,u) = -F_1 \quad \delta x(0) = 0 \\ -\frac{d}{dt}\delta p - \tau A^T\,\delta p - \delta\tau\,A^T\,p = -F_2 \\ \delta u + G\sigma_\varepsilon(B^T\,p)\,B^T\,\delta p = -F_3 \\ \delta x(1) = -F_4 \\ p(1)^T\,\big(-A(1)\,F_4 + B(1)\,\delta u(1)\big) + \varepsilon u(1)^T\,\delta u(1) \\ \qquad\qquad +\delta p(1)^T\,\big(A(1)\,x(1) + B(1)\,u(1)\big) = -F_5, \end{cases} \tag{2.5}$$

where the coordinates of $G\sigma_\varepsilon(B^T p)B^T\delta p$ are given by $G\sigma_\varepsilon((B^T p)_i))(B^T\delta p)_i$.

A possible initialization may consist of choosing $((u)_0,\, \tau_0)$, setting $(x)_0$ as the linear interpolation between x_0 and x_1, and determining $(p)_0$ such that the transversality condition and the adjoint equation are satisfied.

We now briefly summarize those facts from semi-smooth Newton methods which are relevant for this paper. Proofs can be found in [6, 11, 9, 7]. Let X and Z be Banach spaces and let $F : D_F \subset X \to Z$ be a nonlinear mapping with an open domain D_F.

Definition 2.1. The mapping $F : D_F \subset X \to Z$ is called Newton-differentiable at $\mathbf{x}$, if there exists an open neighborhood $N(\mathbf{x}) \in D_F$ and mappings $DF : N(\mathbf{x}) \to \mathcal{L}(X, Z)$ such that

$$\lim_{h \to 0} \frac{1}{|h|_X} |F(\mathbf{x} + h) - F(\mathbf{x}) - DF(\mathbf{x} + h)h|_Z = 0. \tag{2.6}$$

The family $\{DF(s) : s \in N(\mathbf{x})\}$ is called a Newton-derivative of F at $\mathbf{x}$.

Note that F does not need to be Fréchet-differentiable in order to have the property (2.6). In general, there exists a set of Newton-derivatives at $\mathbf{x}$ which becomes a singleton whenever F is Fréchet-differentiable. If the mapping F is Newton-differentiable for each $\mathbf{x}$ in an open subset $\mathcal{U} \subset D_F$, then we say that F is Newton-differentiable on $\mathcal{U}$.

Lemma 2.2. *[Chain rule] Suppose that $H : D_H \subset X \to Y$ is Newton-differentiable at $\mathbf{x} \in D_H$ and $F : Y \to Z$ is Newton-differentiable at $H(\mathbf{x})$. Then the mapping $g = F(H)$ is Newton-differentiable at $\mathbf{x}$.*

Theorem 2.3. *Suppose that $\mathbf{x}^* \in \mathcal{U}$ is a solution to $F(\mathbf{x}) = 0$ and that F is Newton-differentiable in an open set $\mathcal{U}$ containing $\mathbf{x}^*$ with Newton-derivative DF. If further $\{\|DF(\mathbf{x})^{-1}\| : \mathbf{x} \in \mathcal{U}\}$ is bounded, then the Newton-iteration*

$$\mathbf{x}_{k+1} = \mathbf{x}_k - DF(\mathbf{x}_k)^{-1} F(\mathbf{x}_k)$$

converges q-superlinearly to $\mathbf{x}^*$, provided that $|\mathbf{x}_0 - \mathbf{x}^*|_X$ is sufficiently small.

For the statement and the proof of the superlinear convergence of the time-optimal control problem, some further notation is required. For $(x, p, u, \tau) \in D_F$ we define $\mathcal{A} \in \mathbb{R}^{(n+1)\times(n+1)}$ by

$$\mathcal{A} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & 0 \end{pmatrix},$$

where

$$A_{11} = \varepsilon^{-1}\tau \int_0^1 \Phi(1,t)\, B(t)\, \chi_I\, B(t)^T\, \Phi(1,t)^T\, dt \in \mathbb{R}^{n\times n} \tag{2.7}$$

$$
\begin{aligned}
A_{12} = \varepsilon^{-1}\tau \textstyle\int_0^1 \Phi(1,t)\, B(t)\, \chi_I\, B(t)^T \int_t^1 \Phi(t,s)^{-T}\, A(s)^T\, p(s)\, ds\, dt \\
- \textstyle\int_0^1 \Phi(1,t)\big(A(t)\, x(t) + B(t)\, u(t)\big)\, dt \in \mathbb{R}^n
\end{aligned} \tag{2.8}
$$

$$
\begin{aligned}
A_{21} = \big(A(1)\, x(1) + B(1)\, u(1)\big)^T \\
-\big(p(1)^T\, B(1) + \varepsilon\, u(1)^T\big) G\sigma_\varepsilon\big(B(1)^T\, p(1)\big)\, B(1)^T \in (\mathbb{R}^n)^T,
\end{aligned} \tag{2.9}
$$

where $\chi_I = diag(\chi_{I_1}, \ldots, \chi_{I_m})$ and χ_{I_i} is the characteristic function of the set

$$I_i = I_i(p) = \{t : |b_i^T p| < \varepsilon\}, \quad i = 1, \ldots, m$$

which is nonempty for every $p \in \mathcal{U}_{p_\varepsilon}$, i and $t \in (\alpha,\, \alpha + \delta)$. The normality assumption together with (H1) implies that the symmetric matrix A_{11} is invertible with uniformly bounded inverse with respect to $p \in \mathcal{U}_{p_\varepsilon}$ and τ in compact subsets of $(0, \infty)$. In fact, since $I_i(p) \supset (\alpha,\, \alpha + \delta)$ holds for every i and $p \in \mathcal{U}_{p_\varepsilon}$, the matrix χ_I is an identity matrix on $(\alpha,\, \alpha + \delta)$ for every $p \in \mathcal{U}_{p_\varepsilon}$ and hence

$$
\begin{aligned}
A_{11} &\geq \varepsilon^{-1}\, \tau \textstyle\int_\alpha^{\alpha+\delta} \Phi(1,t)\, B(t)\, B(t)^T\, \Phi(1,t)^T\, dt \\
&= \varepsilon^{-1}\, \tau \Phi(1) \textstyle\int_\alpha^{\alpha+\delta} \Phi(t)^{-1}\, B(t)\, B(t)^T\, \Phi(t)^{-T}\, dt\, \Phi(1)^T
\end{aligned} \tag{2.10}
$$

where τ belongs to the family of closed bounded neighborhoods of τ_ε in $\mathbb{R}^+$.

We assume that

$$
\begin{cases}
\text{there exists a bounded neighborhood} \\
\mathcal{U} \subset D_F \subset X \text{ of } (x_\varepsilon,\, p_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon) \text{ and } c > 0 \text{ such that} \\
|A_{21}\, A_{11}^{-1}\, A_{12}| \geq c \text{ for all } (x,\, p,\, u,\, \tau) \in \mathcal{U}.
\end{cases} \tag{H3}
$$

Theorem 2.4. *If normality,* (H1)–(H3) *hold and* $(x_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon)$ *denotes a solution to* (P_ε) *with associated adjoint* p_ε*, then the semi-smooth Newton algorithm converges superlinearly, provided that the initialization is sufficiently close to* $(x_\varepsilon,\, u_\varepsilon,\, \tau_\varepsilon)$.

3. A numerical example

We consider

$$\text{(3.1)} \qquad \begin{cases} \min_{\tau\geq 0} \int_0^\tau dt \\ \text{subject to} \\ \frac{d}{dt}x(t) = A(t)x(t) + B(t)u(t) \\ |u(t)| \leq 1,\ x(0) = x_0,\ x(\tau) = x_1, \end{cases}$$

where

$$A = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, \qquad x_0 = \begin{pmatrix} 1/2 \\ 1/6 \end{pmatrix}, \qquad x_1 = \begin{pmatrix} 0 \\ 0 \end{pmatrix},$$

$$B = \begin{pmatrix} 25t & 0 \\ 0 & 2t \end{pmatrix} \chi_{\left[0,\frac{1}{3}\right)} + \begin{pmatrix} -t & 0 \\ 0 & 2t \end{pmatrix} \chi_{\left[\frac{1}{3},\frac{1}{2}\right)} + \begin{pmatrix} -t & 0 \\ 0 & -t \end{pmatrix} \chi_{\left[\frac{1}{2},\frac{2}{3}\right]} + \begin{pmatrix} t & 0 \\ 0 & -t \end{pmatrix} \chi_{\left(\frac{2}{3},1\right]}.$$

To solve (3.1) numerically a time discretization based on the Crank-Nicolson method on the equidistant mesh with the mesh size $h = 1/(N+1)$ is applied to (2.5). The initialization for the state was chosen as a semicircle connecting x_0 and x_1. Then $u(1)$ and $u(1-h)$ were chosen to be active, τ such that the state equation held and p was chosen so that the transversality condition and the adjoint equation were satisfied. The Newton system (2.5) was solved iteratively and the iteration was stopped when the equation residual was smaller than 10^{-6} in the L^2-norm. The optimal time of steering the state form x_0 to the origin is $\tau^* = 0.272752$. This reference measure was obtained by discretization for $N = 8192$ and $c = 500$, since the exact solution is not available.
In this paper we choose to regularize σ by the ramp functions

$$\text{(3.2)} \qquad \sigma_\varepsilon(s_i) = \begin{cases} -1 & \text{if } s_i \leq -\varepsilon \\ \frac{s_i}{\varepsilon} & \text{if } |s_i| < \varepsilon \\ 1 & \text{if } s_i \geq \varepsilon. \end{cases}$$

In Table 1 we show the number of iterates of the Newton iteration (outer loop) that was required for this procedure with respect to the different $c = \frac{1}{\varepsilon}$. Also in Table 1 we depict the optimal minimal times $\tau^*(c)$. These results are obtained for $N = 512$.
Certainly other alternatives are possible for approximation of the function σ, as

for instance

$$\sigma_{atan}(s_i) = \frac{2}{\pi}\, atan\big(\varepsilon^{-1}\, s_i\big). \tag{3.3}$$

This family of $C^\infty-$ functions also has the property that it converges to σ as $\varepsilon \to 0^+$, but it appears to be less appropriate for the purpose of approximating the discontinuous switching structure of the optimal controls. We refer to Table 2, which corresponds to the results in Table 1, only with the approximation procedure (3.2) changed to (3.3).

c	1	10	20
# It	34	37	29
$\tau^*(c)$	0.3115	0.2731	0.2731

TABLE 1. Approximation σ_ε.

c	1	10	50	100	500
# It	19	140	50	84	168
$\tau^*(c)$	0.4765	0.3012	0.277	0.2746	0.2733

TABLE 2. Approximation σ_{atan}.

For the purpose of comparison, the same values for c in the first two computational processes were taken. Clearly c has to be taken significantly larger in the case of (3.3) than for (3.2) to obtain comparable results. The final time for the σ_ε approximation with $c = 10$ is quite accurate when compared to the optimal τ^* and further computations do not considerably improve the accuracy. This is not the case for the σ_{atan} approximation where the result improves as c is increased. With $c = 500$ we obtain the same four-digit accuracy as for $c = 10$ in the σ_ε approximation.

The small discrepancy between this converged value and τ^* is considered to be the discretization error. In some cases, typically at the beginning of the iterations and for the lowest values of c the full Newton step was too large. Therefore we used a one-dimensional line search based on a quadratic polynomial interpolation for the L^2- norm of the residual combined with an Armijo rule.

The graphs for the corresponding controls for σ_ε approximation and $N = 512$ are given in Figure 1. The plots were obtain for $c = 10$. Since the example

(3.1) has two coordinate for the control value, $m = 2$, the upper plots present the first coordinate of the control and the lower ones the second coordinate of u. The first coordinate jumps two times while the second one only one time.

It. no	25	26	27	28	29	30	31	32
c_k	0.94	0.8069	0.5947	0.3814	0.3772	0.3023	0.0307	0.0009

TABLE 3. Approximation σ_ε for $c = 10$ and $N = 512$.

Table 3 shows the quotients $c_k = \dfrac{|u^{k+1} - u^*(c)|_{L^2}}{|u^k - u^*(c)|_{L^2}}$, where $u^*(c)$ is the solution to the discretize version of (3.1) using the ramp function (3.2) for $c = 10$ and $N = 512$. It shows that the algorithm is in fact superlinearly convergent.

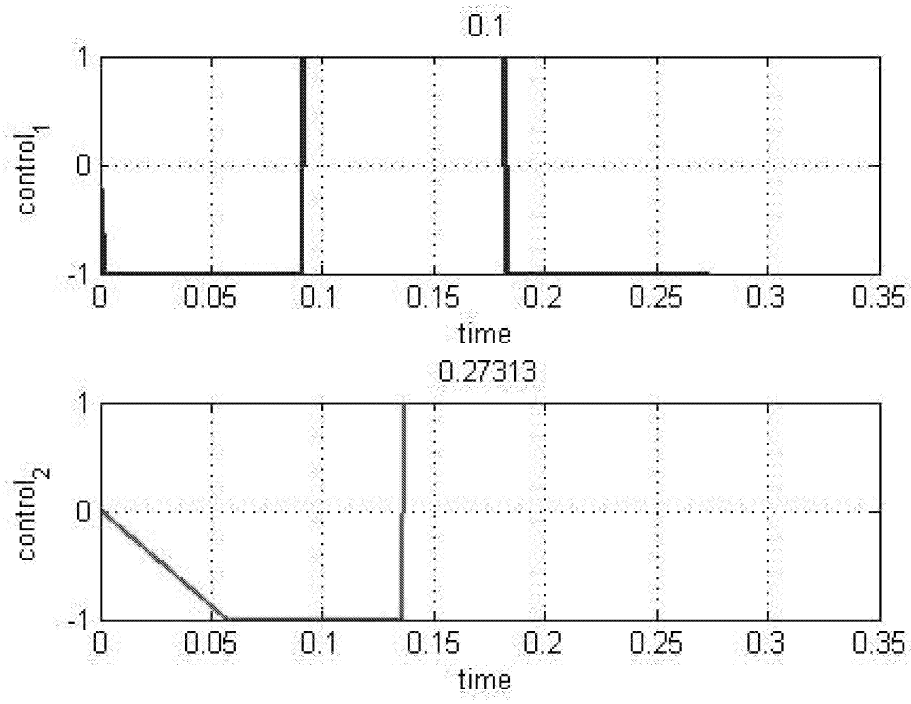

FIGURE 1. Approximation σ_ε for $N = 512$ and $c = 10$.

Precision	*Hardware*	*It.no*	*Time*
Single	Intel Pentium4 2.8GHz	3871	50614.2 ms
Single	Nvidia 8800 GT	3825	996.941 ms
Single	AMD Phemon 9950 2.6GHz	3401	20162.9 ms
Single	Nvidia GTX 280	3859	1073.21 ms
Double	AMD Phemon 9950 2.6GHz	1342	9347.58 ms
Double	Nvidia GTX 280	1342	405.809 ms

TABLE 4. Difference between a CPU and GPU real-time computing at each iteration of Newton's method.

3.1. **CUDA Architecture.** The algorithm for solving the time optimal control problem (3.1) has the outer loop which present the Newton iteration of the semi-smooth Newton method. Each Newton loop has the inner loop which solves the sparse linear system using a CGNR method.

In the Table 1 for $c = 10$ we needed 37 Newton's loop but each Newton loop needs a huge number (3871) of iterations in inner loop to find an appropriate Newton step, see Table 4. Firstly a code was written in the C/C++ language and additionally a GPU enabled version of the code was developed using Nvidia's CUDA technology. The key component for the GPU implementation was an efficient sparse matrix-vector multiplication kernel for the CGNR iteration. Two workstations with two different graphics boards were used for the numerical tests. Overall the single precision benchmark gives a speedup of 51x and 19x. While the double precision benchmark gives a speedup of 23x.

References

[1] C. Büskens and H. Maurer, Sqp-methods for solving optimal control problems with control and state sonstraints:adjoint variables, sensitivity analysis and real-time, *Journal of Computational and Applied Mathematics*, **120**, 2000, 85–108.

[2] H. J. Pesch Ch. Büskens and S. Winderl, Real-time solutions for bang-bang and singular optimal control problems, *Online Optimization of Large Scale Systems*, 2001, 129–142.

[3] H. O. Fattorini, *Infinite dimensional linear control systems: The time optimal and norm optimal problems*, Elsevier, Amsterdam, 2005.

[4] Wendell H. Fleming and Raymond W. Rishel, Deterministic and stochastic optimal control, *Bulletin of the American Mathematical Society*, **82**, 1976, 869–870.

[5] H. Hermes and J. LaSalle, *Functional analysis and time optimal control*, Academic Presse, 1969.

[6] K. Ito and K. Kunisch, *On the Lagrange multiplier approach to variational problems and applications.*

[7] K. Ito and K. Kunisch, The primal-dual active set method for nonlinear optimal control problems with bilateral constraints, *SIAM Journal on Control and Optimization*, **43**, 2004, 357–376.

[8] G. L. Lippi, J.-H. R. Kim and H. Maurer, Minimzing the transition time in lasers by optimal control methods. single-mode semiconductor lasers with homogenous transverse profile, *Physica. D* **191**, 2004, 238–260.

[9] K. Ito, M. Hintermüller and K. Kunisch, The primal–dual active set strategy as a semi–smooth newton method, *SIAM Journal on Optimization*, **13**, 2002, 865–888.

[10] H. Maurer and N. P. Osmolovskii, Second order sufficient optimality conditions for time-optimal bang-bang control, *SIAM J. Control and Optimization*, **42**, 2004, 2239–2263.

[11] M. Hintermüller and K. Kunisch, *Pde-constrained optimization subject to pointwise control and zero- and first-order state constraints*, *Siam Journal on Optimization*, **20**, 2009, 1133–1156.

Institute for Mathematics
and Scientific Computing,
University of Graz,
Heinrichstrasse 36
A-8010 Graz, Austria

E-Mail:jelena.rubesa@uni-graz.at