JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 22, No. 2, March–April 1999

Engineering Notes

ENGINEERING NOTES are short manuscripts describing new developments or important results of a preliminary nature. These Notes cannot exceed 6 manuscript pages and 3 figures; a page of text may be substituted for a figure and vice versa. After informal review by the editors, they may be published within a few months of the date of receipt. Style requirements are the same as for regular contributions (see inside back cover).

Treatment of Control Constraints in Finite Element Solution of Optimal Control Problems

Michael S. Warner* and Dewey H. Hodges†
Georgia Institute of Technology, Atlanta, Georgia 30332-0150

I. Introduction

IN Ref. 1, high-order finite elements in time were used to solve optimal control problems with control constraints. In that work, control inequality constraints were converted to equality constraints via the use of slack variables. It was discovered that overall error in the finite element solution correlated with errors in the so-called switching times, where the control constraints become active or inactive.

In that same reference, the methodology to handle problems with discontinuities in the state variables and/or system equations was also explored for finite elements in time. In that case, the end points of the different phases, or time intervals with continuous states and system equations, could be variables in the problem.

In the current work, these two ideas are combined, resulting in a scheme that implements high-order finite elements in time to approximately solve problems with control inequality constraints of known switching structure as multiphase problems. This technique is demonstrated on a problem from Ref. 2, with comparisons shown between handling the control-constrained problems as having a single phase or multiple phases.

II. Optimal-Control Problems

As in Ref. 1, the systems being studied are governed by a set of n_x general, nonlinear state equations,

$$\dot{x} = f(x, u, t), \qquad x \in R^{n_x}, \quad u \in R^{n_u}, \quad t \in [0, t_f] \tag{1}$$

in terms of the states x, the controls u, and the time t. Here we are assuming a single phase, so that the states and system equations are continuous for all t from zero to the final time t_f.

A set of n_{bc} general, nonlinear boundary conditions on the states are specified at some combination of the initial and final times in the form

$$\Psi[x(0), x(t_f), t_f] = 0, \qquad \Psi \in R^{n_{bc}} \tag{2}$$

Received Sept. 18, 1997; revision received Oct. 19, 1998; accepted for publication Oct. 24, 1998. Copyright © 1998 by Michael S. Warner and Dewey H. Hodges. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission.

*Graduate Research Assistant, School of Aerospace Engineering; currently Space Systems Engineer, National Air Intelligence Center, Data Exploitation Directorate, U.S. Air Force, Wright-Patterson Air Force Base, OH. Member AIAA.

†Professor, School of Aerospace Engineering. Fellow AIAA.

The cost functional for this problem can include a scalar penalty on the states at the endpoints of the time interval and/or an integral penalty on the states, controls, and time:

$$J = \phi[x(0), x(t_f), t_f] + \int_0^{t_f} L(x, u, t)\, dt \tag{3}$$

All admissible control histories are assumed to be bounded and piecewise continuous, while also meeting n_g control inequality constraints g formulated as

$$g_i[x(t), u(t)] \le 0, \qquad g \in R^{n_g}, \quad t \in [0, t_f], \quad i \in [1, n_g] \tag{4}$$

where each g_i is a function of at least one control, noting that multiple constraints may restrict a single control.

The constraint is considered active at time t if $g_i[x(t), u(t)] = 0$, which is enforced through use of slack variables k, such that Eq. (4) is replaced by the equality constraints

$$g_i[x(t), u(t)] + k_i^2(t) \equiv g_i + K_i = 0, \qquad i \in [1, n_g] \tag{5}$$

The optimal control equations can be found in Ref. 2, among other sources. Of interest here are the equations used to determine the controls, the first set of which are the optimality conditions in terms of the Hamiltonian H:

$$\frac{\partial H}{\partial u_i}(x, \lambda, u, t) = 0, \qquad i \in [1, n_u] \tag{6}$$

where $H \equiv L + \lambda^T f + \mu^T (g + K)$ for costates λ and control constraint Lagrange multipliers μ.

Then, control constraint equations

$$g_i + K_i = 0, \qquad 2\mu_i k_i = 0, \qquad i \in [1, n_g] \tag{7}$$

provide the remaining necessary equations. The latter of Eqs. (7) implies strict complementarity for μ_i and k_i.

In the finite element approximations of the optimal control equations,[1] the time interval is broken up into N elements, which can be of different lengths. The states, costates, controls, slack variables, and control constraint Lagrange multipliers are each represented on each element as linear combinations of special Jacobi polynomials outlined in Ref. 3, which are of arbitrary degree p. The coefficients of these Jacobi polynomials are then the unknowns in the nonlinear finite element equations, which are summarized in Ref. 1, with the number of equations depending on n_u, n_x, n_g, N, and p.

III. Implementation and Results

The main problem being studied is from Ref. 2. The problem is to minimize the cost function

$$J = \frac{1}{2}x^2(t_f) + \frac{1}{2}\int_0^{t_f} u^2\, dt \tag{8}$$

where x and u are scalars. The system dynamics are governed by

$$\dot{x} = h(t)u \tag{9}$$

for some function $h(t)$, subject to the constraint that the magnitude of the control function always be less than one. This is expressed as two control inequality constraints:

$$g_1 = u - 1 \le 0, \qquad g_2 = -(u + 1) \le 0 \tag{10}$$

One combination of parameters such that an exact solution is available[4] is if the final time is chosen to be 10, the initial condition is a given constant, and the function $h(t)$ is chosen to be

$$h(t) = 1 + t - \tfrac{3}{17}t^2 \tag{11}$$

The resulting exact value for the state at the final time is $x(t_f) = -\frac{17}{39}$, which is the exact value of the corresponding (constant) costate as well, whereas the optimal control is

$$u(t) = \begin{cases} -x(t_f)h(t) & 0 \le t \le 2 \\ 1 & 2 \le t \le \frac{11}{3} \\ -x(t_f)h(t) & \frac{11}{3} \le t \le 8 \\ -1 & 8 \le t \le 10 \end{cases} \tag{12}$$

The exact time history for the state is a polynomial given in Ref. 1 in terms of u.

Thus, $t = 2, \frac{11}{3}$, and 8 are switching times when the control comes on or off a constraint. Because of the strict complementarity, the corresponding μ and k are transitioning onto or off of zero, respectively, at these times.

However, in this finite element formulation, a single set of equations applies at all points in time. With no a priori way to know where the switching points are within the finite element mesh, no guarantee exists that an element boundary will fall on any switching point. Thus, in any element that straddles an optimal switching point, because μ or k could only be zero or nonzero over the entire element, errors are introduced.

To help quantify these errors, the Fortran code called GENCODE (as described in Ref. 1) was used to solve the finite element equations corresponding to this problem. In the discretization of the time interval, GENCODE allows for an arbitrary initial finite element mesh. Then, once a given finite element solution has been obtained, elements can be subdivided or combined, but no mechanism exists to move the element boundaries.

This problem was run for various degrees of approximating polynomials (or shape functions) and element lengths, in each case starting with 12 elements, so that none of the nodes happened to fall on an optimal switching point. Figures 1 and 2 show the performance of the different order shape functions, with each data point representing twice as many elements as the point to its left. Errors do reduce as the mesh is refined and the shape function orders are increased, but the performance of the code was better on problems without control constraints.

IV. Multiphase Problems

In an attempt to reduced the errors in the slack-variable formulation, we explored treating the control-constrained problem as a multiphase problem. In Ref. 1, the hp-version finite element equations are presented for two-phase problems in which the differential equations and integral penalty are allowed to be discontinuous at a single point in time t_1, which can be fixed or free. Likewise, extra boundary conditions on the states and the scalar part of the cost

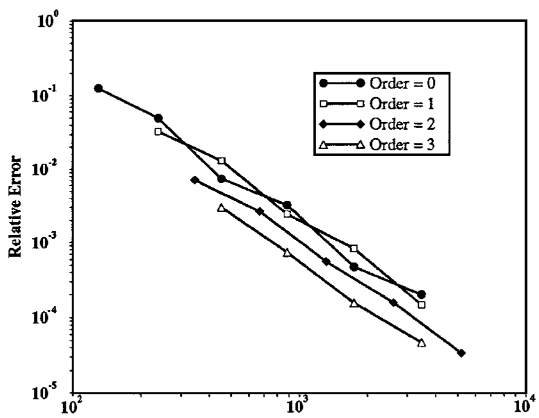

Fig. 1 Relative error in states and costates vs number of parameters, old way.

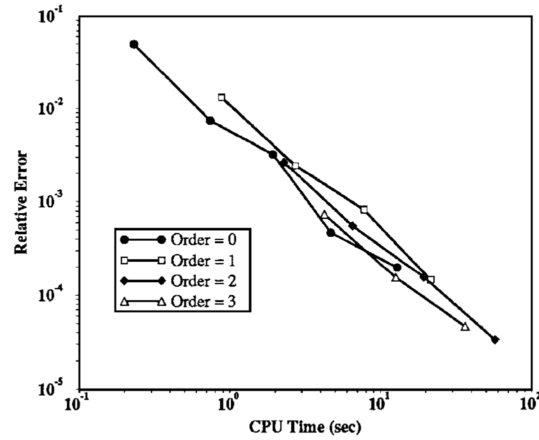

Fig. 2 Relative error in states and costates vs CPU time, old way.

functional can be specified just before the intermediate time t_1^-, just after the intermediate time t_1^+, or at some combination of those times. The formulation generalizes for the case of many intermediate times, resulting in a multiphase problem.

Of interest here is when the control inequality constraints are different in the two phases (though again this generalizes when there are more phases), such that

$$G(x, u) = \begin{cases} g_1(x, u) \le 0, & g_1 \in R^{n_{g1}}, \quad t \in [0, t_1) \\ g_2(x, u) \le 0, & g_2 \in R^{n_{g2}}, \quad t \in (t_1, t_f] \end{cases} \tag{13}$$

where either n_{g_1} or n_{g_2} could be zero. In general, these constraints could be enforced through use of slack variables as before.

The Hamiltonian is now distinct within each phase:

$$H_1 \equiv L + \lambda^T f + \mu^T(g_1 + K)$$

$$H_2 \equiv L + \lambda^T f + \mu^T(g_2 + K)$$

where the subscript denotes the phase number. Now, the optimal solution can have discontinuities in the Hamiltonian of the following form, after generalizing the formulation to include n_s switching times:

$$H_{i+1}\left(t_i^+\right) = H_i\left(t_i^-\right) + \frac{\partial \Phi}{\partial t_i}, \qquad i \in [1, n_s - 1] \tag{14}$$

Thus, if the intermediate times are neither penalized nor specified, which is the case when all that changes across a phase boundary is the nature of the control constraints, then the Hamiltonian is also continuous, despite having a different functional form on both sides of that time.

Just as the actual equations are the same as for the single-phase problem in most ways, the finite element equations are essentially the same as before. The form of the equations is the same within each phase, taking care that the appropriate constraints are used for each phase, with the jump condition on the Hamiltonian [Eq. (14)], providing the extra needed equations to find the unknown phase-boundary times.

V. Implementation and Results

To demonstrate how GENCODE handles problems with control constraints formulated as multiphase problems, the problem in the preceding section was formulated as a multiphase problem. To do this, extra boundary conditions were from enforcing continuity in the states at the three intermediate times. The switching structure of the control constraints was determined by first running the code on the problem set up with slack variables.

The code was then run for three elements per phase for various shape function orders. Figures 3 and 4 show the errors as functions of CPU time and number of parameters, with the best such plot from the old way of handling a control constraint also shown. For the parameters plot, this is a third-order solution, whereas for the CPU time plot it is zeroth order. Each plot uses the solution with

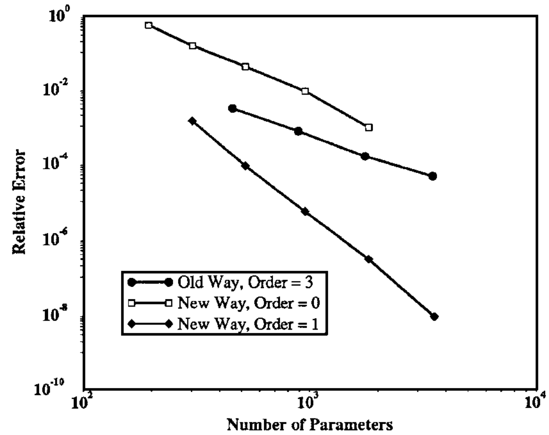

Fig. 3 Relative error in states and costates vs number of parameters, comparison of old and new ways.

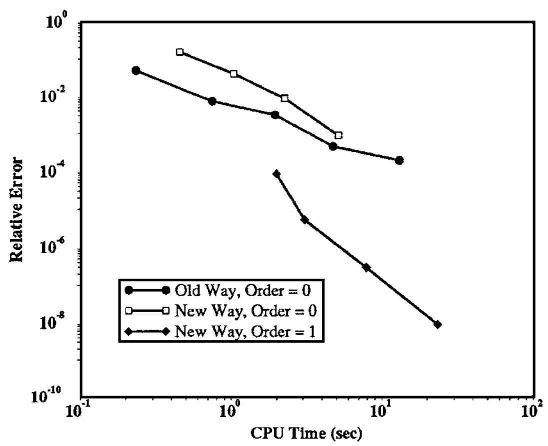

Fig. 4 Relative error in states and costates vs CPU time, comparison of old and new ways.

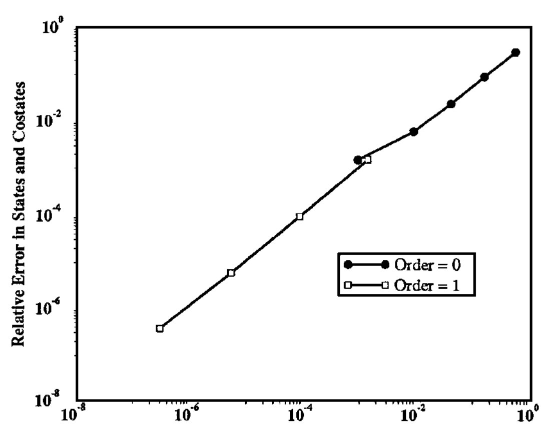

Fig. 5 Switching time accuracy vs overall accuracy.

three elements per phase as a starting point, with each data point corresponding to doubling the number of elements per phase.

Only the zeroth- and first-order shape function plots are shown for the new way, because the second-order shape function solution had zero error all along the time history, even for only two elements per phase. Although the CPU time necessary to implement the new method was higher for the zeroth-order shape functions than for the old way (reflecting the effect of the significantly increased size of the Jacobian), the multiple phase first-order solution is clearly far superior to the best results from the single-phase methodology. Thus, the key to solving this problem accurately and efficiently is clearly in the switching points, as seen in Fig. 5. Even for the three-element-per-phase solution with first-order shape functions, the switching times were all within 0.2% of the optimal value, moving an order of magnitude closer with each refinement. On the other hand, even with 192 elements obtained by uniform refinement and even with the highest-order shape functions, calculated switching times could get no closer than 0.2%.

VI. Conclusions

Two different methods have been illustrated to solve control-constrained optimal control problems using higher-order finite element formulations. In the first method, the optimal control problem is treated as if it had only one phase, with slack variables being used to enforce the control constraints. In the second method, the switching structure of the control constraints is specified ahead of time, and the problem is treated as a multiphase problem with unknown switching times.

The latter method proved to be much more accurate with the switching times being calculated explicitly, with a node point always being at a switching time. Results for a textbook problem show order of magnitude increases in accuracy using the second method, not only for a given number of parameters, but also for a given amount of CPU time.

References

[1] Warner, M. S., and Hodges, D. H.,"Solving Optimal Control Problems Using *hp*-Version Finite Elements in Time," *Journal of Guidance, Control, and Dynamics* (submitted for publication).

[2] Bryson, A. E., and Ho, Y.-C., *Applied Optimal Control—Optimization, Estimation, and Control*, Hemisphere, New York, 1975, Chap. 2.

[3] Hodges, D. H., and Hou, L.-J., "Shape Functions for Mixed *p*-Version Finite Elements in the Time Domain," *Journal of Sound and Vibration*, Vol. 145, No. 2, 1991, pp. 169–178.

[4] Bless, R. R., and Hodges, D. H., "Finite Element Solution of Optimal Control Problems with State-Control Inequality Constraints," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 4, 1992, pp. 1029–1032.

Application of the Nyquist Stability Criterion on the Nichols Chart

Joseph C. Slater*
Wright State University, Dayton, Ohio 45435-0001

Introduction

THE classical implementation of the Nyquist stability criterion is well known. It is most often used as a stability check when eigenvalue analysis of a closed-loop system is not considered to be reliable, or when uncertainty exists in the plant. All that is necessary for determining stability of the closed-loop system is to know the number of poles on and to the right of the imaginary axis of the open-loop system as well as the frequency-response function. Although valuable as a stability check, it is difficult to use as a tool because loop shaping goals guide the control design process. In addition, the Nyquist stability criterion is generally defined on the open-loop Nyquist plot. This is not directly useful for design techniques based on the Nichols chart such as Quantitative Feedback Theory. Cohen et al.[1] have previously extended the Nyquist stability criterion to the Nichols chart. Here a simpler formulation is presented that provides the designer with additional insight useful in determining loop shaping goals for stable controller design. The formulation assists not only in stability analysis, but in prescribing design objectives for achieving stabilizing continuous control designs.

Received Jan. 14, 1998; revision received July 6, 1998; accepted for publication Oct. 6, 1998. Copyright © 1998 by Joseph C. Slater. Published by the American Institute of Aeronautics and Astronautics, Inc., with permission.

*Assistant Professor, Department of Mechanical and Materials Engineering. Member AIAA.