

# Eigenvector approximate dichotomic basis method for solving hyper-sensitive optimal control problems

Anil V. Rao<sup>1,\*,\*†</sup> and Kenneth D. Mease<sup>2,‡</sup>

<sup>1</sup> *Department of Mechanical and Aerospace Engineering, Princeton University, Princeton, NJ 08544, U.S.A.*

<sup>2</sup> *Department of Mechanical and Aerospace Engineering, University of California, Irvine, CA 92697, U.S.A.*

A number of corrections to the text of this paper were inadvertently omitted from the version published in *Optim. Control. Appl. Meth.*, **20**, 59–77 (1999). The publisher apologizes for any inconvenience.

## SUMMARY

The dichotomic basis method is further developed for solving completely hyper-sensitive Hamiltonian boundary value problems arising in optimal control. For this class of problems, the solution can be accurately approximated by concatenating an initial boundary-layer segment, an equilibrium segment, and a terminal boundary-layer segment. Constructing the solution in this composite manner alleviates the sensitivity. The method uses a dichotomic basis to decompose the Hamiltonian vector field into its stable and unstable components, thus allowing the missing initial conditions needed to specify the initial and terminal boundary-layer segments to be determined from partial equilibrium conditions. The dichotomic basis reveals the phase-space manifold structure in the neighbourhood of the optimal solution. The challenge is to determine a sufficiently accurate approximation to a dichotomic basis. In this paper we use an approximate dichotomic basis derived from local eigenvectors. An iterative scheme is proposed to handle the approximate nature of the basis. The method is illustrated on an example problem and its general applicability is assessed. Copyright © 2000 John Wiley & Sons, Ltd.

KEY WORDS: dichotomic transformations; singular perturbations; time scales; optimal control; numerical methods

## 1. INTRODUCTION

Over the past several decades, significant progress has been made in the development of both direct and indirect numerical methods for solving optimal control problems [1]. Direct methods that convert the optimal control problem to a non-linear programming problem using implicit integration have risen to prominence for use in general codes [1–4] because of their versatility and ease of use. However, our focus in this paper is on the development of an indirect method for reasons given below.

\* Correspondence to: Anil V. Rao, Flight Mechanics Department, The Aerospace Corporation, MS M4/957, P.O. Box 92957, Los Angeles, CA 90009. U.S.A. E-mail: avr96@alumni.princeton.edu.

† Senior Member of Technical Staff.

‡ Associate Professor.

Contract/Grant Sponsor: National Science Foundation & California Space Institute.

Many optimally controlled dynamical systems evolve on two or more widely separated time-scales. A subclass of multiple time-scale optimal control problems is the class of completely hyper-sensitive optimal control problems. The solution of a completely hyper-sensitive problem has a characteristic three-segment structure, described qualitatively as ‘take-off’, ‘cruise’, and ‘landing’ by analogy to an optimal airport-to-airport trajectory for a transport aircraft [5]. The optimal solution is primarily determined by the following considerations. The cruise segment is determined by the cost function and the state dynamics, while it is almost independent of the boundary conditions. The take-off segment is determined by the initial conditions, the state dynamics, and the goal of reaching the cruise segment in forward time. The landing segment is determined by the terminal conditions, the state dynamics, and the goal of reaching the cruise segment in backward time. As the time interval increases, the fraction of time spent in the cruise segment increases.

The key to handling the hyper-sensitivity with a direct method of the type mentioned above is to use a higher density of nodes in the take-off and landing segments. While a direct method may produce an accurate solution to a hyper-sensitive optimal control problem, it may not produce insight as to the multiple time-scale structure of the optimally controlled system, insight that could facilitate the development of simple yet effective feedback control laws. Indirect methods require more insight as to the problem characteristics, but simultaneously offer more opportunity to develop such insight.

In an indirect method, candidate optimal (i.e. extremal) solutions are obtained by solving a Hamiltonian boundary-value problem (HBVP). Straightforward indirect methods (e.g. simple shooting) often suffer from ill-conditioning due to ‘extreme sensitivity to initial conditions’ [6]. The indirect multiple-shooting method [7] is a means of overcoming this ill-conditioning. The form of ill-conditioning of interest here arises when the time interval of interest is long relative to the rates of expansion and contraction in certain directions in the neighbourhood of the optimal solution [8]. In the case of fast rates in all directions the HBVP, as well as the corresponding optimal control problem, is called *completely hyper-sensitive*. For problems with fast rates in only some directions, the HBVP and the optimal control problem are called *partially hyper-sensitive*. Partially hyper-sensitive optimal control problems have two or more time-scales and have solutions with boundary, and possibly interior, layers with fast changes but otherwise evolve slowly.

The solution to a completely hyper-sensitive HBVP can be approximated by concatenating an initial boundary-layer segment, an equilibrium segment and a terminal boundary-layer segment; these segments correspond to the take-off, cruise and landing segments described above. Constructing the solution in this composite manner alleviates the sensitivity. The (analytical) singular perturbation method [5, 9, 10] is a formal means of constructing the composite solution using asymptotic expansions. It has the attribute of being applicable to the partially hyper-sensitive case, but it has the weakness of requiring the state dynamics to be given in a special form that essentially requires *a priori* knowledge of the time-scale structure. A Hamilton–Jacobi–Bellman equation based approach has been proposed [11] for completely hyper-sensitive optimal control problems, but this approach is not extendible to the partially hyper-sensitive case.

The method considered in this paper uses a dichotomic basis to decompose the Hamiltonian vector field into its stable and unstable components, thus allowing the missing initial conditions needed to specify the initial and terminal boundary-layer segments to be determined from partial equilibrium conditions. A by-product of this solution method is the dichotomic basis which provides information on the phase space manifold structure in the neighbourhood of the

optimal solution. Moreover, the method is extendible to a broad class of partially hyper-sensitive boundary-value problems. The dichotomic basis method is inspired by the computational singular perturbation (CSP) methodology for stiff initial value problems [12–14]. Previous consideration of the application of CSP to optimal control can be found in Ardema [15]. Previous development and applications of the dichotomic basis method are presented by Rao and Mease [8, 16, 17]. The challenge in making the dichotomic basis method viable is to develop a means of determining a sufficiently accurate approximation to a dichotomic basis. In this paper we determine an approximate dichotomic basis from local eigenvectors and propose an iterative algorithm to handle the approximate nature of the basis. The method is illustrated on an example problem, and its general applicability is assessed.

## 2. HAMILTONIAN BOUNDARY VALUE PROBLEM

In this paper, we are interested in the following class of optimal control problems. Find the piecewise continuous control  $u(t) \in \mathbb{R}^m$  on  $[0, t_f]$  that minimizes the scalar cost

$$J = \int_0^{t_f} \mathcal{L}[x, u] dt \quad (1)$$

subject to the differential constraint

$$\dot{x} = f(x, u) \quad (2)$$

and boundary conditions

$$\begin{aligned} x(0) &= x_0 \\ x(t_f) &= x_f \end{aligned} \quad (3)$$

where  $x(t) \in \mathbb{R}^n$  is the *state*.

The first-order necessary conditions for optimality lead to a Hamiltonian boundary-value problem (HBVP) for the extremal trajectories. This HBVP is composed of the Hamiltonian differential equations

$$\begin{aligned} \dot{x} &= [\partial H^* / \partial \lambda]^T \\ \dot{\lambda} &= -[\partial H^* / \partial x]^T \end{aligned} \quad (4)$$

and the boundary conditions

$$\begin{aligned} x(0) &= x_0 \\ x(t_f) &= x_f \end{aligned} \quad (5)$$

where  $\lambda(t) \in \mathbb{R}^n$  is the *adjoint* and  $H^*(x, \lambda) = \mathcal{L}(x, u^*(x, \lambda)) + \lambda^T f(x, u^*(x, \lambda))$  is the *Hamiltonian* evaluated at the optimal control  $u^*(x, \lambda) = \arg \min_u H(x, \lambda, u)$ . Points  $p = (x, \lambda)$  lie in the  $2n$ -dimensional *Hamiltonian phase space*, or more simply, the *phase space*. Since  $J$  and  $f(x, u)$  do not depend explicitly on time,  $H^*$  is constant along trajectories of equation (4). We use  $\dot{p} = G(p)$  as an alternate expression for the Hamiltonian system in equation (4) and refer to  $G(p)$  as the Hamiltonian vector field.  $G(p)$  is assumed to be continuously differentiable.

The solution to a completely hyper-sensitive HBVP, when viewed as a trajectory in the phase space, lies in a neighbourhood  $N \subset \mathbb{R}^{2n}$  of an equilibrium point  $\bar{p}$  of the vector field, i.e. of a point  $\bar{p}$  such that  $G(\bar{p}) = 0$ . The Hamiltonian nature of  $G$  dictates that the Jacobian  $J = \partial G / \partial p$  evaluated at  $\bar{p}$  will have eigenvalues that are symmetric about the imaginary axis in the complex plane. For a completely hyper-sensitive HBVP, none of the eigenvalues lie on the imaginary axis. Consequently,  $\bar{p}$  is a saddle point.

### 3. COMPLETELY HYPER-SENSITIVE HBVP

For sufficiently large values of  $t_f$ , the HBVP of equation (4) is *completely hyper-sensitive*. The main features of a completely hyper-sensitive HBVP are conveyed by the following example. These features suggest an approximate three-segment solution.

#### 3.1. Motivating example

Consider the following optimal control problem. Minimize the cost

$$J = \int_0^{t_f} (x^2 + u^2) dt \quad (6)$$

subject to the differential constraint

$$\dot{x} = -x^3 + u \quad (7)$$

and boundary conditions

$$\begin{aligned} x(0) &= 1 \\ x(t_f) &= 1.5 \end{aligned} \quad (8)$$

Applying the first-order necessary conditions for optimality leads to the HBVP

$$\begin{aligned} \dot{x} &= -x^3 - \lambda/2, & x(0) &= 1 \\ \dot{\lambda} &= -2x + 3x^2\lambda, & x(t_f) &= 1.5 \end{aligned} \quad (9)$$

This is a completely hyper-sensitive HBVP for sufficiently large  $t_f$ . The important features can be seen by looking at the solution, both as a function of time and as a phase plane trajectory. The solution was computed using the *Sparse Optimal Control Software (SOCS)* [2] and is shown in Figures 1 and 2 for  $t_f = (1, 3, 5, 10, 25)$ . It can be seen that as  $t_f$  increases, the solution attains the aforementioned ‘take-off’, ‘cruise’, and ‘landing’ structure. Notice that  $x = 0$  is an equilibrium of the open-loop system, thus no control effort is required to hold the system at  $x = 0$ . The cost for the equilibrium solution is zero. Therefore,  $x = 0$  is a desirable state (cruise condition) for the system to be in. As  $t_f$  increases, the control objective can be viewed as transitioning from ‘steering the system from the initial conditions to the terminal conditions’ to ‘steering the system from the initial conditions to the equilibrium and then steering the system from the equilibrium to the terminal conditions’. Strictly speaking, this transition never takes place. However, assuming such a transition does occur, the problem can be decomposed into simpler subproblems; the associated error decreases as  $t_f$  increases.

3.2. Phase-space structure

Figure 3 shows the solutions to the example HBVP as phase plane trajectories for  $t_f = (1, 3, 5, 10, 25)$ . Also shown are the stable manifold  $\mathcal{W}_s(\bar{p})$  and the unstable manifold  $\mathcal{W}_u(\bar{p})$  of the saddle point  $\bar{p}$ .  $\mathcal{W}_s(\bar{p})$  is composed of all the trajectories that approach  $\bar{p}$  as  $t \rightarrow \infty$  while  $\mathcal{W}_u(\bar{p})$  is composed of all the trajectories that approach  $\bar{p}$  as  $t \rightarrow -\infty$ . It can be seen that the solutions to the example HBVP for  $t_f = 10$  and  $t_f = 25$  are indistinguishable from the trajectory formed by  $\mathcal{W}_s(\bar{p})$  and  $\mathcal{W}_u(\bar{p})$ . Both  $\mathcal{W}_s(\bar{p})$  and  $\mathcal{W}_u(\bar{p})$  are one-dimensional for the example. For an  $n$ -dimensional state, they would each be  $n$ -dimensional. These manifolds are the non-linear analogs

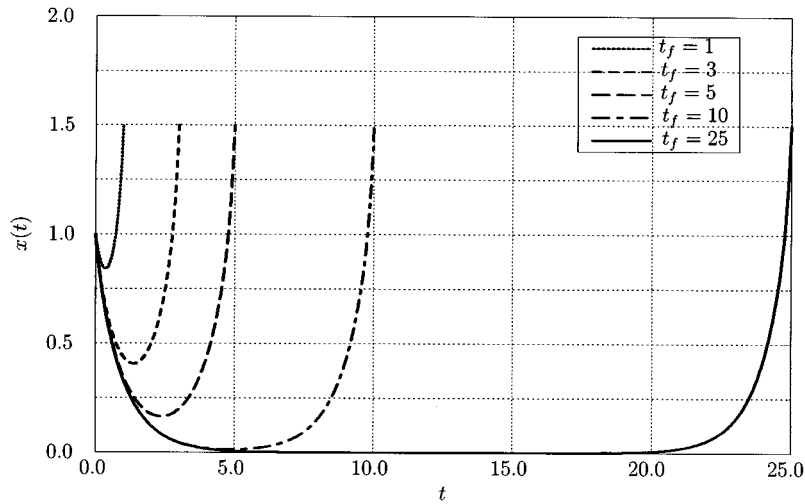


Figure 1.  $x(t)$  vs.  $t$  for example problem of equation (9);  $t_f = (1, 3, 5, 10, 25)$

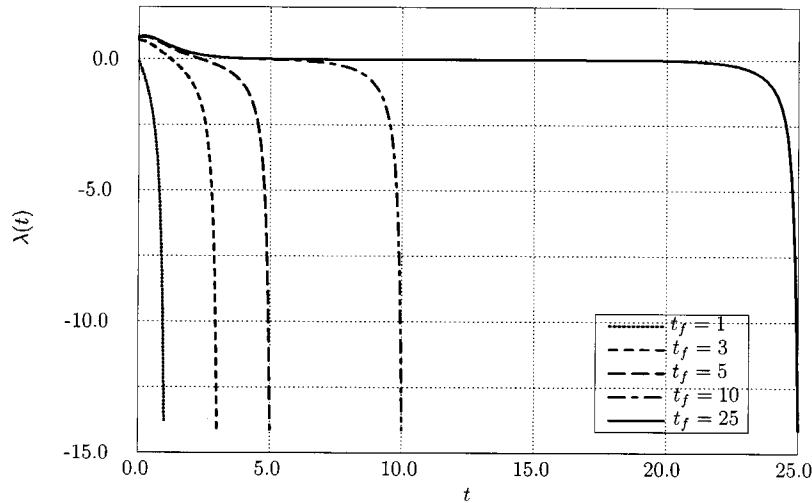


Figure 2.  $\lambda(t)$  vs.  $t$  for example problem of equation (9);  $t_f = (1, 3, 5, 10, 25)$

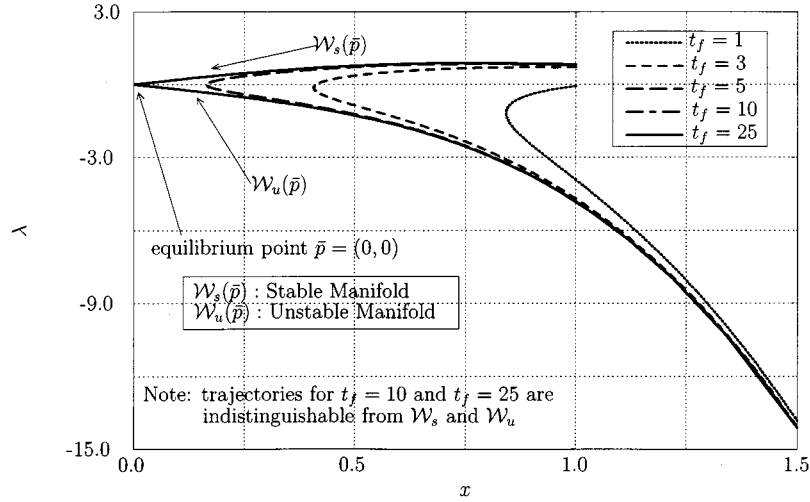


Figure 3.  $\lambda$  vs.  $x$  for example problem of equation (9);  $t_f = (1, 3, 5, 10, 25)$

of the stable and unstable eigenspaces for a linear-time invariant system [18]. They are tangent to the stable and unstable eigenspaces of the linearized dynamics at  $\bar{p}$ . The phase plane trajectories, for sufficiently large  $t_f$ , begin very near  $\mathcal{W}_s(\bar{p})$ , proceed quickly adjacent to  $\mathcal{W}_s(\bar{p})$  toward  $\bar{p}$ , make a slow turn in the vicinity of  $\bar{p}$ , and then proceed quickly away from  $\bar{p}$  adjacent to  $\mathcal{W}_u(\bar{p})$ . The initial boundary-layer segment lies closer and closer to  $\mathcal{W}_s(\bar{p})$  as  $t_f$  increases; the terminal boundary-layer segment lies closer and closer to  $\mathcal{W}_u(\bar{p})$  as  $t_f$  increases.

### 3.3. Three-segment approximate solution

Consider the approximate composite solution  $\hat{p}$  given by

$$\hat{p} = \begin{cases} p_s(t), & 0 \leq t \leq t_{\text{ibl}} \\ \bar{p}, & t_{\text{ibl}} < t < t_{\text{fbl}} \\ p_u(t), & t_{\text{fbl}} \leq t \leq t_f \end{cases} \quad (10)$$

where  $p_s$  is the solution to equation (4) with initial condition  $p_s(0) = (x_0, \lambda_0)$  and  $\lambda_0$  chosen such that  $p_s(0) = (x_0, \lambda_0) \in \mathcal{W}_s(\bar{p})$ ;  $\bar{p} = (\bar{x}, \bar{\lambda})$  is the equilibrium solution; and  $p_u$  is the solution to equation (4) with final condition  $p_s(t_f) = (x_f, \lambda_f)$  with  $\lambda_f$  chosen such that  $p_u(t_f) = (x_f, \lambda_f) \in \mathcal{W}_u(\bar{p})$ . In words, the composite approximation is constructed by concatenating an initial boundary-layer segment on the stable manifold, an equilibrium segment, and a terminal boundary-layer segment on the unstable manifold [11, 17]. The durations of the initial and terminal boundary-layers,  $t_{\text{ibl}}$  and  $t_f - t_{\text{fbl}}$ , must be selected; they must be long enough to allow the boundary-layer segments to reach the equilibrium to sufficient accuracy in forward and backward time, respectively.

Let  $p^*$  denote the solution of interest to the HBVP given by equations (4) and (5). Define the approximation error by

$$\|\hat{p} - p^*\|_\infty = \sup_{t \in [0, t_f]} \|\hat{p}(t) - p^*(t)\|_2 \quad (11)$$

where  $\|\cdot\|_2$  denotes the standard Euclidean norm on  $\mathbb{R}^{2n}$ . If one is interested in a solution that approximates  $p^*$  to within a tolerance  $\varepsilon$  and  $\|\hat{p} - p^*\|_\infty < \varepsilon$ , then the proposed  $\hat{p}$  is a candidate.

The options for determining  $p_s$  are: (i) If  $n = 1$ , then  $H^*(x_0, \lambda_0) = H^*(\hat{x}, \hat{\lambda})$  can be used to determine  $\lambda_0$  given  $x_0$ . Then  $p_s$  can be determined by forward integration of equation (4). Because of the restriction on  $n$ , this is not a general approach. (ii) Finding a small perturbation in the stable eigenspace of  $\bar{p}$  from which backward integration leads to the point on  $\mathcal{W}_s$  that satisfies the initial condition  $x = x_0$ . (iii)  $\mathcal{W}_s$  could in principle be given by  $n$  algebraic equations in  $x$  and  $\lambda$  and these equations could be used to determine  $\lambda_0$  given  $x_0$ . Then  $p_s$  can be determined by forward integration of equation (4). (iv) Separate the Hamiltonian vector field  $\dot{p} = G(p)$  into its stable and unstable components  $G(p) = G_s(p) + G_u(p)$ . Use the partial equilibrium condition  $G_u(p_0) = 0$  to determine  $\lambda_0$  given  $x_0$ . Then  $p_s$  can be determined by forward integration of equation (4). These options also apply to determining  $p_u$  with the direction of time reversed.

A dichotomic basis method for implementing option (iv) has been presented by Rao and Mease [8, 16, 17]. We consider this method further in the remainder of this paper. The dichotomic basis method is the only one of the four options just mentioned that has potential for more general multiple time-scale (partially hyper-sensitive) HBVPs. For partially hyper-sensitive problems, there is a well-defined manifold structure and the concept of partial equilibrium conditions is still appropriate [19]. Option (ii) is not appropriate because there is no saddle point. Option (iii) is valid but it is not likely realizable.

#### 4. APPROXIMATE DICHOTOMIC BASIS METHOD

We focus on developing a method for computing  $p_s$ , because the same method will work for  $p_u$  simply by reversing time. In this section, the dichotomic basis and approximate dichotomic basis methods are presented. These methods are reviewed here to set the stage for the new developments presented in the following section.

##### 4.1. Dichotomic Basis

Restricting our attention to completely hyper-sensitive HBVPs, at each point in the phase space along the optimal trajectory  $p^*$ , there is an  $n$ -dimensional contracting subspace and an  $n$ -dimensional expanding subspace. This means that a neighbouring optimal trajectory that begins in the contracting subspace will approach  $p^*$  in forward time, while a neighbouring optimal trajectory that begins in the expanding subspace will approach  $p^*$  in backward time (i.e. in forward time it will depart from  $p^*$ ). These subspaces in general vary in direction along  $p^*$ . We assume that this property also holds along trajectories in a neighbourhood  $N \subset \mathbb{R}^{2n}$  of  $p^*$  and that the approximate solution  $\hat{p}$  from equation (10) lies in  $N$ . Let the columns of the matrix  $D(p) \in \mathbb{R}^{2n \times 2n}$  form a continuously differentiable basis for  $\mathbb{R}^{2n}$  on  $N$ . A vector, such as  $G(p)$ , in the phase space can be written in terms of  $D$  as

$$G(p) = D(p)v(p) \tag{12}$$

where the components of  $v \in \mathbb{R}^{2n}$  are the components of  $G$  in the new basis. Differentiating along  $p(t)$ , we have

$$\dot{v} = (D^{-1}JD - D^{-1}\dot{D})v = \Lambda v \tag{13}$$

where  $J = \partial G/\partial p$  is the Jacobian of  $G(p)$ .  $D(p)$  is called a *dichotomic basis* in the neighbourhood  $N$ , if the following two properties are satisfied [8]:

1. For each  $p \in N$ ,  $\Lambda(p)$  has the block-triangular form

$$\Lambda(p) = \begin{bmatrix} \Lambda_s(p) & \Lambda_{su}(p) \\ 0 & \Lambda_u(p) \end{bmatrix} \quad (14)$$

where  $\Lambda_s(p) \in \mathbb{R}^{n \times n}$ ,  $\Lambda_u(p) \in \mathbb{R}^{n \times n}$ , and  $\Lambda_{su}(p) \in \mathbb{R}^{n \times n}$ .

2. Along any segment  $p_{[t_1, t_2]}$  of a trajectory  $p$  of the Hamiltonian vector field  $G(p)$  lying in  $N$ , the transition matrices  $\Phi_s^p(t, 0)$  and  $\Phi_u^p(t, 0)$  corresponding to  $\Lambda_s$  and  $\Lambda_u$ , defined such that  $\Phi_s^p(0, 0) = I$  and  $\Phi_u^p(0, 0) = I$ , satisfy the inequalities

$$\|\Phi_s^p(t, 0) [\Phi_s^p(\tau, 0)]^{-1}\| \leq K_1 \|D(p(\tau))\| \|D^{-1}(p(t))\| e^{-\alpha(t-\tau)}, \quad t_2 > t \geq \tau > t_1 \quad (15)$$

$$\|\Phi_u^p(t, 0) [\Phi_u^p(\tau, 0)]^{-1}\| \leq K_1 \|D(p(\tau))\| \|D^{-1}(p(t))\| e^{-\alpha(\tau-t)}, \quad t_1 \leq t \leq \tau < t_2 \quad (16)$$

where  $K_1 > 0$  and  $\alpha > 0$  are scalars that can vary on  $N$ . This property ensures that  $\Phi_s^p$  contracts vectors exponentially in forward time while  $\Phi_u^p$  contracts vectors exponentially in backward time. Since we are working with finite time intervals, it is important that the exponential bounds are tight at  $t = \tau$ . See Reference [8] for further details.

A dichotomic basis decouples the expanding dynamics from the contracting dynamics. Denoting contracting (stable) and expanding (unstable) by the subscripts 's' and 'u', respectively, a dichotomic basis can be split as

$$D(p) = [D_s(p) \quad D_u(p)] \quad (17)$$

where the columns of  $D_s(p) \in \mathbb{R}^{2n \times n}$  span the contracting subspace and the columns of  $D_u(p) \in \mathbb{R}^{2n \times n}$  span the expanding subspace. Correspondingly, the vector  $v$  splits as

$$v = \begin{bmatrix} v_s \\ v_u \end{bmatrix} \quad (18)$$

where  $v_s \in \mathbb{R}^n$  and  $v_u \in \mathbb{R}^n$ .

For a point  $p$  on the stable manifold  $\mathcal{W}_s$ ,  $D_s(p)$  coincides with the tangent space to  $\mathcal{W}_s$ . Because  $\mathcal{W}_s$  is an invariant manifold [18],  $G(p)$  must lie in the column span of  $D_s(p)$  for  $p \in \mathcal{W}_s$ . It follows that the components  $v_u(p)$  of  $G(p)$  must be zero for  $p \in \mathcal{W}_s$ . If the differential equations for  $v$  are adjoined to the differential equation for  $p$  and  $G(p)$  is represented in the dichotomic basis  $D$ , we have

$$\begin{bmatrix} \dot{p} \\ \dot{v}_s \\ \dot{v}_u \end{bmatrix} = \begin{bmatrix} D_s(p) & D_u(p) \\ \Lambda_s(p) & \Lambda_{su}(p) \\ 0 & \Lambda_u(p) \end{bmatrix} \begin{bmatrix} v_s \\ v_u \end{bmatrix} \quad (19)$$

If equations (19) and (4) are initialized consistently, their solutions will be identical. It is clear from the structure of the system in equation (19) that if  $v_u$  is initially zero it will remain zero. Because  $v_u$  is the projection of  $G(p)$  into the unstable subspace at a point  $p$ , the only way to make  $v_u$  zero is to choose  $p$  properly. In particular, if  $p$  is chosen to lie in  $\mathcal{W}_s$ , then  $v_u$  will be zero. At the initial time, because  $x(0) = x_0$  is specified, it is  $\lambda(0)$  that is adjusted to make  $v_u$  zero such that  $p(0) = (x_0, \lambda(0)) \in \mathcal{W}_s$ .



With  $v_u = 0$ , equation (19) reduces to

$$\begin{aligned}\dot{p} &= D_s(p)v_s \\ \dot{v}_s &= \Lambda_s(p)v_s\end{aligned}\tag{20}$$

Using the notation  $p(0) = (x(0), \lambda(0)) = p_0$ , the solution  $p_s(t)$  is found by integrating equation (20) with the initial conditions

$$\begin{aligned}x(0) &= x_0 \\ \lambda(0) \text{ found from solving } v_u(0) &= D_u^\dagger(p_0)G(p_0) = 0\end{aligned}\tag{21}$$

$$v_s(0) = D_s^\dagger(p_0)G(p_0)$$

where  $D_s^\dagger(p) \in \mathbb{R}^{n \times 2n}$  and  $D_u^\dagger(p) \in \mathbb{R}^{n \times 2n}$  are the first and second, respectively,  $n$  rows of  $D^{-1}$ :

$$D^{-1}(p) = \begin{bmatrix} D_s^\dagger(p) \\ D_u^\dagger(p) \end{bmatrix}\tag{22}$$

Any value of  $p_0$  that satisfies the equation  $D_u^\dagger(p_0)G(p_0) = 0$  implies that  $\lambda(0)$  lies in the stable manifold of  $\bar{p}$  at  $x(0)$ . Integrating equation (20) from  $t = 0$  to  $t = t_{\text{ibl}}$  so that  $\|p(t_{\text{ibl}}) - \bar{p}\| < \varepsilon$ , where  $\varepsilon > 0$  is a specified tolerance, produces the initial boundary-layer approximation  $p_s$ .

#### 4.2. Approximate dichotomic basis method

In this subsection, we describe a method for constructing  $p_s$  using a basis that only approximately decouples the stable and unstable dynamics. We begin by defining an approximate dichotomic basis.

Let  $A(p) = [A_s(p)A_u(p)]$  be a non-dichotomic basis on  $N$  where  $A_s(p) \in \mathbb{R}^{2n \times n}$  and  $A_u(p) \in \mathbb{R}^{2n \times n}$ . Furthermore, let  $A^{-1}(p)$  be written in terms of the two matrices  $A_s^\dagger(p) \in \mathbb{R}^{n \times 2n}$  and  $A_u^\dagger(p) \in \mathbb{R}^{n \times 2n}$  such that

$$A^{-1}(p) = \begin{bmatrix} A_s^\dagger(p) \\ A_u^\dagger(p) \end{bmatrix}\tag{23}$$

and let  $h$  represent the co-ordinate vector for the basis  $A$  with  $n$ -dimensional components  $h_s$  and  $h_u$

$$h = \begin{bmatrix} h_s \\ h_u \end{bmatrix}\tag{24}$$

In terms of  $A(p)$ , the differential equations for  $p$  and  $h$  are

$$\begin{bmatrix} \dot{p} \\ \dot{h}_s \\ \dot{h}_u \end{bmatrix} = \begin{bmatrix} A_s(p) & A_u(p) \\ \Lambda_s(p) & \Lambda_{su}(p) \\ \Lambda_{us}(p) & \Lambda_u(p) \end{bmatrix} \begin{bmatrix} h_s \\ h_u \end{bmatrix}\tag{25}$$

where

$$\Lambda = \begin{bmatrix} \Lambda_s(p) & \Lambda_{su}(p) \\ \Lambda_{us}(p) & \Lambda_u(p) \end{bmatrix} = A^{-1}JA - A^{-1}\dot{A}\tag{26}$$

Let  $\Gamma_s^p(t, 0)$  and  $\Gamma_u^p(t, 0)$  be the transition matrices of  $\Lambda_s$  and  $\Lambda_u$ , respectively, from equation (26). Because the columns of  $A(p)$  form a non-dichotomic basis, in general, the transition matrices  $\Gamma_s^p(t, 0)$  and  $\Gamma_u^p(t, 0)$  have both contracting and expanding behaviour along trajectory segments in  $N$  and  $\Lambda_{us}(p) \neq 0$  for  $p \in N$ . In terms of the dichotomic and non-dichotomic bases, the vector  $G(p)$  can be written as

$$G(p) = D_s(p)v_s + D_u(p)v_u = A_s(p)h_s + A_u(p)h_u \quad (27)$$

Suppose we attempt to place the initial point  $p_0$  on the stable manifold by choosing  $\lambda(0)$  so that  $h_u(0) = 0$ . Because the basis field  $A$  is non-dichotomic, the calculated initial point will not be on the stable manifold. The size the unstable component  $v_u$  is a measure of the error. The unstable components of  $G$  at the beginning and end of the trajectory are

$$v_u(0) = D_u^\dagger(p_0)A_s(p_0)h_s(0) \quad (28)$$

and

$$v_u(t_{\text{ibl}}) = \Phi_u^p(t_{\text{ibl}}, 0)D_u^\dagger(p_0)A_s(p_0)h_s(0) \quad (29)$$

The norm of  $v_u(t_{\text{ibl}})$  can be bounded by

$$\|v_u(t_{\text{ibl}})\| = \|\Phi_u^p(t_{\text{ibl}}, 0)D_u^\dagger(p_0)A_s(p_0)h_s(0)\| \leq \|\Phi_u^p(t_{\text{ibl}}, 0)\| \|D_u^\dagger(p_0)A_s(p_0)\| \|h_s(0)\| \quad (30)$$

where  $\|\Phi_u^p(t_{\text{ibl}}, 0)\|$  and  $\|D_u^\dagger(p_0)A_s(p_0)\|$  denote matrix norms induced by the vector norm. Note that  $\|D_u^\dagger(p_0)A_s(p_0)\|$  determines the size of  $v_u$  and is a measure of how well the basis  $A$  approximates the dichotomic basis  $D$  at point  $p_0$ , because  $\|D_u^\dagger(p_0)D_s(p_0)\| = 0$ . In general  $\|\Phi_u^p(t, 0)\|$  is exponentially increasing with  $t$ ; thus, the size of  $t_{\text{ibl}}$  relative to the rate of growth determines how small  $\|D_u^\dagger(p_0)A_s(p_0)\|$  must be. The key benefit of the composite approximation approach is that error growth, an inherent feature of integrating a Hamiltonian system, is only taking place over the time interval  $t_{\text{ibl}}$ , which is much smaller than  $t_r$ . The factor  $\|D_u^\dagger(p_0)A_s(p_0)\|$  provides additional attenuation of the growing component of the solution.

Suppose that  $A(p)$  is an approximate dichotomic basis and that the  $h_u(\cdot)$  component of  $G(p)$  is known along the desired trajectory  $p_s$  over the time interval  $[0, t_{\text{ibl}}]$ . Then, using  $h_u(\cdot)$  as an input, the trajectory can be found by integrating the system of equations

$$\begin{bmatrix} \dot{p} \\ \dot{h}_s \end{bmatrix} = \begin{bmatrix} A_s(p) & A_u(p) \\ \Lambda_s & \Lambda_{su} \end{bmatrix} \begin{bmatrix} h_s \\ h_u \end{bmatrix} \quad (31)$$

in forward time with the initial conditions

$$\begin{aligned} x(0) &= x_0 \\ \lambda(0) &\text{ found from solving } A_u^\dagger(p_0)G(p_0) = h_u(0) \\ h_s(0) &= A_s^\dagger(p_0)G(p_0) \end{aligned} \quad (32)$$

Similarly, given  $x(t_{\text{ibl}}) = x_{\text{ibl}}$  and using  $h_s(\cdot)$  as an input, the trajectory can be found by integrating the system of equations

$$\begin{bmatrix} \dot{p} \\ \dot{h}_u \end{bmatrix} = \begin{bmatrix} A_s(p) & A_u(p) \\ \Lambda_{us} & \Lambda_u \end{bmatrix} \begin{bmatrix} h_s \\ h_u \end{bmatrix} \quad (33)$$

in backward time with the terminal conditions

$$\begin{aligned} x(t_{\text{ibl}}) &= x_{\text{ibl}} \\ \lambda(t_{\text{ibl}}) \text{ found from solving } A_s^\dagger(p(t_{\text{ibl}}))G(p(t_{\text{ibl}})) &= h_s(t_{\text{ibl}}) \\ h_u(t_{\text{ibl}}) &= A_u^\dagger(p(t_{\text{ibl}}))G(p(t_{\text{ibl}})) \end{aligned} \quad (34)$$

However, since neither  $h_s$  nor  $h_u$  along  $p_s$  is known *a priori*, we will solve equations (31)–(34) by a successive approximation approach given by Algorithm 1.

*Algorithm 1*

Let  $A(p)$  be an approximate dichotomic basis in the neighbourhood  $N$ . Choose a convergence level  $\delta > 0$  and a matching tolerance  $\varepsilon > 0$ .

- (i) Choose  $t_{\text{ibl}}$  and  $h_u(\cdot)$  on  $t \in [0, t_{\text{ibl}}]$  ( $h_u(\cdot) \equiv 0$ , for example).
- (ii) Generate the initial conditions at  $t = 0$  using equation (32).
- (iii) Integrate equation (31) forward from  $t = 0$  to  $t = t_{\text{ibl}}$  using  $h_u(\cdot)$  from (i) on first iteration or the value of  $h_u(\cdot)$  obtained from (v) on subsequent iterations.
- (iv) Generate the terminal conditions from equation (34) using  $x(t_{\text{ibl}}) = x_{\text{ibl}}$  and  $h_s(t_{\text{ibl}})$  from the forward integration in the previous step.
- (v) Integrate equation (33) backward from  $t = t_{\text{ibl}}$  to  $t = 0$  using  $h_s(\cdot)$  from (iii).
- (vi) Repeat (iii)–(v) until  $\|p^{(k)}(t_{\text{ibl}}) - p^{(k-1)}(t_{\text{ibl}})\| < \delta$  where  $p^{(k)}(\cdot)$  is the  $k$ th iterate of (v). If  $\|p^{(k)}(t_{\text{ibl}}) - \bar{p}\| < \varepsilon$ , then stop. Otherwise, start again at step (i) with a larger value of  $t_{\text{ibl}}$ .

5. COMPUTING AN APPROXIMATE DICHOTOMIC BASIS

Algorithm 1 can be used if an approximate dichotomic basis is known *a priori*; however, this is not usually the case. In this section we describe a method that computes an approximate dichotomic basis in conjunction with Algorithm 1. The strategy proposed here is similar to the strategy used in the computational singular perturbation (CSP) methodology for stiff initial value problems [12, 13].

5.1. Approximate dichotomic basis using eigenvectors

The Jacobian  $J(p)$  of the Hamiltonian vector field  $G(p)$  has the form

$$J(p) = \begin{bmatrix} F(p) & -K(p) \\ -Q(p) & -F^T(p) \end{bmatrix} \quad (35)$$

where  $F(p) \in \mathbb{R}^{n \times n}$ ,  $K(p) \in \mathbb{R}^{n \times n}$ , and  $Q(p) \in \mathbb{R}^{n \times n}$ . Let the columns of the matrix  $A(p) \in \mathbb{R}^{2n \times 2n}$  be the eigenvectors of  $J(p)$ . Because the equilibrium solution  $\bar{p} = (\bar{x}, \bar{\lambda})$  is a saddle point,  $J(\bar{p})$  has no eigenvalues on the imaginary axis and  $A(\bar{p})$  can be split as  $A(\bar{p}) = [A_s(\bar{p}) A_u(\bar{p})]$  where column span  $\{A_s(\bar{p})\}$  is the  $n$ -dimensional stable eigenspace and column span  $\{A_u(\bar{p})\}$  is the  $n$ -dimensional unstable eigenspace. On the equilibrium solution,  $\bar{p}$ , the term  $A^{-1}\dot{A} \equiv 0$  and the matrix  $\Lambda(\bar{p})$  has the form

$$\Lambda(\bar{p}) = A^{-1}(\bar{p})J(\bar{p})A(\bar{p}) = \begin{bmatrix} -Z & 0 \\ 0 & Z \end{bmatrix} \quad (36)$$

where the eigenvalues of  $Z$  are in the open right-half of the complex plane. Consequently, the eigenvectors of  $J(\bar{p})$  form a dichotomic basis along the equilibrium solution  $\bar{p}$ .

If the following requirement about the eigenvalue structure of  $J(p)$  is met for all  $p \in N$ ,

*Requirement 1*

At every point  $p \in N$ , the eigenvalues of the Jacobian matrix  $J(p)$  lie off the imaginary axis.

Then the eigenvectors of  $J(p)$  may provide an effective approximate dichotomic basis for points in  $N$ . Requirement 1 ensures that the Jacobian has well-defined stable and unstable eigen-directions at every point  $p \in N$ . The eigenvectors have been used as a first approximation to decouple slow and fast modes when solving stiff initial value problems in chemical kinetics using the computational singular perturbation (CSP) methodology [12, 13]. Similarly, the eigenvectors are used here to approximately decouple the contracting and expanding dynamics of an Hamiltonian system. A strategy is now developed to construct an approximate dichotomic basis.

The eigenvector matrix  $A(p)$  of  $J(p)$  can be split so that  $A(p) = [A_s(p) A_u(p)]$  where the columns of  $A_s(p)$  are the eigenvectors corresponding to the eigenvalues with negative real parts and the columns of  $A_u(p)$  are the eigenvectors corresponding to the eigenvalues with positive real parts. In the case of repeated and/or complex conjugate eigenvalues, generalized real eigenvectors are used (e.g. a complex conjugate pair of eigenvectors  $v_R \pm iv_I$  are replaced by the two real vectors  $v_R$  and  $v_I$ , respectively). Along a trajectory  $p(\cdot)$ , the eigenvector basis does not decouple the contracting and expanding behaviour because, while  $A$  will block-diagonalize  $A^{-1}JA$ , the term  $A^{-1}\dot{A}$  will in general *not* be zero and  $\Lambda$  of equation (26) will have non-zero off-diagonal blocks.

One possibility is to use the eigenvectors to form an approximate dichotomic basis along a trajectory  $p(\cdot)$ . However, this strategy is computationally intensive. More importantly, continuous re-computation of the eigenvectors is not required. Instead, it is possible to divide the time interval  $[0, t_{\text{ibl}}]$  into several subintervals and use a fixed eigenvector basis over each subinterval. This leads to the following piecewise constant basis. Let  $\{t_1, \dots, t_N\}$  be values of  $t \in [0, t_{\text{ibl}}]$  where the eigenvectors are computed and let  $t_{N+1} = t_{\text{ibl}}$ . Furthermore, let  $\{p(t_1), \dots, p(t_N)\} = \{p_1, \dots, p_N\}$  be points along  $p(\cdot)$ . For each subinterval,  $[t_i, t_{i+1}]$ ,  $i = 1, 2, \dots, N$ , the eigenvectors of  $J(p_i)$  are used as a basis. Consequently,  $A(p)$  has the form

$$A(p(t)) = A(p_i), \quad t \in [t_i, t_{i+1}], \quad i = 1, \dots, N \quad (37)$$

Moreover, along each subinterval  $[t_i, t_{i+1}]$ , the term  $A^{-1}\dot{A} \equiv 0$  and the matrix  $\Lambda$  of equation (26) has the form

$$\Lambda = A^{-1}(p(t))J(p(t))A(p(t)), \quad i = 1, \dots, N \quad (38)$$

where  $A(p(t))$  is obtained from equation (37).  $\Lambda$  is not block-diagonal in general because  $J(p)$  changes along  $p(t)$ . The number of times  $N$  that the basis vectors should be changed is highly problem dependent. Each subinterval can have a different length.

## 5.2. Initialization and switching strategy

The eigenvector basis is initialized as follows. First, an initial  $h_u$  profile on  $[0, t_{\text{ibl}}]$  is chosen;  $h_u(t) \equiv 0$  is a reasonable choice if the eigenvectors provide an accurate approximation to

a dichotomic basis. Then,  $\lambda(0)$  is found by solving

$$A_u^\dagger(p_0)G(p_0) = h_u(0) \tag{39}$$

which gives  $p_0 = (x(0), \lambda(0))$ . The initial eigenvector matrix  $A(p_0)$  is determined while solving for  $\lambda(0)$  since  $A_u^\dagger$  is needed in equation (37). With  $p_0$  and  $A(p_0)$  determined,  $h_s(0) = A_s^\dagger(p_0)G(p_0)$  and equation (31) can be integrated forward, switching to a new eigenvector basis at the chosen times.

At an eigenvector switch point, it is necessary to restart the integration because there may be discontinuities in  $h_s$  and  $h_u$ . Let  $t_{sw}$  be an eigenvector switch time and let  $(\cdot)^-$  and  $(\cdot)^+$  denote the left and right limits of a variable at  $t_{sw}$ , respectively. During a switch, the value of  $p(t)$  is continuous, i.e.  $p^-(t_{sw}) = p^+(t_{sw}) = p(t_{sw})$ . When integrating equation (31) forward in time, the values of  $h_s^-(t_{sw})$  and  $h_s^+(t_{sw})$  are given, respectively, by

$$h_s^-(t_{sw}) = (A_s^\dagger)^- G(p(t_{sw})) \tag{40}$$

and

$$h_s^+(t_{sw}) = (A_s^\dagger)^+ G(p(t_{sw})) \tag{41}$$

Because  $(A_s^\dagger)^- \neq (A_s^\dagger)^+$ , in general  $h_s^-(t_{sw}) \neq h_s^+(t_{sw})$ . The value  $h_s^+(t_{sw})$  is used as a starting condition for the next subinterval. Similarly, when integrating equation (33) backward in time, the values of  $h_u^-(t_{sw})$  and  $h_u^+(t_{sw})$  are given, respectively, by

$$h_u^-(t_{sw}) = (A_u^\dagger)^- G(p(t_{sw})) \tag{42}$$

and

$$h_u^+(t_{sw}) = (A_u^\dagger)^+ G(p(t_{sw})) \tag{43}$$

After the switch, the value  $h_u^-(t_{sw})$  is used as a starting condition for the next subinterval of  $t$ . Because of the successive approximation approach we are using, computing  $h_s$  during the forward integration and  $h_u$  during the backward integration, there can be discontinuities in  $\dot{p}$  at the switch times while the solution is converging; however, if and when the solution has converged,  $\dot{p}$  will be continuous, because  $G(p)$  is continuously differentiable by assumption.

### 5.3. Eigenvector approximate dichotomic basis method

Using the piecewise constant eigenvector basis, Algorithm 1 can be modified to give Algorithm 2 as follows. Requirement 1 must be satisfied for the algorithm to be viable.

#### Algorithm 2

Choose a convergence level  $\delta > 0$  and a matching tolerance  $\varepsilon > 0$ .

- (i) Choose  $t_{ibl}$ ,  $h_u(\cdot)$  on  $t \in [0, t_{ibl}]$  ( $h_u(\cdot) \equiv 0$ , for example), and times  $t_i \in [0, t_{ibl}]$ ,  $i = 1, 2, \dots, N$ , to switch the eigenvectors.
- (ii) Generate the initial conditions at  $t = 0$  using equation (32). The initial eigenvectors  $A(p_0)$  of  $J(p_0)$  are determined in the process.
- (iii) Integrate equation (31) forward from  $t = 0$  to  $t = t_{ibl}$  using  $h_u(\cdot)$  from (i) on first iteration or the value of  $h_u(\cdot)$  obtained from (v) on subsequent iterations. Compute a new set of eigenvectors  $A(p(t_i))$ ,  $i = 1, 2, \dots, N$ , along the trajectory.
- (iv) Generate the terminal conditions from equation (34) using  $x(t_{ibl}) = x_{ibl}$  and  $h_s(t_{ibl})$  from the forward integration in the previous step.

- (v) Integrate equation (33) backward from  $t = t_{\text{ibl}}$  to  $t = 0$  using  $h_s(\cdot)$  from (iii) and the previously computed piecewise constant  $A(p(\cdot))$ . Whenever a switch point is encountered, re-initialize  $h_u$  according to equation (42).
- (vi) Repeat (iii)–(v) until  $\|p^{(k)}(t_{\text{ibl}}) - p^{(k-1)}(t_{\text{ibl}})\| < \delta$  where  $p^{(k)}(\cdot)$  is the  $k^{\text{th}}$  iterate of (v). If  $\|p^{(k)}(t_{\text{ibl}}) - \bar{p}\| < \varepsilon$ , then stop. Otherwise, increase  $t_{\text{ibl}}$  and repeat (ii)–(v).

## 6. APPLICATION OF METHOD

Algorithm 2 is now applied to the example HBVP of equation (9) to obtain the boundary-layer solutions  $p_s$  and  $p_u$ . The composite approximation is constructed by concatenating  $p_s$  and  $p_u$  with the equilibrium solution. For both  $p_s$  and  $p_u$ , a matching tolerance of  $\varepsilon = 10^{-3}$  and a convergence level  $\delta = 10^{-3}$  are used. The numerical integration is done in MATLAB using the integration routine ode113. To obtain  $p_u$ , Algorithm 2 is applied to a time-reversed version of the dynamics. Letting  $\sigma = t_f - t$  and denoting differentiation with respect to  $\sigma$  by  $(\cdot)'$ , we have

$$\begin{aligned}x' &= x^3 + \lambda/2 \\ \lambda' &= 2x - 3x^2\lambda\end{aligned}\tag{44}$$

The terminal boundary-layer solution thus obtained can be transformed back to a function of  $t$  using  $t = t_f - \sigma$  for any desired value of  $t_f$ .

To obtain  $p_s(t) = (x_s(t), \lambda_s(t))$ , the initial value  $\lambda(0) = \lambda_0$  is found by solving equation (39) which gives a starting value  $\lambda(0) = 1.0704$ . The eigenvalues of  $J(p_0)$ , where  $p_0 = (x_0, \lambda_0) = (1, 1.0704)$ , are  $\{\mu_1, \mu_2\} = \{2.606, -2.606\}$ . Based on these local eigenvalues and those at the equilibrium point, we chose  $t_{\text{ibl}} = T = 20$ . It was found that changing the eigenvectors at  $t = 0, 2, 4, 6$ , and  $8$  was sufficient. Furthermore, a new set of eigenvectors was computed after iterations 1 and 10. For this problem, 20 iterations were required to meet the specified convergence level. Figures 4 and 5 show  $x(t)$  vs.  $t$  and  $\lambda(t)$  vs.  $t$ , respectively, for iterations 1, 10, and 20 (iteration 20 is the converged trajectory). It is seen that each iterate levels off as  $t \rightarrow t_{\text{ibl}}$ , indicating that the unstable component of  $G$  has been sufficiently eliminated. Moreover, the discontinuity in  $\dot{p}_s$ , evident during the early iterations, disappears by iteration 20.

To obtain  $p_u(\sigma) = (x_u(\sigma), \lambda_u(\sigma))$ , the solution of equation (37) is  $\lambda(0) = -11.632$ . The eigenvalues of  $J(p_0)$ , where  $p_0 = (1.5, -11.632)$ , are  $\{\mu_1, \mu_2\} = \{9.95, -9.95\}$ . Based on these local eigenvalues and those at the equilibrium point, we chose  $t_f - t_{\text{fbl}} = \sigma_{\text{fbl}} = T = 20$ . For this problem it was found that changing the eigenvectors at  $\tau = 0, 3, 6, 9, 12$ , and  $15$  was sufficient. Furthermore, a new set of eigenvectors was computed after iterations 1 and 10. For this problem, 20 iterations were required to meet the specified convergence level. Figures 6 and 7 show  $x(\sigma)$  vs.  $\sigma$  and  $\lambda(\sigma)$  vs.  $\sigma$ , respectively, for iterations 1, 10, and 20 (iteration 20 is the converged trajectory). Similar to the results for  $p_s$ , each iterate levels off as  $\sigma \rightarrow \sigma_{\text{fbl}}$  and the discontinuity in  $\dot{p}'_u$ , evident in the early iterations, disappears by iteration 20.

Using the converged solutions for  $p_s$  and  $p_u$  with  $p_u$  transformed back to a function of  $t$ , the composite approximation is constructed for  $t_f = 100$ . The results for other sufficiently large values of  $t_f$  are essentially similar; the only difference is the duration of the equilibrium segment. The initial boundary-layer segment  $p_s$  is used on the interval  $t \in [0, 20]$ , the equilibrium segment is used on the interval  $t \in [20, 80]$ , and the terminal boundary-layer segment  $p_u$  is used on the interval  $t \in [80, 100]$ . Figures 8 and 9 show the composite approximations for  $x(t)$  vs.  $t$  and  $\lambda(t)$  vs.  $t$ , respectively. If the composite approximations were plotted alongside the solutions generated using SOCS, the difference would be virtually indistinguishable.

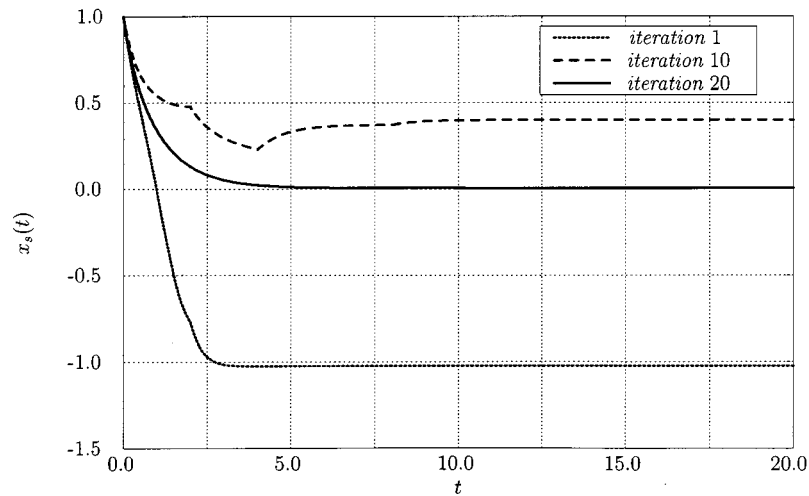


Figure 4. Solution iterates of  $x_s(t)$  vs.  $t$  for the initial boundary-layer of the example problem of equation (9) using Algorithm 2

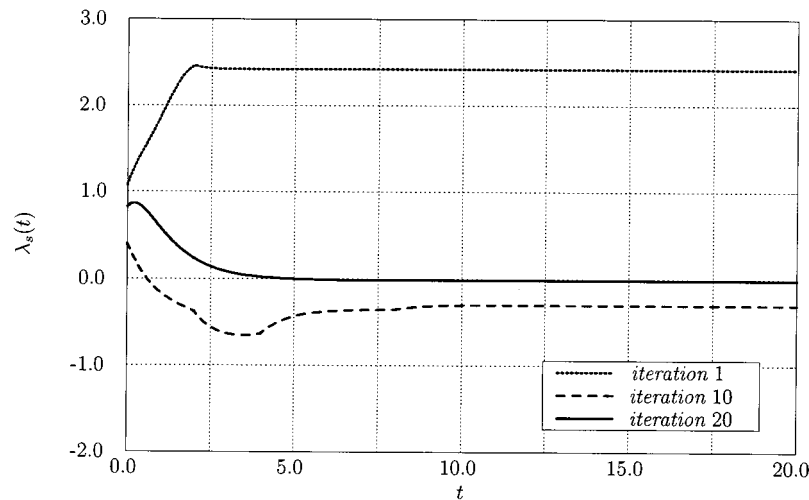


Figure 5. Solution iterates of  $\lambda_s(t)$  vs.  $t$  for initial boundary-layer of the example problem of equation (9) using Algorithm 2

## 7. DISCUSSION OF METHOD

The example problem conforms to Requirement 1; in a region of the  $(x, \lambda)$  plane sufficiently large to contain the composite solution and the iterates that led to it, the eigenvalues of  $J(p)$  are real and non-zero. Furthermore, the fact that all solution iterates level off shows that the hyper-sensitivity has been eliminated over the time interval of interest.

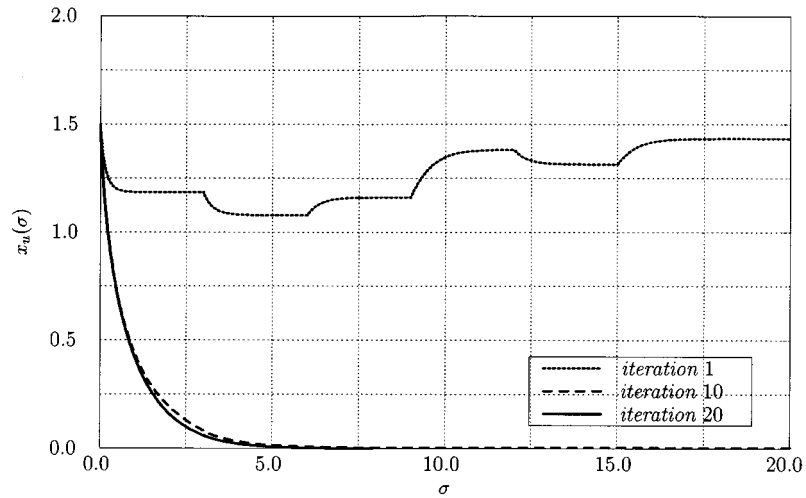


Figure 6. Solution iterates of  $x_u(\sigma)$  vs.  $\sigma$  for the final boundary-layer of the example problem of equation (9) using Algorithm 2

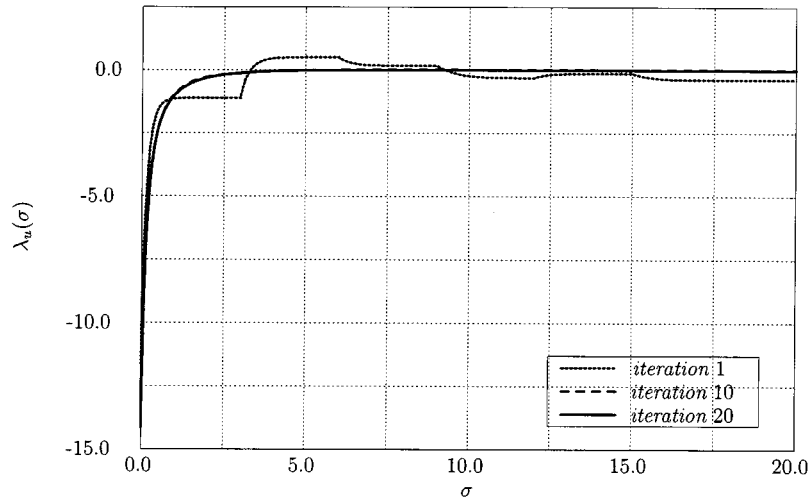


Figure 7. Solution iterates of  $\lambda_u(\sigma)$  vs.  $\sigma$  for the terminal boundary-layer of the example problem of equation (9) using Algorithm 2

Compared with a constant basis, the piecewise constant basis better approximates the local contracting and expanding directions. Algorithm 2 was first applied to obtain  $p_s$  and  $p_u$  using a constant eigenvector basis from  $J(p_0)$  (results not presented). It was found that Algorithm 1 did not converge for  $p_s$ , while the convergence of Algorithm 1 in computing  $p_u$  was extremely slow (of the order of 1000 iterations for a constant basis as opposed to the 20 iterations using a piecewise constant basis).



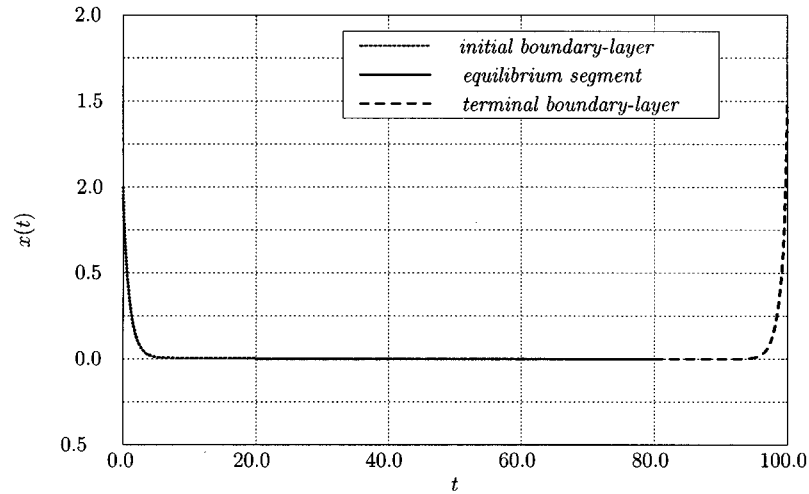


Figure 8. Composite trajectory for  $t_f = 100$  of  $x(t)$  vs.  $t$  for the example problem of equation (9) using the converged initial boundary-layer solution  $x_s(t)$  on the interval  $t \in [0, 20]$ , the equilibrium solution  $\bar{x}$  on the interval  $t \in [20, 80]$ , and the terminal boundary-layer solution  $x_u(t)$  on the segment  $t \in [80, 100]$ .

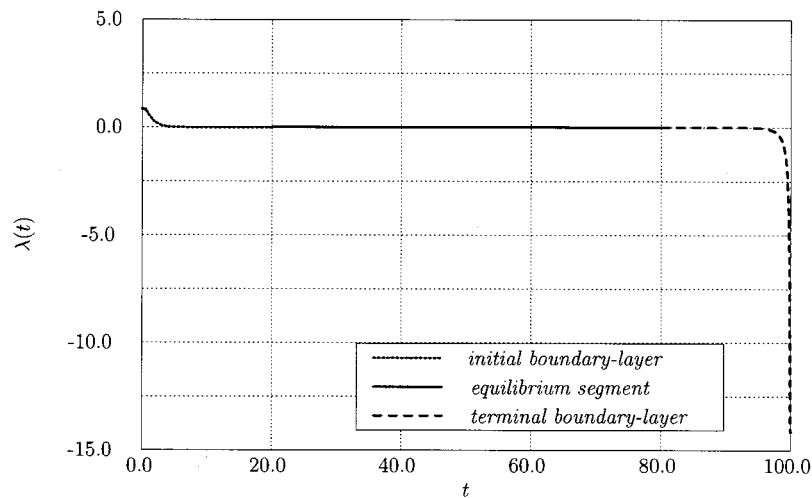


Figure 9. Composite trajectory for  $t_f = 100$  of  $\lambda(t)$  vs.  $t$  for the example problem of equation (9) using the converged initial boundary-layer solution  $\lambda_s(t)$  on the interval  $t \in [0, 20]$ , the equilibrium  $\bar{\lambda}$  on the interval  $t \in [20, 80]$ , and the terminal boundary-layer solution  $\lambda_u(t)$  on the segment  $t \in [80, 100]$ .

Comparing the required basis vector switch times with the curvature of the stable and unstable manifolds (see Figure 3), we find a direct correlation. The basis needs to be updated more frequently when the manifold being tracked is changing direction in the phase space, suggesting that the switching strategy should incorporate information about the curvature of the manifold. The development of such a switching strategy is beyond the scope of this paper.

Requirement 1 is not generally met by completely hyper-sensitive HBVPs. For the example system, there are regions of the phase space where the requirement is not met. Hence, using local eigenvectors is not always applicable. The dichotomic basis method can still be applied, but the approximate dichotomic basis must be determined differently.

One needs some *a priori* knowledge that an optimal control problem is completely hyper-sensitive before the dichotomic basis method should be applied. This knowledge can come from a solution obtained by a direct method. If the solution has the characteristic take-off, cruise, landing structure, then the dichotomic basis method is applicable. Also, the particular saddle point that is influencing the nature of the solution can be identified. Completely hyper-sensitive HBVPs can involve multiple saddle points and (hetero-clinic) orbits that connect them. The solution from a direct method can be used to identify such structure and the basic approach considered in this paper can be adapted to handle it.

A motivating reason for using the dichotomic basis method is to gain insight into the phase space manifold structure in the neighbourhood of the optimal solution. A dichotomic basis provides this information. It identifies the contracting and expanding subspaces at points in the phase space. It provides conditions satisfied by points on the stable and unstable manifolds. With an approximate dichotomic basis, the HBVP can be solved by our iterative algorithm, but the information on the manifold structure is only approximate. Once the approximate solution has been determined, the size of the  $h_u$  component of  $G$  along the initial boundary-layer segment and the size of the  $h_s$  component of  $G$  along the terminal boundary-layer segment indicate how accurately the basis approximates a dichotomic basis. If the accuracy is not sufficient, the eigenspaces at  $\bar{p}$  can be propagated along the initial and terminal boundary-layer segments using a differential Riccati equation [8]. Another option is to compute a dichotomic basis at one or more points along the approximate solution using Lyapunov exponents and the associated direction vectors [20].

## 8. CONCLUSIONS

A version of the dichotomic basis method for solving completely hyper-sensitive Hamiltonian boundary-value problems has been developed in which local eigenvectors are used to construct an approximate dichotomic basis. An iterative algorithm was given for computing the initial and terminal boundary-layer segments which when combined with an equilibrium segment form a composite approximate solution. The method was illustrated on a simple example.

Although the class of completely hyper-sensitive problems is restrictive, it is an important step to solving the large class of multiple time-scale optimal control problems with fast boundary and interior layers. A dichotomic basis can be used to compute composite approximate solutions for this class of problems. The challenge is to develop a way to generate a sufficiently accurate approximation to a dichotomic basis. Constructing an approximate dichotomic basis from local eigenvectors is one approach, but it is not always applicable; other methods must be developed.

## ACKNOWLEDGEMENTS

This research was sponsored by the National Science Foundation and the California Space Institute. The authors would like to thank Professor S.-H. Lam of Princeton University who provided valuable insight during the course of this research. The authors would also like to thank Dr. John T. Betts of the Boeing Company for allowing us to use his Sparse Optimal Control Software code for comparison.

## REFERENCES

1. Betts JT. Survey of numerical methods for trajectory optimization. *Journal of Guidance Control and Dynamics* 1998; **21**(2):193–207.
2. Betts JT, Huffman WP. Sparse Optimal Control Software, SOCS. *Mathematics and Engineering Analysis Library Report, MEA-LR-085*, Boeing Information and Support Services, P.O. Box 3797, Seattle, WA, 98124-2297, 15 July 1997.
3. Betts JT, Frank PD. A sparse nonlinear optimization algorithm. *Journal of Optimization Theory and Applications* 1994; **82**(3).
4. Vlases WG, Paris SW, Lajoie RM, Martens PJ, Hargraves CR. Optimal trajectories by implicit simulation. *Technical Report WRDC-TR-90-3056, Wright-Patterson Air Force Base*, Boeing Aerospace and Electronics, 1990.
5. Kokotovic PV, Khalil HK, O'Reilly J. *Singular Perturbation Methods in Control: Analysis and Design*. Academic Press: San Diego, 1986.
6. Bryson AE, Ho YC. *Applied Optimal Control*. Hemisphere: New York, 1975.
7. Oberle, HJ, Grimm W. BNDSCO—A program for the numerical solution of optimal control problems. *Internal Report No. 515-89/22*, Institute for Flight Systems Dynamics, DLR, Oberpfaffenhofen, Germany, 1989.
8. Rao AV, Mease KD. Dichotomic basis approach to solving hyper-sensitive optimal control problems. *Automatica* 1999; **35**(4):633–642.
9. Kelley HJ. Aircraft maneuver optimization by reduced-order approximation. In *Control and Dynamic Systems*, vol. 10. Leondes CT (ed.), 1973; 131–178.
10. Ardema MD. An introduction to singular perturbations in non-linear control. In *Singular Perturbations in Systems and Control*, Ardema MD (ed.), 1983; 1–92.
11. Anderson BDO, Kokotovic PV. Optimal control problems over large time intervals. *Automatica* 1987; **23**(3):355–363.
12. Lam SH. Using CSP to understand complex chemical kinetics. *Combustion Science and Technology* 1993; **89**(5–6): 375–404.
13. Lam SH, Goussis DA. The CSP method of simplifying kinetics. *International Chemical Kinetics* 1994; **26**:461–486.
14. Mease KD. Geometry of computational singular perturbations. *Proceedings of the IFAC Nonlinear Control Design Symposium*, Tahoe City, 1995.
15. Ardema MD. Computational singular perturbation method for optimal control. *Proceedings of the American Control Conference*, San Diego, 1990.
16. Rao AV, Mease KD. A new method for solving optimal control problems. *Proceedings of the AIAA Guidance, Navigation and Control Conference*, Baltimore, 1995; 818–825.
17. Rao AV. Extension of the computational singular perturbation method to optimal control. *PhD Dissertation*, Princeton University, 1996.
18. Guckenheimer J, Holmes P. *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer: New York, 1990.
19. Mease KD. An approach to solving two time-scale trajectory optimization problems. *IFAC Workshop on Control Applications of Optimization*, Haifa, Israel, December 1995.
20. Bharadwaj S, Wu M, Mease KD. Identifying time-scale structure for simplified guidance law development. *AIAA Guidance, Navigation, and Control Conference*, New Orleans, 1997.
21. Seywald H, Cliff EM, Well KH. Range optimal trajectories for an aircraft flying in the vertical plane. *Journal of Guidance Control and Dynamics* 1994; **17**(2):398–398.
22. Seywald H. Identification of the optimal switching structure for range-optimal atmospheric flight trajectories. *Optimal Control Applications and Methods* 1997; **18**(3):159–177.
23. Arnol'd VI. *Ordinary Differential Equations*. Springer: Berlin, 1992.
24. Rao AV, Mease KD. Minimum time-to-climb trajectories using a modified sweep method. *Proceedings of the AIAA Guidance, Navigation and Control Conference*, Baltimore, 1995; 826–833.