**Proceeding of the 2004 American Control Conference**
**Boston, Massachusetts June 30 - July 2, 2004**



# Direct Transcription Solution of Inequality Constrained Optimal Control Problems

John T. Betts Stephen L. Campbell Anna Engelsone

*Abstract*— **Direct transcription is a popular way to solve the complex optimal control problems that arise in industry. With a direct transcription approach, the problem is fully discretized and then the discrete problem is solved numerically. Recently it has been shown that the theory for direct transcription differs in several key ways from the theory for other approaches. These differences have implications for numerical algorithms and the interpretation of solutions to practical problems. This paper examines some of those differences.**

## I. Introduction

Optimal control problems arise in essentially all areas. In practical applications there are almost always a number of constraints. These constraints can arise either on physical grounds or from design or operating restrictions. The constraints can be equalities or inequalities and can involve the control, the state, or both.

A number of approaches have been developed to numerically solve optimal control problems. Methods that rely on the necessary conditions in some form can be very useful on particular problems. However, these methods suffer from the fact that adding new constraints can require deriving new necessary conditions. Also, in many complex problems actually getting the necessary conditions in a useful form can be a very difficult task. For example, there can be a number of constraints going active and inactive with a complex switching structure [2]. It may not be possible to get even a reliable estimate on the number of switches much less which constraints they involve. We shall point out another difficulty later in this paper.

One of the alternatives that is frequently used in industry is direct transcription because it is often the easiest way to formulate complex problems [1]. The control problem is fully discretized in time (and in space if it is a PDE control problem), and then the resulting finite dimensional optimization problem is passed to a nonlinear programming (NLP) code. Upon solution of the NLP problem, the solution is evaluated and if found wanting, the temporal mesh is refined and the new larger NLP problem is solved. While simply described, a direct transcription code is actually quite complex because of the need to have sophisticated mesh refinement algorithms, state of the art NLP solvers, and the exploitation of sparse matrix and other structures.

John Betts is with Mathematics and Engineering Analysis, The Boeing Company, P.O. Box 3707 MS 7L-21, Seattle, Washington 98124-2207.

Steve Campbell is with the Department of Mathematics, North Carolina State University, Raleigh, NC 27695-8205 USA. slc@math.ncsu.edu

Anna Engelsone is with the Operations Research Program, North Carolina State University, Raleigh, NC 27695-7913.

Over the last decade there has been a considerable amount of research on direct methods for the solution of optimal control problems including those that arise from the control of partial differential equations (PDEs). We note here only [9], [10], [11], [19], [20], [24] and [15], [16]. There has also been considerable work on the convergence of the discretizations used when constraints are present, see for example, [17], [18]. This research has been important in advancing the understanding and implementation of direct transcription algorithms. Our emphasis here is somewhat different. We will be discussing how the theory from the areas of optimal control and the numerical solution of constrained differential equations (DAEs) requires careful, and in some cases a different, interpretation when applied to direct transcription methods.

For the last several years we have been working on improvements [5], [7], [6], [8] of Boeing's optimal control software project, SOCS (Sparse Optimal Control Software) [4]. During this effort several things have become apparent. First, provided one has sophisticated mesh refinement strategies and state of the art NLP solvers and sparse software, direct transcription works very well on a wide variety of problems. It provides solutions of many problems that would be hard to solve by other methods. Secondly, theory that fully explains this is lacking. Thirdly, some of the existing theory for optimal control needs to be modified when talking about direct transcription and failure to take this into account can lead to wrongful analysis and misinterpretation of results. This third point is the focus of this paper. While our observations come from working with SOCS, they apply to any direct transcription code and in that sense are fully general.

Section 2 briefly summarizes some very recent results which are probably new to the ACC audience. Section 3 presents new observations and new computational experience with a PDE control problem. To simplify the discussion we will refer to approaches other than direct transcription as indirect methods. Portions of this paper appear in the technical report [3].

## II. The Discretization Need not "Converge"

A differential algebraic equation (DAE) is a mixed system of differential and algebraic equations

$$F(x', x, t, u) = 0$$

where the Jacobian of $F$ with respect to $x'$ is identically singular. Often physical models themselves are DAEs [12].

**0-7803-8335-4/04/$17.00 ©2004 AACC**

However, even if the original model is explicit, the addition or activation of constraints means that DAE theory will play a role. While space prohibits a review of DAE theory here, we note that an integer called the index measures how far the differential equation is from being an explicit differential equation. An ordinary differential equation (ODE) is index zero. For simple problems, the index is the number of differentiations needed to produce an ODE.

SOCS has a number of discretizations available. The default is to use the Lobatto IIIA formulas trapezoid (TR) and Hermite-Simpson (HS) which are second and fourth order Runge-Kutta methods. Typically SOCS begins with TR since experience has shown that it is usually easier to then get a feasible solution, and then uses HS. However, sometimes it is necessary to start with HS [6]. Theoretical results exist for these discretization when applied to DAEs [12], [21], [23]. In general, the results say that the index needs to be no more than three, and in the case of TR it needs to be two or less. The equations also always need to have a special structure if the index is greater than one.

If the DAE is arising because of equality constraints, then both direct and indirect approaches require that the discretization be able to integrate the constrained dynamics. As a consequence both approaches are limited to low index (less than or equal to three), and SOCS is limited theoretically to index two. In practice, SOCS is limited to index one because of the need to find feasible solutions. Some direct and indirect methods have been successfully applied to index three mechanical systems by utilizing specially designed integrators and exploiting the special structure of the equations of motion.

With state inequality constraints, the situation is completely different. When a constraint is active, the indirect method still needs to be able to solve the resulting DAE numerically on the interval the constraint is active on. Thus its discretization must be able to integrate that DAE. Surprisingly this is not true for direct transcription approaches. We have observed them successfully solve a number of optimal control problems where the discretization was not convergent on the DAE that resulted when the constraint was active.

In an indirect method the failure of the discretization occurs in part because small errors get amplified as the integration proceeds. That is, the discretization error growth equation is unstable. However, in a direct approach the optimizer can choose to make small deviations from the active inequality constraint. The instability of the discretization means that this small perturbation can cancel out the large error growth. The result is that the optimization algorithm can get a good approximation of the optimal solution even though the discretization is not convergent in the classical sense. Thus the interpretation and consequences of the unstable error growth equation are reversed when working with direct transcription methods instead of indirect methods.

This behavior was pointed out in [14], with convergence results in [7]. It should be pointed out that this analysis does not use the discrete Lagrange multipliers since we have observed that they may not converge even though SOCS is finding the solution correctly.

In the example presented in the next section we will see that SOCS solves problems with inequality constraints of order over 20. If such an inequality constraint is active then the resulting DAE would have index at least 21. Yet there is no known discretization that is practical or convergent for DAEs of index greater than 4! Clearly the standard DAE theory is not correctly describing what is happening with inequality constraints.

## III. The Necessary Conditions can be Misinterpreted

In this section we consider a test problem which is a boundary controlled heat equation with an inequality constraint on the temperature profile. The discretized version of this problem leads to an optimal control problem with an inequality constraint whose order increases with the fineness of the spatial discretization.

Consider the boundary control of the one dimensional heat equation such as an insulated metal rod. $u(x,t)$ is the temperature at point $x$ at time $t$. The spatial and temporal intervals are $0 \leq x \leq \pi$, and $0 \leq t \leq 5$ respectively. The control objective is to keep the temperature profile close to zero but still have the temperature profile stay above a time varying profile $g$. The controls are the temperatures at the ends of the rod. Our control problem is to find state $u$ and control $v_0, v_\pi$ to minimize $J$ where

$$J(u, v_0, v_\pi) = \int_0^\pi \int_0^5 u(x,t)^2 dx dt + \int_0^5 q_1 v_0^2(t) + q_2 v_\pi^2(t) dt \tag{1}$$

subject to the constraints

$$u_t = u_{xx} \tag{2a}$$
$$u(0,x) = u_0(x) \tag{2b}$$
$$u(t,0) = v_0(t) \tag{2c}$$
$$u(t,\pi) = v_\pi(t) \tag{2d}$$
$$u(x,t) \geq g(x,t) \tag{2e}$$

For fixed scalars $c, a$, we will take $g$ to be

$$g(x,t,a,c) = c \sin x \sin\left(\frac{\pi t}{5}\right) - a \tag{3}$$

If the values of $a, c$ are clear from the discussion, we shall omit them from the notation.

As is common practice, the PDE control problem will be solved by discretizing using the method of lines and then applying an ODE optimization package. Let $N$ be a positive integer and $x_i = i\frac{\pi}{N}$. This partitions the $x$ interval into $N$ equal intervals and uses $N+1$ grid points. Let $u_i(t) = u(t, x_i)$ for $i = 1, \ldots N-1$ be the value of the temperature at time $t$ and position $x_i$. The values of $u$ at the endpoints $x_0, x_N$ are take to be the control variables; $v_0 = u(0,t),\ v_\pi = u(\pi, t)$. We retain the temporal derivative $u_t$ in (2) and approximate $u_{xx}$ using centered differences. Let $\delta = \frac{\pi}{N}$.

The original constraints (2) then become the inequality constrained ODE

$$u_1' = \frac{1}{\delta^2}(u_2 - 2u_1 + v_0) \tag{4a}$$

$$u_i' = \frac{1}{\delta^2}(u_{i+1} - 2u_i + u_{i-1}), \quad \text{for} \quad 2 \leq i \leq N-2 \tag{4b}$$

$$u_{N-1}' = \frac{1}{\delta^2}(v_\pi - 2u_{N-1} + u_{N-2}) \tag{4c}$$

$$u_i(t) \geq g(x_i, t) \tag{4d}$$

The $x$ integration in the original cost (1) is approximated by the trapezoid rule so that the new cost is

$$\delta \left( \int_0^5 v_0^2(t) + v_\pi^2(t) + \sum_{i=1}^{N-1} 2u_i(t)^2 \ dt \right) + \int_0^5 q_1 v_0^2(t) + q_2 v_\pi^2(t) \ dt \tag{5}$$

The control problem (4) with cost (5) has parameters $a, c, N, q_i$. Let $\overline{q}_i = q_i + \delta$.

We take $\overline{q}_1 = \overline{q}_2 = 10^{-3}$ which is typical of the situation where the control weight is really for numerical regularization in the PDE problem and we are using a fine spatial mesh. We also take the initial temperature profile to be zero, $u(0,x) = 0$. This problem has a very interesting feature. We have solved this problem for a variety of values of $N$ and $a, c$. We assume that $N$ is even since it simplifies the discussion but the numerical results are similar for $N$ odd. The only difference is that the constraints are active at two spatial points if $N$ is odd. We initially take $c = 1, a = 0.7$. Then a typical solution profile for $u$ is given in Figure 1. The controls are in Figure 2. The controls differ by very little as $N$ changes once $N > 5$.

Since the problem is symmetric with respect to $x$, so is the solution. Examining the numerical solutions, which were computed to an accuracy of $10^{-7}$, the only place the constraints were active is at $x_{N/2}$ and then only for some $t_i$. Figure 3 graphs the grid points where the constraints were active and is typical of what one sees for different $N$.

If the constraint (4d) is active, (4) is a DAE in $v_i, u_i$ and the index is $\frac{N}{2} + 1$. Equivalently, the order of the constraint is $\frac{N}{2}$ and goes to infinity as $N$ does. We have solved this problem with $N$ up to 40. Note that this means that we are solving a problem with order 20 inequality constraint and

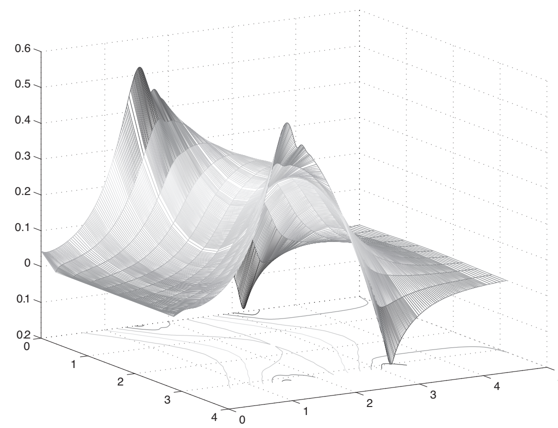


Fig. 1. Optimal state $u(x,t)$ for problem (1),(2) using approximation (4),(5) with $N = 10$.

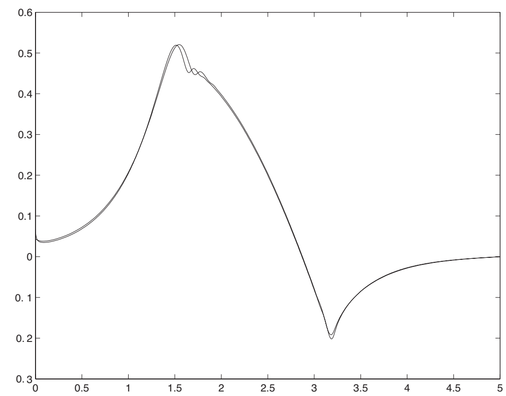


Fig. 2. Optimal control $v$ for approximation (4),(5) with $N = 10$ and $N = 31$.

the index 21 DAE that results if the constraint is active on an interval.

*1) What is an arc?:* Figures 4 and 5 show the plot of the constraint surface as a function of $x, t$ and the solution at $x_{N/2}$. Figure 4 shows the entire constraint surface and $x_{N/2}$. Figure 5 shows the $N = 10$ case just at $x_{N/2}$. Both figures seem to show a smooth transition onto the constraint surface, riding the surface, and then a departure. The graphs appear the same for other values of $N$ greater than three.

However, when these results were shown at a workshop, it was pointed out that the computational answers might not be correct because there was a classical result in the literature which said that under certain mild appearing technical assumptions, which hold for this problem, that the solution of an odd order state constrained problem of order 3 or higher cannot ride a constraint over a nonzero length interval [22]. The only possibility is one or more touch and goes. That is, the intervals a constraint is active have zero length. Subsequently, this problem was solved by a very different approach by a different group. They got answers similar to ours.

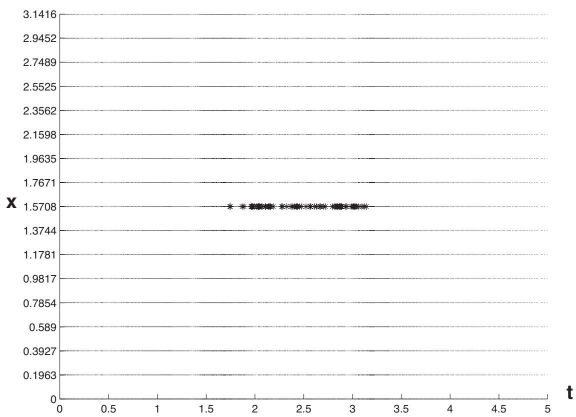


Fig. 3. Points $(x_i, t_i)$ where the constraints were active for $N = 16$ when solving problem (4),(5).

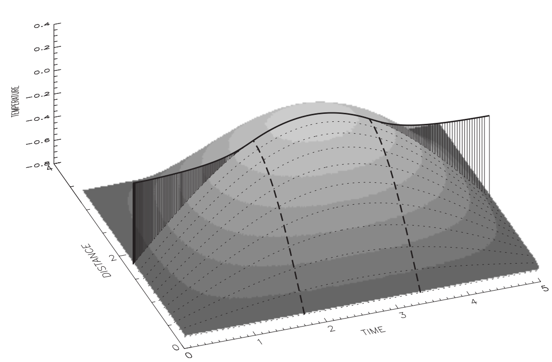


Fig. 4. $u_{N/2}(t)$ and the constraint surface with $N = 10$ when solving (4),(5). Dashed lines indicate apparent ends of the constraint arc.

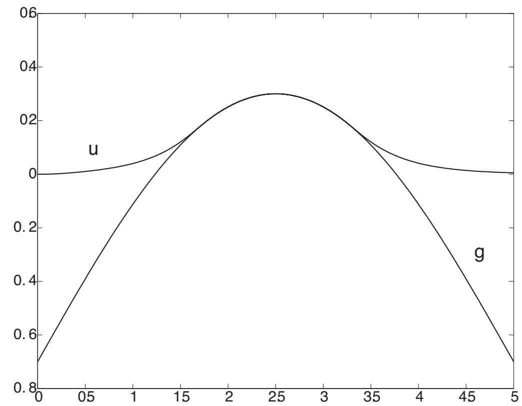


Fig. 5. $u_{N/2}(t)$ and the constraint surface for $N = 10$ when solving problem (4),(5).

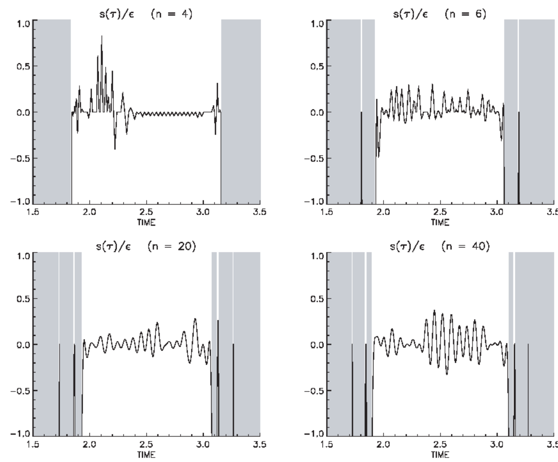


Fig. 6. Normalized deviation from the constraint for several $N$ when solving problem (4),(5). Vertical axis scale is in terms of multiples of $10^{-7}$.

It is important to understand what is happening numerically on this problem in order to interpret the computational result. While Figures 4 and 5 seem to show an interval on which the constraint is active we observe in Figure 3 that there is a small gap which is several mesh points wide near each end. This is typical. In order to examine this more carefully we plot the deviation from the constraint $s(t)$ normalized to the square root of machine precision $\epsilon$ which is the SOCS tolerance. That is, we plot $\min\{s(t)/\epsilon, 1\}$. This expands the deviation within one magnitude of the precision. The result is in Figure 6 for several values of $N$.

What we see in Figure 6 is apparently a riding of the constraint over an interval containing $[2, 3]$ along with one or two small touch and goes before and after the interval with the touch and goes not being visible to the naked eye but showing up numerically. The touch and goes are indicated by the vertical black half lines in the shaded area.

But does this tell the full story? Consider Figure 7 which shows a slightly wider window. It depicts the deviation from the constraint for two values of $N$. The graph for $N = 31$ has been displaced upward to make comparison easier. The touch and goes now appear as the sharp peaks at either end. We see that the peaks are very small, of height only about $10^{-6}$. However, what is of equal interest is the behavior in between. We see that the two curves look like noisy copies of the same curve. For this particular set of parameters, the problem appears to have an even finer structure. However, the allowable tolerances in SOCS do not permit us to directly resolve this behavior by calculating to higher accuracy.

We are in the process of examining this problem more carefully but what we know now is the following.

It is true that the analytical solution does not ride the constraint just as the theory predicts. However, the theoretical deviates from the constraint by less than the accuracy to which the problem is being solved. The SOCS solution is correct to the very high accuracy that is requested which is $10^{-7}$.

Another way to view the ill conditioning of this problem

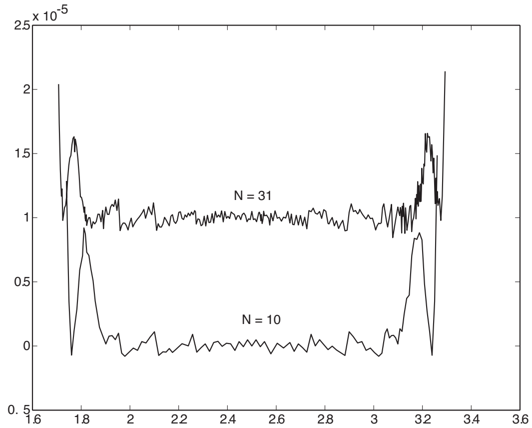


Fig. 7. Constraint error for $N = 10$ and $N = 31$ when solving problem (4),(5).

is to observe that for this example the theory shows there is not an active constraint arc. Yet there is an interval on which the solution state is within $10^{-7}$ of the constraint. Thus, for example, if $N = 20$, the solution of the optimal control probelm is very close to the solution of an index 21 DAE and the numerical solution is very close to solving that index 21 DAE even though ithe discretization would be wildly divergent if actually applied as a integrator to that index 21 DAE.

Note also that if one were to try and use the necessary conditions to define the problem, then one would have to add a large number of unknown activation times. Then one would get an extremely ill conditioned problem to solve. It is also not clear how many of the tiny touch and goes actually exist. In this situation we see that it is advantageous to not have to strictly adhere to the necessary conditions. Direct transcription finds a solution to the desired accuracy which most would agree is the behavior that one would see physically. In some sense, direct transcription relaxes the problem and finds a solution accurate up to the level requested thereby avoiding having to deal with some of the technical issues posed by the necessary conditions.

## IV. Acknowledgements

Research supported in part by National Science Foundation Grants DMS-0101802, DMS-0209695, and ECS-0114095.

## References

[1] J. T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, Philadelphia, 2001.

[2] J. T. Betts, *Parametric tool path trajectory optimization*, Boeing Technical Document, SSGTECH-98-006, March, 1998, The Boeing Corporation, Seattle.

[3] J. T. Betts and S. L. Campbell, *Discretize Then Optimize*, Technical Document Series, &CT-TECH-03-01, Mathematics and Computing Technology, Phantom Works, Boeing, Seattle, 2003.

[4] J. T. Betts and W. P. Huffman, *Sparse Optimal Control Software SOCS*, Mathematics and Engineering Analysis Technical Document MEA-LR-085, Boeing Information and Support Services, July, 1997.

[5] N. Biehn, J. Betts, S. Campbell, and W. Huffman, *Compensating for Order Variation in Mesh Refinement for Direct Transcription Methods*, J Comp. Appl. Math., 125 (2000), 147–158.

[6] J. T. Betts, S. L. Campbell, and K. Nagapradeep, *Initialization of Direct Transcription Optimal Control Software*, Proc. IEEE Conf. Decision & control, 2003, 3802–3807.

[7] John Betts, Neil Biehn, and S. L. Campbell, *Convergence of nonconvergent IRK discretizations of optimal control problems with state inequality constraints*, SIAM J. Sci. Comp., SIAM J. Sci. Comp., 23 (2002), 1982–2008.

[8] John Betts, Neil Biehn, Stephen Campbell, and William Huffman, *Compensating for order variation in mesh refinement for direct transcription methods II: computational experience*, J. Comp. Appl. Math., 143 (2002), 237–261.

[9] L. T. Biegler, *Optimization strategies for complex process models*, Advances in Chemical Engineering, 18 (1992), 197–256.

[10] L. T. Biegler and S. Kameswaran, *Equivalence between direct and indirect methods for DAE-constrained optimization*, Carnegie Mellon University, preprint, 2003.

[11] H. G. Bock, M. M. Diehl, D. B. Leineweber and J. P. Schlder, *A direct multiple shooting method for real-time optimization of nonlinear DAE processes*. Nonlinear model predictive control (Ascona, 1998), 245–267, Progr. Systems Control Theory, 26, BirkhŁuser, Basel, 2000.

[12] K. E. Brenan, S. L. Campbell, and L. R. Petzold, *Numerical Solution of Initial-Value Problems in Differential-Algebraic Equations*, SIAM, 1996.

[13] A. E. Bryson, Jr., W. F. Denham, and S. E. Dreyfus, *Optimal programming problems with inequality constraints I: necessary conditions for extremal solutions*, J. AIAA 1 (1963), 2544–2550.

[14] S. Campbell, N. Biehn, L. Jay, and T. Westbrook, *Some comments on DAE theory for IRK methods and trajectory optimization*, J. Comp. Applied Math, 120 (2000), 109–131.

[15] P. D. Christofides and P. Daoutidis, *Feedback control of hyperbolic PDE systems*, AIChE Journal, 42 (1996), 3063–3086.

[16] P. D. Christofides and P. Daoutidis, *Finite dimensional control of parabolic PDE systems using approximate inertial manifolds*, J. Math. Anal. Appl., 216 (1997), 398–420.

[17] A. Dontchev and W. Hager, *The Euler approximation in state constrained optimal control*, Math. Comp., 70 (2001), 173–203.

[18] A. Dontchev, W. Hager and V. Veliov, *Second-order Runge-Kutta approximations in constrained optimal control*, SIAM J. Numer. Anal., 38 (2000), 202–226.

[19] P. J. Enright and B. A. Conway, *Discrete approximations to optimal trajectories using direct transcription and nonlinear programming*, AIAA Journal of Guidance Control and Dynamics, 15 (1992), 994–1002.

[20] W. F. Feehery and P. I. Barton, *Dynamic simulation and optimization with inequality path constraints*, Comp. and Chem Engng, 20 (1996), S707–S712. FeBa96aFeBa96a

[21] E. Hairier and G. Wanner, *Solving Ordinary Differential Equations II, Stiff and Differential Algebraic Problems*, Springer-Verlag, 1991.

[22] D. H. Jacobsen, M. M. Bee, and J. L. Speyer, *New necessary conditions of optimality for control problems with state variable inequality constraints*, J. Math. Anal. Appl. 35 (1971), 255-284.

[23] L. Jay, *Convergence of a class of Runge-Kutta methods for differential-algebraic systems of index 2*, BIT, 33 (1993), 137-150.

[24] S. Vasantharajan and L. T. Biegler, *Simultaneous strategies for optimization of differential-algebraic systems with enforcement of error criteria*, Computers Chem. Engng., 14 (1990), 1083–1100.