JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 15, No. 2, March-April 1992

Optimal Control of Two-Time-Scale Systems with State-Variable Inequality Constraints

A. J. Calise* and J. E. Corban†
Georgia Institute of Technology, Atlanta, Georgia 30332

The established necessary conditions for optimality in nonlinear control problems that involve state-variable inequality constraints are applied to a class of singularly perturbed systems. The distinguishing feature of this class of two-time-scale systems is a transformation of the state-variable inequality constraint, present in the full-order problem, to a constraint involving states and controls in the reduced problem. It is shown that, when a state constraint is active in the reduced problem, the boundary-layer problem can be of finite time in the stretched time scale. Thus, the usual requirement for asymptotic stability of the boundary-layer system is not applicable and cannot be used to construct approximate boundary-layer solutions. Several alternative solution methods are explored and illustrated with simple examples.

I. Introduction

STATE-VARIABLE inequality constraints are commonly encountered in the study of dynamic systems. The study of rigid body aircraft dynamics and control is certainly no exception. For instance, a maximum allowable value of dynamic pressure is usually prescribed for aircraft with supersonic capability. This limit is required to insure that the vehicle's structural integrity is maintained. Given a typical state-space description of the vehicle dynamics, this limit constitutes an inequality constraint on vehicle state.

State inequality constraints have been studied extensively by researchers in the field of optimal control. First-order necessary conditions for optimality when general functions of state are constrained have been obtained.[1-3] However, the construction of solutions via this set of conditions proves difficult. Most practitioners seeking an open-loop control solution rely on direct approaches to optimization that employ penalty functions for satisfaction of state inequality constraints.[4] As a rule, algorithms employing such methods are computationally intense and slow to converge. Consequently, they are not well suited for real-time implementation.

As discussed in the literature, the use of singular perturbation techniques in the study of aircraft trajectory optimization can, through order reduction, lead to both open- and closed-loop solutions that are computationally efficient. These methods can sometimes be used to circumvent difficulties associated with enforcing a state inequality constraint as well.[5] As an example, consider the minimum time intercept problem of Ref. 6 that employs a model of the F-8 aircraft. A near-optimal feedback solution is obtained via singular perturbation theory that includes consideration of an inequality constraint on dynamic pressure. In the zero-order reduced solution, algebraic constraints are obtained when the perturbation parameter ϵ, which premultiplies the so-called fast dynamic equations, is set to zero. These constraints can be used to eliminate the fast states (in this case altitude and flight-path angle) from the reduced problem. One can choose, however, to retain one or more of the fast states and to eliminate instead a corresponding number of the original control variables. In such a case, the retained fast state variables are treated as controls, and the original state constraint becomes a constraint involving both state and control. This transformation of the constraint function from one dependent only on state to one dependent on both state and control leads to considerable simplification when seeking a solution to the reduced problem. In subsequent analysis of boundary layers, altitude resumes its status as a state variable, and dynamic pressure once again becomes a function of state alone. However, because the reduced solution for the example F-8 aircraft does not lie on the constraint boundary during ascent, the inequality constraint on dynamic pressure does not have to be considered in the boundary-layer analysis. The problem of treating a pure state constraint is thus avoided. Of note is the fact that modern supersonic fighter aircraft (such as the F-15) do ride the dynamic pressure constraint boundary during the ascent leg of the minimum time to intercept path.

In addition to the example just cited, dynamic pressure bounds are encountered during fuel-optimal climb for supersonic transports,[7] for rocket-powered launch vehicles such as the U.S. Space Shuttle,[8] and for single-stage-to-orbit air-breathing launch vehicles.[9] If, in applying singular perturbation methods when seeking a feedback solution to any of these problems, the reduced solution climb path lies directly on the dynamic pressure constraint boundary for a portion of the flight, then it is necessary to consider boundary-layer transitions onto the constrained arc. This problem has not received attention in the literature.

In the absence of a state-variable inequality constraint (i.e., when it is inactive), the initial boundary-layer solution for the class of systems being considered is an infinite time process. A solution is sought that asymptotically approaches the reduced solution. However, it will be shown that, when a state constraint is active in the reduced solution, the boundary-layer problem can be of finite time in the stretched time variable. Thus, the usual requirement for asymptotic stability of the boundary-layer system is not applicable and cannot be used to construct an approximate boundary-layer solution.[10] The presence of an active state inequality constraint also introduces the possibility of discontinuous costate variables at the juncture between constrained and unconstrained arcs. A Valentine transformation can be used to convert the constrained problem to an equivalent unconstrained problem of increased dimension. Smoothness is regained in the process, but to little or no advantage when seeking a solution for real-time implementation. This point is discussed in greater depth later.

It is shown by example that one can analytically obtain a solution to a simple singularly perturbed example involving a

Presented as Paper 90-3450 at the AIAA Guidance, Navigation, and Control Conference, Portland, OR, Aug. 20-22, 1990; received Sept. 14, 1990; revision received April 30, 1991; accepted for publication May 13, 1991.

*Professor, School of Aerospace Engineering. Associate Fellow AIAA.

†Post Doctoral Researcher, School of Aerospace Engineering, and President, Guided Systems Technologies, 430 Tenth Street, NW, Suite N107, Atlanta, GA 30318. Member AIAA.

state inequality constraint by dealing with the possible costate discontinuities directly. In addition, a general procedure for constructing a near-optimal boundary-layer transition onto the constraint boundary is formulated for use when an exact analytic solution cannot be obtained. The resulting approximation proves quite satisfactory when compared with the optimal solution, at least for small perturbations away from the reduced solution.

This paper proceeds as follows. Section II provides a brief review of the first-order necessary conditions that have been obtained for state-variable inequality constrained problems in optimal control. Section III discusses the optimal control of singularly perturbed systems subject to state-variable inequality constraints in general and in particular examines the features of state inequality constrained boundary layers when the reduced solution lies on the constraint boundary. Section IV presents the solutions to three example problems. The first two are linear and simply illustrate the problem features discussed in earlier sections. A nonlinear example is then used to illustrate a general feedback strategy for the construction of near-optimal boundary-layer transitions onto a constrained arc. The approximate solution is then compared directly with the optimal solution. Section V completes the paper by providing some concluding remarks.

II. State-Variable Inequality Constrained Problems in Optimal Control

The introduction of a state-variable inequality constraint of the form

$$S(z,t) \leq 0 \tag{1}$$

can lead to considerable difficulty when attempting to obtain an optimal control solution to a general nonlinear problem. In Eq. (1), the variable z denotes the system's state vector and t denotes time. One approach to incorporating state-variable inequality constraints in the formulation of first-order necessary conditions for optimality consists of constructing successive total time derivatives of S until explicit dependence on the control u appears. If p time derivatives are required, then Eq. (1) is referred to as a pth order state-variable inequality constraint. The relation $S^p(z,u,t)=0$ is then adjoined to the Hamiltonian as a constraint to be enforced when $S=0$. In this instance, the superscript p denotes the pth total time derivative of S. This approach introduces additional "tangency" conditions that must be satisfied at points of entry onto and exit from a constrained arc. These conditions are given by

$$N(z,t) = \begin{bmatrix} S(z,t) \\ S^1(z,t) \\ \vdots \\ S^{p-1}(z,t) \end{bmatrix} = 0 \tag{2}$$

Equation (2) constitutes a set of interior boundary conditions that must be met at each juncture between a constrained and unconstrained arc. To satisfy these interior boundary conditions, one must allow for the possibility of discontinuities in the costate variables at the junctions.[1]

An alternative set of necessary conditions can be obtained by adjoining the constraint function in Eq. (1), rather than its pth derivative, to the Hamiltonian and then employing a separating hyperplane theorem.[2] The resulting necessary conditions prove simpler and "sharper" than those of Ref. 1, but the possibility for discontinuous costate variables remains. The relationship between these sets of necessary conditions is defined in Ref. 3, which also identifies the conditions that have to be added to those of Ref. 1 to fill the gap between them. Other approaches to treating such problems include employing a transformation technique in which a slack variable is used to transform the state constrained problem into an equivalent unconstrained problem of higher dimension and various direct numerical methods of solution.[4,11]

III. Optimization of Singularly Perturbed Systems Subject to State Inequality Constraints

Consider the singularly perturbed optimal control formulation:

$$\frac{dx}{dt} = f(x,y,u,t) \quad x(t_0)=x_0 \quad x\in R^n \quad u\in R^1 \tag{3}$$

$$\epsilon\frac{dy}{dt} = g(x,y,u,t) \quad y(t_0)=y_0 \quad y\in R^m \tag{4}$$

with an index of performance and scalar inequality constraint of the form

$$J = \phi[x(t_f),\ \epsilon y(t_f)] \qquad S(x,y)\leq 0 \tag{5}$$

where ϵ is a small parameter. It is further assumed that $g_u^T g_u > 0$, $S_y^T S_y > 0$ (the superscript T denotes transpose and the subscripts denote partial differentiation). Zero-order necessary conditions for optimality of the associated reduced and boundary-layer problems in the absence of state-variable inequality constraints are readily available.[12] In addition to basic smoothness assumptions on both f and g, sufficient conditions for the existence of a boundary-layer solution carry the requirement that there exist an isolated root of $g=0$ along the reduced solution to serve as an asymptotically stable equilibrium point for the boundary-layer dynamics. Furthermore, the initial condition on y must lie in the domain of attraction of this root at the initial time, which amounts to a controllability requirement in this context. Similar requirements exist at the final time.

The extension of the necessary conditions for state constrained problems discussed in the previous section to include the singular perturbation formulation in Eqs. (3-5) is as follows. Adjoining the constraint in Eq. (5) directly, the Hamiltonian is given by

$$H = \lambda_x f + \lambda_y g + \nu S \tag{6}$$

where $\nu(t)\geq 0$ when $S=0$, and $\nu(t)=0$ when $S<0$. The additional first-order necessary conditions for optimality are

$$\frac{d\lambda_x}{dt} = -H_x \quad \lambda_x(t_f)=\phi_x(t_f) \quad H_u = 0 \tag{7}$$

$$\epsilon\frac{d\lambda_y}{dt} = -H_y \qquad \lambda_y(t_f)=\phi_y(t_f)/\epsilon \tag{8}$$

with the jump conditions

$$\lambda_x(t_i^+) = \lambda_x(t_i^-) - \sigma(t_i,\ \epsilon)S_x(t_i) \qquad H(t_i^+) = H(t_i^-) \tag{9}$$

$$\lambda_y(t_i^+) = \lambda_y(t_i^-) - \sigma(t_i,\ \epsilon)S_y(t_i)/\epsilon \tag{10}$$

where $\sigma(t_i,\ \epsilon)\geq 0$. The t_i represent the times at which a constrained and unconstrained arc meet. In Eqs. (9) and (10), the scalar multiplier σ is $o(\epsilon)$. Note that, in the limit as ϵ goes to zero, there are no jumps in the slow costate variables. In contrast, the fast costates can exhibit finite jumps proportional to S_y. To arrive at the condition in Eq. (10), it is important to note that the interpretation for λ_y in a singular perturbation formulation is different from that of a nonsingular formulation, namely,

$$\lambda_y(t_0) = \left[\frac{\partial J}{\partial y}(t_0)\right]/\epsilon \tag{11}$$

The reduced problem and associated necessary conditions are obtained by setting $\epsilon = 0$ in Eqs. (3-10). The fast states act as control variables as a consequence of setting $\epsilon = 0$ in Eq. (8). Hence, the constraint S can be viewed as a function of both states and controls, and the possibility of jumps in the slow costates is eliminated. The reduced solution necessary conditions are given by

$$\frac{dx}{dt} = f(x,y,u,t) \quad x(t_0) = x_0 \quad g(x,y,u,t) = 0 \tag{12}$$

$$H = \lambda_x f + \lambda_y g + \nu S \tag{13}$$

$$\frac{d\lambda_x}{dt} = -H_x \quad \lambda_x(t_f) = \phi_x(t_f) \quad H_y = H_u = 0 \tag{14}$$

where $\nu(t) \geq 0$ when $S = 0$, and $\nu(t) = 0$ when $S < 0$.

The zero-order initial boundary-layer solution associated with Eqs. (3-5) is obtained by introducing the time stretching transformation $\tau = (t - t_0)/\epsilon$ and again setting ϵ to zero. On this time scale, the slow states and costates are essentially frozen at their reduced solution values. The initial boundary-layer necessary conditions are given by

$$\frac{dx}{d\tau} = 0 \Rightarrow x(\tau) = x^0 \qquad \frac{d\lambda_x}{d\tau} = 0 \Rightarrow \lambda_x(\tau) = \lambda_x^0 \tag{15}$$

$$\frac{dy}{d\tau} = g(x^0,y,u,\tau) \qquad \frac{d\lambda_y}{d\tau} = -H_y \tag{16}$$

$$H_{\mathrm{BL}} = \lambda_x^0 f + \lambda_y g + \nu S \qquad \nu(t) \geq 0 \tag{17}$$

$$H_{\mathrm{BL}_u} = 0 \qquad S(x,y) \leq 0 \tag{18}$$

where $[\,]^0$ denotes the reduced solution evaluated at $t = t_0$. When the reduced solution trajectory does not lie on the constraint boundary, the boundary-layer solution, if it exists, is expected to be an infinite time process and to asymptotically approach the reduced solution. A local stability test, accomplished by linearizing the boundary-layer necessary conditions about the reduced solution, is usually employed to test for the existence of a solution. Since the linearized dynamics [with the control eliminated using Eq. (18)] in general have eigenvalues symmetrically arranged about the origin, the sufficiency condition is reduced to the requirement that none of the resulting eigenvalues lie on the imaginary axis.[10] When this condition holds true, a feedback solution for small perturbations away from the reduced solution can be constructed by selecting the initial conditions on the perturbations in λ_y to lie in the subspace spanned by the eigenvectors associated with the stable eigenvalues.

When the reduced solution lies on the state constraint boundary, it may no longer act as an equilibrium point of the boundary-layer dynamics. This feature is evident upon examination of Eqs. (14) and (16). A critical requirement is that $d\lambda_y/d\tau$ approach zero as the reduced solution is approached. In the unconstrained case, this is a natural consequence of the fact that $H_y = 0$ is satisfied with $\nu(t_0) = 0$ by the reduced solution. However, when the reduced solution lies on a constraint boundary, the situation is considerably more complicated. It may happen that the state constraint is encountered before the end of the boundary-layer transition, in which case the reduced solution would serve as an equilibrium point in a reduced state space associated with riding the constraint boundary. If on the other hand the state constraint is not encountered until the end of the boundary-layer transition, then the reduced solution is very likely not an equilibrium point. In particular, this occurs when $S^i(x^0,y) = 0$, $i = 0, \ldots, p-1$, has an isolated root $y(x^0)$, which is often the case in aircraft flight mechanics problems. As a consequence, $d\lambda_y/d\tau$ does not approach zero at the end of the boundary-layer transition, but instead "jumps" to zero due to the jumps that occur in λ_y and νS_y. This further implies that the boundary-layer arc is of finite time in the τ time scale. In this case, the traditional asymptotic stability analysis and method of matched asymptotic expansions cannot be used in constructing an approximate solution.

In this paper, several alternate means for constructing such an approximation are presented. In particular, the costate jumps that can occur and the boundary-layer final time are used as free parameters in order to satisfy continuity conditions in the state variables at the end of the boundary-layer response. Note that the continuity condition on H in Eq. (9) is guaranteed if the problem is regular. In this case both the states and the control are continuous, hence $g = 0$ at the end of the boundary-layer transition. Though not addressed in this paper, a similar finite time phenomenon has been observed when considering boundary-layer transitions onto singular arcs.[13]

As mentioned earlier, one approach to treating these problems that avoids the difficulties associated with costate discontinuities consists of using a Valentine transformation to convert the constrained problem to an equivalent unconstrained one of higher dimension.[11] The inequality constraint in Eq. (5) is converted to an equality constraint by the introduction of a slack variable $\alpha(t)$ as follows:

$$S(x,y,t) + \tfrac{1}{2}\alpha^2(t) = 0 \tag{19}$$

Differentiating Eq. (19) p times with respect to time, the following set of equations is obtained:

$$S^1 + \alpha\alpha_1/\epsilon = S_x f + S_y g/\epsilon$$

$$+ S_t + \alpha\alpha_1/\epsilon = 0 \qquad \epsilon\frac{d\alpha}{dt} = \alpha_1$$

$$S^2(x,y,t) + \alpha_1^2/\epsilon^2 + \alpha\alpha_2/\epsilon^2 = 0 \qquad \epsilon\frac{d\alpha_1}{dt} = \alpha_2$$

$$S^3(x,y,t) + 3\alpha_1\alpha_2/\epsilon^3 + \alpha\alpha_3/\epsilon^3 = 0 \qquad \epsilon\frac{d\alpha_2}{dt} = \alpha_3$$

$$\vdots$$

$$S^p(x,y,u,t) + \{\text{terms involving } \alpha_{p-1}, \ldots, \alpha_1, \epsilon\}$$

$$+ \alpha\alpha_p/\epsilon^p = 0 \qquad \epsilon\frac{d\alpha_{p-1}}{dt} = \alpha_p \tag{20}$$

where the presence of ϵ in Eq. (20) is representative of the fact that α and all of the α_i must be fast variables to maintain the equality in Eq. (19). Using the pth equation in Eq. (20), one can solve for the control u to obtain

$$u = G(x,y,\alpha_{p-1}, \ldots, \alpha_1, \alpha\alpha_p, \epsilon, t) \tag{21}$$

Using Eq. (21) and treating $\alpha, \ldots, \alpha_{p-1}$ as additional fast state variables, the following unconstrained problem with α_p as the new control variable is obtained

$$\frac{dx}{dt} = f(x,y,G,t)$$

$$\epsilon\frac{dy}{dt} = g(x,y,G,t)$$

$$\epsilon\frac{d\alpha}{dt} = \alpha_1$$

$$\epsilon\frac{d\alpha_1}{dt} = \alpha_2$$

$$\vdots$$

$$\epsilon\frac{d\alpha_{p-1}}{dt} = \alpha_p \tag{22}$$

with the index of performance again given by

$$J = \phi[x(t_f),\ \epsilon y(t_f)] \tag{23}$$

Note that the dimension of Eq. (22) can be reduced using Eq. (19) and the first $p-1$ equalities in Eq. (20) to eliminate some of the original fast state variables. The initial conditions $\alpha(t_0), \ldots, \alpha_{p-1}(t_0)$ are chosen to satisfy Eq. (19) and the first $p-1$ equations in the set (20):

$$\begin{aligned} \alpha(t_0) &= \pm\{-2S[x(t_0),y(t_0),t_0]\}^{1/2} \\ \alpha_1(t_0) &= -\epsilon S^1[x(t_0),y(t_0),t_0]/\alpha(t_0) \\ \alpha_2(t_0) &= -\epsilon^2\{S^2[x(t_0),y(t_0),t_0] + \alpha_1^2(t_0)\}/\alpha(t_0) \\ &\vdots \end{aligned} \tag{24}$$

With this choice of boundary conditions, Eqs. (19) and (20) are satisfied for all time on the interval of interest for any control function α_p; that is, any function α_p will produce an admissible trajectory. Thus the original constrained problem of Eqs. (3-5) is replaced by the transformed unconstrained problem of Eqs. (22-24). In general, however, this technique simply trades one difficulty for another. Constrained arcs that occur in the reduced solution of the original problem correspond to singular arcs in the the transformed variables.[11] Linearization of the boundary-layer dynamics and exploitation of asymptotic stability properties again is not possible since the first variation with respect to α_p vanishes along the reduced solution, and α_p is in general discontinuous at the juncture of singular and nonsingular arcs.

IV. Examples

Several simple examples are now presented. The first illustrates the use of the Valentine transformation and the finite time nature of the boundary-layer solutions. The second example is solved without resorting to the Valentine transformation and illustrates the complete analytic solution to a singularly perturbed problem of optimal control involving a state-variable inequality constraint. A third nonlinear example is introduced for which a complete analytic solution is unavailable. In this case, a feedback strategy is employed to construct an approximation to the boundary-layer transition onto the constraint arc. This strategy is based on a Taylor's series expansion to first order about a nonequilibrium point. The costate jump that can occur and the boundary-layer final time are used as free parameters to satisfy continuity conditions in the state variables at the end of the boundary-layer response. The optimal solution of the third example was generated numerically for direct comparison to this feedback approximation.

Example 1

Consider the singularly perturbed dynamical system,

$$\frac{dx}{dt} = y - u^2 \qquad \epsilon\frac{dy}{dt} = u \qquad S = y - 1 \le 0 \tag{25}$$

with the initial conditions specified as $x(0) = y(0) = 0$. The final value $x(t_f) > 0$ is specified and the objective is to minimize the time required to reach this specified end condition. The state constraint in this case is first order. Introduction of a slack variable and transformation results in the equivalent unconstrained dynamics:

$$\frac{dx}{dt} = 1 - \alpha^2/2 - (\alpha\alpha_1)^2 \qquad x(0) = 0 \tag{26}$$

$$\epsilon\frac{d\alpha}{dt} = \alpha_1 \qquad \alpha(0) = (2)^{1/2} \tag{27}$$

where y has been eliminated using Eq. (19).

Reduced Solution

The reduced solution Hamiltonian is given by

$$H = \lambda_x[1 - \alpha^2/2 - (\alpha\alpha_1)^2] + \lambda_\alpha\alpha_1 + 1 = 0 \tag{28}$$

where $\alpha_1^0 = 0$ as a consequence of setting $\epsilon = 0$ in Eq. (27). Evaluation of first-order necessary conditions for optimality results in the following:

$$H = 0 \quad \text{and} \quad H_\alpha = 0 \quad \Rightarrow \quad \lambda_x^0 = 1, \quad \alpha^0 = 0 \tag{29}$$

$$H_{\alpha_1} = 0 \quad \Rightarrow \quad \lambda_\alpha^0 = 0 \tag{30}$$

from which it is evident that

$$x^0(t) = t \quad y^0(t) = 1 \quad u^0 = 0 \tag{31}$$

Boundary-Layer Solution

Introducing the time scale transformation $\tau = t/\epsilon$ and again setting ϵ to zero, the boundary-layer dynamics become

$$\frac{d\alpha}{d\tau} = \alpha_1 \qquad \alpha(0) = (2)^{1/2} \tag{32}$$

The boundary-layer Hamiltonian is given by

$$H_{BL} = \alpha^2/2 + (\alpha\alpha_1)^2 + \lambda_\alpha\alpha_1 = 0 \tag{33}$$

and the associated necessary conditions are

$$\frac{d\lambda_\alpha}{d\tau} = -\alpha(1 + 2\alpha_1^2) \tag{34}$$

$$H_{\alpha_1} = 0 \quad \Rightarrow \quad \alpha_1 = -\lambda_\alpha/2\alpha^2 \tag{35}$$

$$H = 0 \quad \Rightarrow \quad \lambda_\alpha = -(2)^{1/2}\alpha^2 \tag{36}$$

Substituting Eq. (36) into Eq. (35) yields

$$\alpha_1(\tau) = -1/(2)^{1/2} \ne \alpha_1^0 \quad \alpha(\tau) = (2)^{1/2} - \tau/(2)^{1/2} \tag{37}$$

When the boundary-layer trajectory reaches the constraint boundary, $\alpha(\tau_f) = 0$ and Eq. (37) yields $\tau_f = 2$. Transforming back, we obtain a feedback solution for u in terms of y,

$$u = (1 - y)^{1/2} \tag{38}$$

Integration of Eq. (25) on the τ time scale using Eq. (38) gives

$$y(\tau) = 1 - (1 - \tau/2)^2 \tag{39}$$

Note that while α_1 is discontinuous at $\tau = \tau_f$, the original control is continuous.

Example 2

Now consider a third-order version of the preceding example:

$$\frac{dx}{dt} = y_1 - u^2 \quad \epsilon\frac{dy_1}{dt} = y_2 \quad \epsilon\frac{dy_2}{dt} = u \tag{40}$$

$$S = y_1 - 1 \le 0 \tag{41}$$

with initial conditions $x(t_0)$, $y_1(t_0) < 1$, and $y_2(t_0)$ given. As before, the final condition on x is prescribed and greater than $x(t_0)$, the final conditions on y_1 and y_2 are free, and the objective is to minimize the final time. In this case, the state constraint is of order two. In contrast to example 1, the solution to this problem is carried out in terms of the original variables.

Reduced Solution

The reduced solution Hamiltonian is given by

$$H = \lambda_x(y_1 - u^2) + \lambda_{y_1}y_2 + \lambda_{y_2}u + \nu(y_1 - 1) + 1 = 0 \quad (42)$$

where $y_2^0(t)$ and $u^0(t)$ are zero as a consequence of setting $\epsilon = 0$ in Eq. (41). Evaluation of first-order necessary conditions for optimality results in the following:

$$H = H_{y_1} = 0 \quad \Rightarrow \quad \nu^0(t) = 1 \quad y_1^0(t) = 1 \quad \lambda_x^0(t) = -1$$

$$H_{y_2} = 0 \quad \Rightarrow \quad \lambda_{y_1}^0(t) = 0$$

$$H_u = 0 \quad \Rightarrow \quad \lambda_{y_2}^0(t) = 0 \quad (43)$$

Thus all of the variables except x are constant, the control is maintained at zero, the initial conditions on y_1 and y_2 are not satisfied, y_1 is on the constraint boundary, and x is propagated to its specified final value according to the relation

$$x^0(t) = x(t_0) + t \quad (44)$$

Boundary-Layer Solution

Introducing the time scale transformation $\tau = t - t_0/\epsilon$ and again setting ϵ to zero, the boundary-layer dynamics become

$$\frac{dy_1}{d\tau} = y_2 \qquad y_1(\tau_f) = y_1^0(t_0) = 1 \quad (45)$$

$$\frac{dy_2}{d\tau} = u \qquad y_2(\tau_f) = y_2^0(t_0) = 0 \quad (46)$$

where the terminal conditions on the fast states are chosen to match their reduced solution values. The boundary-layer Hamiltonian is given by

$$H = \lambda_x^0(y_1 - u^2) + \lambda_{y_1}y_2 + \lambda_{y_2}u + \nu(y_1 - 1) + 1 = 0 \quad (47)$$

and the associated necessary conditions are

$$\frac{d\lambda_{y_1}}{d\tau} = 1 \qquad \lambda_{y_1}(\tau_f) = \Theta \text{ (i.e., free)} \quad (48)$$

$$\frac{d\lambda_{y_2}}{d\tau} = -\lambda_{y_1} \qquad \lambda_{y_2}(\tau_f) = \lambda_{y_2}^0(t_0) = 0 \quad (49)$$

$$H_u = -2\lambda_x^0 u + \lambda_{y_2} = 0 \quad \Rightarrow \quad u = -\lambda_{y_2}(\tau)/2 \quad (50)$$

Note that when the boundary-layer solution reaches the constraint boundary, it has also reached the reduced solution, i.e., it is not possible for the boundary-layer system to ride the constraint boundary before reaching the reduced solution. Thus, the initial boundary-layer problem can be viewed as an unconstrained two-point-boundary-value problem in which the terminal conditions must match the state and costate conditions just before entering onto the constrained arc. Integration of the state and costate dynamic equations following elimination of the control u using Eq. (50) and assuming $t_0 = 0$ yields

$$y_1(\tau) = \tau^4/48 + \lambda_{y_1}(0)\tau^3/12 - \lambda_{y_2}(0)\tau^2/4 + y_2(0)\tau + y_1(0) \quad (51)$$

$$y_2(\tau) = \tau^3/12 + \lambda_{y_1}(0)\tau^2/4 - \lambda_{y_2}(0)\tau/2 + y_2(0) \quad (52)$$

$$\lambda_{y_1}(\tau) = \lambda_{y_1}(0) + \tau \quad (53)$$

$$\lambda_{y_2}(\tau) = -\tau^2/2 - \lambda_{y_1}(0)\tau + \lambda_{y_2}(0) \quad (54)$$

Evaluating Eqs. (51-54) at τ_f results in four algebraic equations in four unknowns: $\lambda_{y_1}(0)$, $\lambda_{y_2}(0)$, $\lambda_{y_1}(\tau_f)$, and τ_f. The solution to this system is obtained as follows:

$$\lambda_{y_1}(0) = [y_2(0) - t_f^3/6]4/t_f^2 \quad (55)$$

$$\lambda_{y_2}(0) = \tau_f^2/2 + \lambda_{y_1}(0)\tau_f \quad (56)$$

$$\lambda_{y_1}(\tau_f) = \lambda_{y_1}(0) + \tau_f \quad (57)$$

where

$$\tau_f^4 + 48y_2(0)\tau_f + 144[y_1(0) - 1] = 0 \quad (58)$$

An analytic solution to Eq. (58) is available using standard handbook methods and yields one real positive solution for τ_f when $y_1(0) < 1.0$. For the particular choice of initial conditions of $x(t_0) = y_1(t_0) = y_2(t_0) = 0$, the solution is given by

$$\tau_f = 2(3)^{1/2} = 3.464$$

$$\lambda_{y_1}(0) = -2\tau_f/3 = -2.309 \qquad \lambda_{y_2}(0) = -\tau_f^2/6 = -2.000$$

The costate λ_{y_1} is discontinuous at the juncture between the unconstrained (boundary layer) and constrained (reduced solution) arcs as evidenced by its finite value at τ_f:

$$\lambda_{y_1}(\tau_f) = 2(3)^{1/2}/3 = 1.155$$

In essence, the time τ_f corresponds to a juncture time t_i in Eq. (10). Note the following.

1) The control solution in Eq. (50) can be put in feedback form by solving the boundary-layer problem for an arbitrary initial condition and evaluating $\lambda_{y_2}(\tau)$ at $\tau = 0$.

2) The control solutions in Eqs. (38) and (50) are exact in the sense that they satisfy the necessary conditions for their corresponding full-order problems. Thus all of the higher order correction terms are zero. This is a consequence of the fact that both the dynamics and the constraints in these examples are independent of the slow state variable.[14]

3) The full-order solution for the costate variables obtained using the necessary conditions of Ref. 2 agree exactly with the reduced solution along a constrained arc. If the necessary conditions of Ref. 1 are used to solve the corresponding full-order problems for these examples, then the fast costate variables are not constant along a constrained arc. Thus, the reduced solution does not even approximate the full-order solutions for the costate variables obtained using these conditions. However, the reduced solutions for the state and control time histories are valid approximations.

Solutions to the system of Eqs. (55-58) for a number of different initial conditions are presented in Table 1. Representative state, costate, and control histories are presented in Figs. 1-4. Note that x simply propagates to its specified final value according to Eq. (44) after the solution for y_1 reaches the constraint boundary. The x trajectories achieve large negative values initially for cases 4 and 5 and are not shown. Cases in which $x(t_f)$ is reached before y_1 reaches the constraint boundary indicate that the singular perturbation approximation is inappropriate. In general, for $\epsilon = 1$ it is possible to nondimensionalize the dynamics so that $\epsilon = 1/x(t_f)$.

Example 3

Consider the addition of a nonlinear term in the dynamic system of example 2 as follows:

$$\frac{dx}{dt} = y_1 - u^2 \quad \epsilon\frac{dy_1}{dt} = y_2 + y_1y_2 \quad \epsilon\frac{dy_2}{dt} = u \quad (59)$$

$$S = y_1 - 1 \le 0 \quad (60)$$

As in example 2, the state constraint is of order two.

Reduced Solution

The reduced solution Hamiltonian is given by

$$H = \lambda_x(y_1 - u^2) + \lambda_{y_1}(y_2 + y_1y_2) + \lambda_{y_2}u + \nu S + 1 = 0 \quad (61)$$

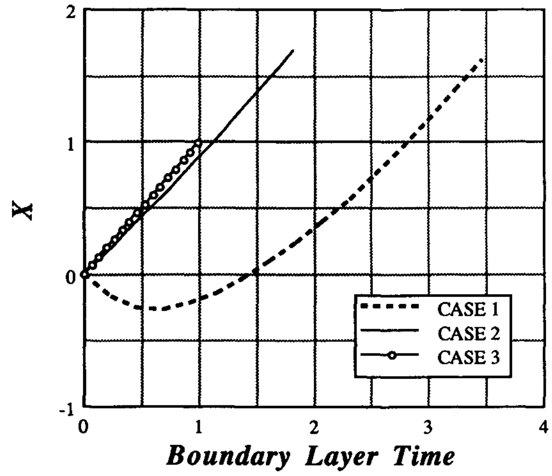

Fig. 1 Example 2: x vs τ for several different sets of initial conditions (see Table 1).

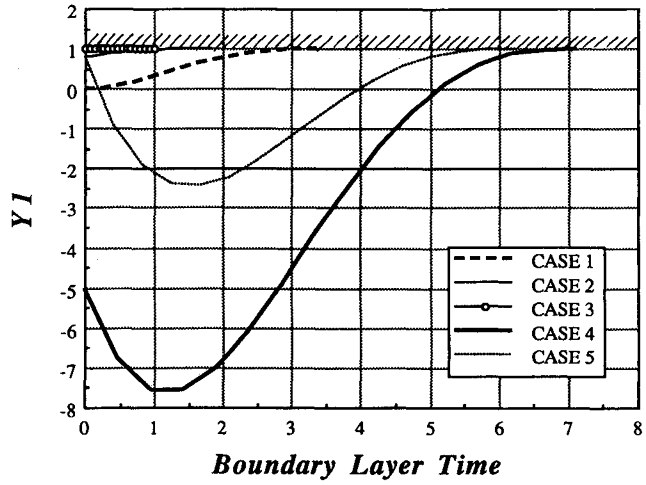

Fig. 2 Example 2: y_1 vs τ for several different sets of initial conditions (see Table 1).

where $y_2^0(t)$ and $u^0(t)$ are again zero as a consequence of setting $\epsilon = 0$ in Eq. (59). The reduced solution is identical to that of example 2. All of the variables except x are constant and the control is maintained at zero.

Boundary-Layer Solution

Introducing the time scale transformation $\tau = t - t_0/\epsilon$ and again setting ϵ to zero, the boundary-layer dynamics become

$$\frac{dy_1}{d\tau} = y_2 + y_1 y_2 \qquad y_1(\tau_f) = 1 \tag{62}$$

$$\frac{dy_2}{d\tau} = u \qquad y_2(\tau_f) = 0 \tag{63}$$

where the terminal conditions on the fast states are chosen to match their reduced solution values. The boundary-layer Hamiltonian is given by

$$H = \lambda_x^0(y_1 - u^2) + \lambda_{y_1}(y_2 + y_1 y_2) + \lambda_{y_2} u + \nu S + 1 = 0 \tag{64}$$

and the associated necessary conditions are

$$\frac{d\lambda_{y_1}}{d\tau} = 1 - \lambda_{y_1} y_2 \qquad \lambda_{y_1}(\tau_f) = \Theta \text{ (i.e., free)} \tag{65}$$

$$\frac{d\lambda_{y_2}}{d\tau} = -\lambda_{y_1}(1 + y_1) \qquad \lambda_{y_2}(\tau_f) = y_2^0 = 0 \tag{66}$$

Table 1 Representative solutions of Eqs. (55-58) in example 2

Case	$y_1(0)$	$y_2(0)$	$\lambda_{y_1}(0)$	$\lambda_{y_2}(0)$	$\lambda_{y_1}(\tau_f)$	τ_f
1	0.00	0.00	−2.309	−2.000	1.155	3.464
2	0.80	0.20	−0.982	−0.121	0.849	1.830
3	0.99	0.01	−0.620	−0.123	0.371	0.991
4	−5.00	−5.00	−5.143	−11.260	1.980	7.122
5	0.80	−5.00	−4.681	−9.717	1.573	6.254

$$H_u = -2\lambda_x^0 u + \lambda_{y_2} = 0 \quad \Rightarrow \quad u = -\lambda_{y_2}(\tau)/2 \tag{67}$$

Because an analytical solution for the state and costate dynamic equations (following elimination of the control) could not be found, we must resort to numerical methods of solution or seek approximations to the optimal solution.

As discussed earlier, the reduced solution does not act as an equilibrium point for the boundary-layer system. However, one can write a Taylor's series expansion for the boundary-layer system about a nonequilibrium point. We choose as such a point the state and costate conditions just before the juncture of the unconstrained boundary-layer trajectory and the reduced solution that lies on the constraint boundary. The corresponding time is denoted as τ_f, which is representative of the time t_i^- in Eq. (10). All of the conditions at this point are known except for the value of $\lambda_{y_1}(\tau_f)$, which is free to jump from a finite value to its reduced solution value of zero at the

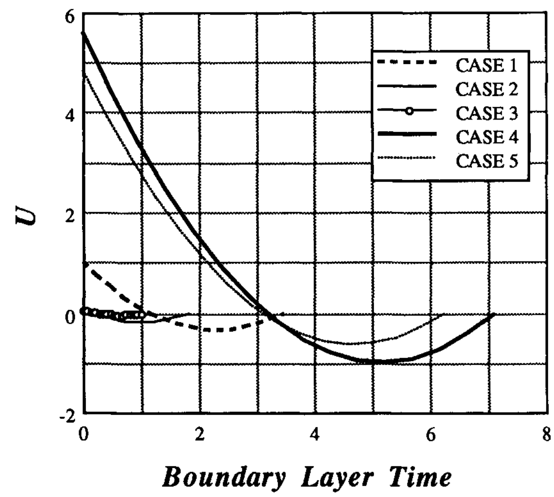

Fig. 3 Example 2: u vs τ for several different sets of initial conditions (see Table 1).

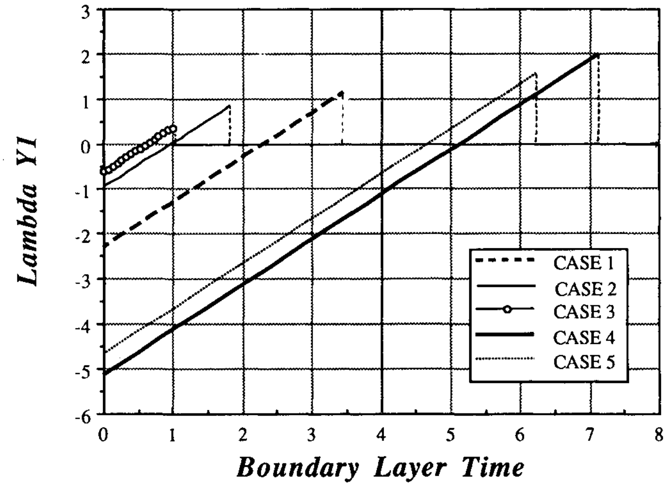

Fig. 4 Example 2: λ_{y_1} vs τ for several different sets of initial conditions (see Table 1).

end of the boundary-layer transition. To first order, the perturbation dynamics are described by

$$\frac{d\delta z}{d\tau} \cong f[z(\tau_f)] + \left.\frac{\partial f}{\partial z}\right|_{z(\tau_f)} \delta z \tag{68}$$

where the vector z is composed of the boundary-layer states and costates and the vector f represents the right-hand sides of their corresponding differential equations given by Eqs. (62), (63), (65), and (66). The perturbations are defined as follows:

$$\delta y_1 = y_1(\tau) - 1.0 \tag{69}$$

$$\delta y_2 = y_2(\tau) \tag{70}$$

$$\delta\lambda_{y_1} = \lambda_{y_1}(\tau) - \Theta \tag{71}$$

$$\delta\lambda_{y_2} = \lambda_{y_2}(\tau) \tag{72}$$

The relation (68) is of the general form

$$\frac{d\delta z}{d\tau} = A\,\delta z + B \tag{73}$$

where

$$A = \begin{vmatrix} 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & -1/2 \\ 0 & -\theta & 0 & 0 \\ -\theta & 0 & -2 & 0 \end{vmatrix} \tag{74}$$

$$B^T = [0\ 0\ 1\ -2\theta] \tag{75}$$

The solution to Eq. (73) is well known to be of the form

$$\delta z(\tau) = e^{A(\tau-\tau_0)}\,\delta z(\tau_0) + \int_{\tau_0}^{\tau} e^{A(\tau-\zeta)} B\, d\zeta \tag{76}$$

Note that the eigenvalues of A in this case are all zero. The state transition matrix is obtained analytically, and the response of the perturbation equations given an arbitrary set of initial conditions can be written as

$$\delta y_1 = \tau^4/12 + e\tau^3/3 - d\tau^2/2 + 2b\tau + a \tag{77}$$

$$\delta y_2 = \tau^3/6 + e\tau^2/2 - d\tau/2 + b \tag{78}$$

$$\delta\lambda_{y_1} = -\theta\tau^4/24 - e\theta\tau^3/6 + d\theta\tau^2/4 + (1-\theta b)\tau + c \tag{79}$$

$$\delta\lambda_{y_2} = -\tau^2 - 2e\tau + d \tag{80}$$

where for convenience we have defined

$$a \equiv \delta y_1(t_0) \qquad b \equiv y_2(t_0) \tag{81}$$

$$c \equiv \lambda_{y_1}(t_0) \qquad d \equiv \lambda_{y_2}(t_0) \qquad e \equiv \theta(1 + a/2) + c \tag{82}$$

We wish to drive these perturbations to zero at $\tau = \tau_f$. Thus, the left-hand sides of Eqs. (77–80) are set to zero. The initial conditions $y_1(t_0)$ and $y_2(t_0)$ are known. It is left to determine the corresponding final time τ_f (i.e., time t_i^-), θ, and the costate initial conditions. The complete analytic solution of the system of Eqs. (78–81) can be derived as follows. Solve Eq. (80) for the unknown value of d in terms of e and τ_f:

$$d = \tau_f^2 + 2e\tau_f \tag{83}$$

Substitute Eq. (83) into Eq. (77) to obtain

$$5\tau_f^4 + 8e\tau_f^3 - 24b\tau_f - 12a = 0 \tag{84}$$

Substitute Eq. (83) into Eq. (78) to obtain

$$2\tau_f^3 + 3e\tau_f^2 - 6b = 0 \tag{85}$$

Solve Eq. (85) for $e\tau_f^2$:

$$e\tau_f^2 = -2\tau_f^3/3 + 2b \tag{86}$$

Finally, substitute Eq. (86) into Eq. (84) to obtain

$$\tau_f^4 + 24b\tau_f + 36a = 0 \tag{87}$$

The complete analytic solution of the quartic polynomial (87) is available by standard handbook methods and yields four possible values of τ_f. Note that all allowable combinations of initial conditions fall into two general categories, those for which $a < 0$ and $b > 0$ and those for which $a < 0$ and $b < 0$. For either case, one can show that only one real positive solution for τ_f will occur.

The initial costate values determined in this way can be used to approximate the initial costates for the nonlinear boundary-layer system. By periodically updating these estimates while numerically integrating the nonlinear boundary-layer dynamics forward in time, one can generate an approximate solution to the boundary-layer problem. As for examples 1 and 2, the control solution Eq. (67) is exact in the sense that it satisfies the necessary conditions for the full-order problem. Thus all of the higher order correction terms are zero.

Table 2 presents the solution to Eqs. (78–81) at time τ_f for various combinations of initial conditions. The feedback strategy described above was employed to generate approximations to the optimal boundary-layer transitions. Representative state, costate, and control time histories are depicted in Figs. 5–8 for the case where $x(t_0) = y_1(t_0) = y_2(t_0) = 0$. For comparison, the optimal solution was generated numerically using a multiple shooting algorithm. The full-order dynamic system was considered and the associated unconstrained two-point-boundary value problem solved for the case when $t_0 = 0$, $x(t_0) = y_1(t_0) = y_2(t_0) = 0$, t_f free, $y_1(t_f) = 1$, $y_2(t_f) = 0$ and $x(t_f)$ free. The optimal state, costate, and control time histories are superimposed over the approximations presented in Figs. 5–8. In the optimal case, y_1 reaches the constraint boundary in 2.767 s. At that time x has a value of 1.117 and λ_{y_1} is 0.856. In contrast, at 2.8 s the approximated boundary-layer transition has not quite reached the reduced solution. The value of x is 1.139, y_1 is 0.999, y_2 is 0.010, and the time to go is estimated as 0.18 s. Using linear interpolation, the value of x at 2.767 s is 1.107; thus the x approximation is slightly less than the optimal value of x as it should be. An open-loop simulation of the nonlinear dynamic system starting with the approximation obtained via Eqs. (78–81) is also depicted in Figs. 5–8. These trajectories clearly illustrate the need to regularly update the approximate solution.

Closing Remark on the Examples

These examples perhaps give the misleading impression that when solving the boundary-layer problem, the number of equations always equals the number of unknowns. This is in fact not the case. If example 2 is extended to a third-order constraint by defining $\epsilon dy_2/dt = y_3$, $\epsilon dy_3/dt = u$, then two additional constraining equations arise (one for each new fast

Table 2 Representative solutions of the linearized boundary-layer system, Eqs. (78–81), of example 3 at time τ_f

Case	$y_1(0)$	$y_2(0)$	$\lambda_{y_1}(0)$	$\lambda_{y_2}(0)$	$\lambda_{y_1}(\tau_f)$	τ_f
1	0.00	0.00	−2.041	−2.000	0.817	2.449
2	0.80	0.20	−1.075	−0.264	0.687	1.144
3	0.99	0.01	−0.667	−0.089	0.268	0.668
4	−5.00	−5.00	−1.019	−13.500	1.469	5.427
5	0.80	−5.00	−4.828	−12.210	1.243	4.952

state and costate variable), but only one additional unknown is introduced, $\lambda_{y_3}(0)$. The question arises as to the significance of this outcome. If the necessary conditions of Ref. 2 are used to analyze this new problem, then the conclusion that results is that constrained arcs are not admissible. If the necessary conditions of Ref. 1 are used, then constrained arcs are admissible (they satisfy all of the first-order necessary conditions). However, if the additional conditions identified in Ref. 3 as filling

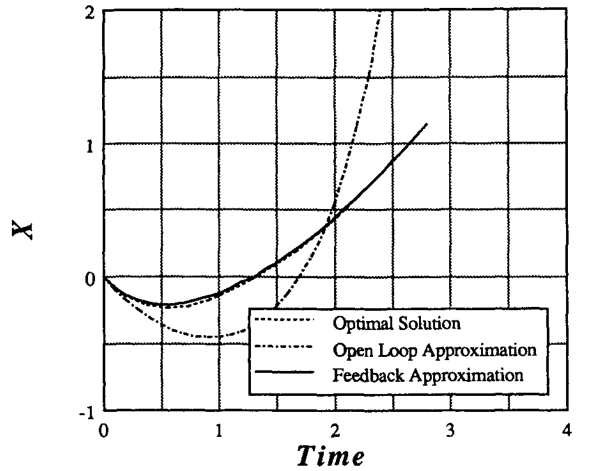

Fig. 5 Example 3: optimal and approximate x vs t when $x(0) = y_1(0) = y_2(0) = 0$.

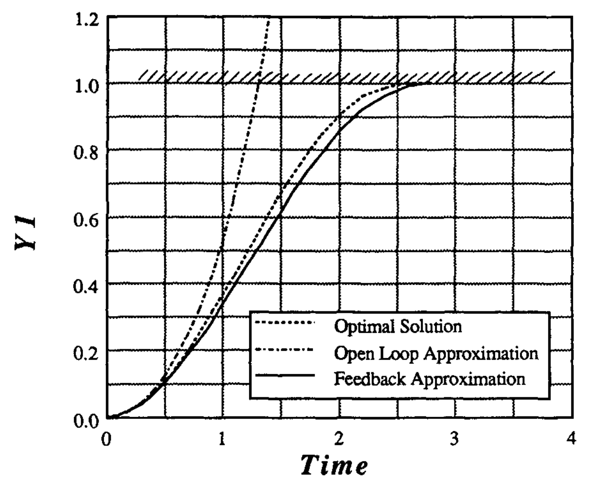

Fig. 6 Example 3: optimal and approximate y_1 vs t when $x(0) = y_1(0) = y_2(0) = 0$.

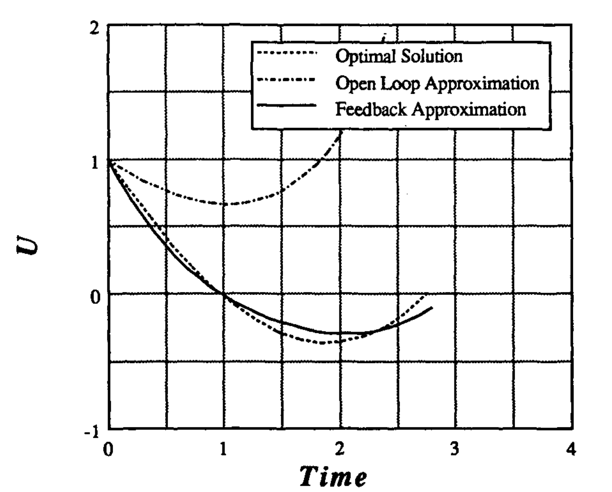

Fig. 7 Example 3: optimal and approximate u vs t when $x(0) = y_1(0) = y_2(0) = 0$.

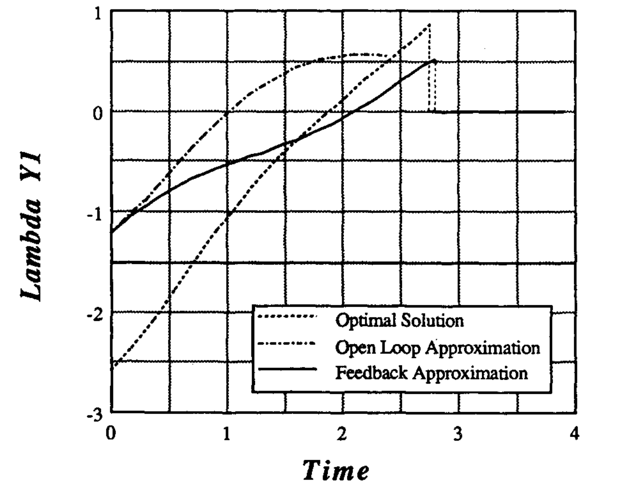

Fig. 8 Example 3: optimal and approximate λ_{y_1} vs t when $x(0) = y_1(0) = y_2(0) = 0$.

the gap between the necessary conditions of Refs. 1 and 2 are examined, then the equality condition in Eq. (19) in Ref. 3 corresponding to $k = 2$, $p = 3$ cannot be met. Thus it appears that having more equations than unknowns in the boundary-layer problem is connected with additional equality conditions that arise from filling the gap between the necessary conditions of Refs. 1 and 2. In this instance, it is not clear if the reduced solution (which rides that constraint) remains a valid $\mathcal{O}(\epsilon)$ approximation to the optimal solution [which does not ride the constraint but may be in an $\mathcal{O}(\epsilon)$ neighborhood of the constraint]. In any case, the existence of a solution to the constraining equations obtained from the boundary-layer necessary conditions appears to be necessary for singular perturbation analysis of two-time-scale optimal control problems with state-variable inequality constraints.

V. Conclusions

Singularly perturbed optimal control problems with state-variable inequality constraints can exhibit complex boundary-layer phenomenon. In particular, the boundary-layer transitions associated with such problems can be of finite time when the state constraint is first encountered at the end of the boundary-layer transition. Valentine's transformation can be used to avoid the problems associated with discontinuous costate time histories, but at the expense of introducing a singular arc and discontinuities in the transformed control variable. Because of the finite time nature of the boundary-layer solutions, the traditional asymptotic stability analysis and method of matched asymptotic expansions cannot be used in constructing an approximate solution. Instead, the costate jumps that can occur and the boundary-layer final time must be used as free parameters to satisfy continuity conditions in the state variables at the end of the boundary-layer response. This technique has proved quite satisfactory when used to construct an approximate solution for a relatively simple nonlinear example, at least for small perturbations away from the reduced solution.

Acknowledgments

This research was supported by the NASA Langley Research Center under Grant NAG-1-922. The authors also wish to thank Nickos Markopoulos for reducing the constraining equations of example 3 to a quartic.

References

[1]Bryson, A. E., Jr., and Ho, Y.-C., *Applied Optimal Control*, Hemisphere, New York, 1975.

[2]Jacobson, D. H., Lele, M. M., and Speyer, J. L., "New Necessary Conditions of Optimality for Control Problems with State-Variable Inequality Constraints," *Journal of Mathematical Analysis and Applications*, Vol. 35, Aug. 1971, pp. 255–284.

[3]Kreindler, E., "Additional Necessary Conditions for Optimal Control with State-Variable Inequality Constraints," *Journal of Optimization Theory and Applications,* Vol. 30, No. 2, 1982, pp. 241-251.

[4]Boykin, W. H., Jr., and Bullock, T. E., "State Constraints and Singular Solutions to Penalty Function Optimization Problems," *AIAA Journal,* Vol. 10, No. 2, 1972, p. 137-141.

[5]Calise, A. J., "Singular Perturbation Methods for Variational Problems in Aircraft Flight," *IEEE Transactions on Automatic Control,* Vol. AC-21, No. 3, 1976, pp. 345-353.

[6]Calise, A. J., and Moerder, D. D., "Singular Perturbation Techniques for Real Time Aircraft Trajectory Optimization and Control," NASA CR 3597, Aug. 1982.

[7]Falco, M., and Kelly, H. J., "Aircraft Symmetric Flight Optimization," *Control and Dynamic Systems, Advances in Theory and Applications,* Vol. 10, Academic Press, New York, 1973 (see Fig. 14, p. 114).

[8]McHenry, R. L., Brand, T. J., Long, A. D., Cockrell, B. F., and Thibodeau, J. R., III, "Space Shuttle Ascent Guidance, Navigation, and Control," *Journal of the Astronautical Sciences,* Vol. 27, No. 1, 1979, pp. 1-38.

[9]Hargraves, C. R., and Paris, S. W., "Direct Trajectory Optimization Using Nonlinear Programming and Collocation," *Journal of Guidance, Control, and Dynamics,* Vol. 10, No. 4, 1987, pp. 338-342.

[10]Ardema, M. D., "Linearization of the Boundary Layer Equations of the Minimum Time to Climb Problem," *Journal of Guidance and Control,* Vol. 2, No. 5, 1979, pp. 434-436.

[11]Jacobson, D. H., and Lele, M. M., "A Transformation Technique for Optimal Control Problems with a State Variable Inequality Constraint," *IEEE Transactions on Automatic Control,* Vol. AC-14, No. 5, 1969, pp. 457-464.

[12]Ardema, M. D., "Singular Perturbations in Flight Mechanics," NASA TM X-62, 380, Aug. 1974; revised July 1977.

[13]Ardema, M. D., "Nonlinear Singularly Perturbed Optimal Control Problems with Singular Arc," *Automatica,* Vol. 16, No. 1, 1980, pp. 99-104.

[14]Calise, A. J., "A New Boundary Layer Matching Procedure for Singularly Perturbed Systems," *IEEE Transactions on Automatic Control,* Vol. AC-23, No. 3, 1978, pp. 434-438.