Brief Paper

Dichotomic basis approach to solving hyper-sensitive optimal control problems[1]

A.V. Rao[a],*, K.D. Mease[b,2]

[a] *Department of Mechanical and Aerospace Engineering, Princeton University, USA*
[b] *Department of Mechanical and Aerospace Engineering, University of California, Irvine, USA*

Abstract

As a step toward developing a general method for determining the underlying geometric structure of two time-scale optimally controlled nonlinear systems, we define a degenerate class of two time-scale optimal control problems, called completely hyper-sensitive problems, and propose an indirect solution method for this class of problems. The method uses a dichotomic basis to split the Hamiltonian vector field into its stable and unstable components. An accurate approximation to the optimal solution is constructed by matching the initial and terminal boundary-layer segments with the equilibrium solution. A variation of the method for the case of an approximate dichotomic basis is also developed and is applied to a nonlinear spring-mass problem. The challenging problem of determining a dichotomic basis or a sufficiently accurate approximation to one is discussed only briefly, but some potential solutions are identified. © 1999 Elsevier Science Ltd. All rights reserved.

Keywords: Optimal control; Dichotomic transformation; Hamiltonian systems

1. Introduction

In this paper, we are concerned with *nonlinear* optimal control problems and are motivated not only by the desire to obtain a solution for a particular set of boundary conditions, but also to gain insight into the structure of the optimally controlled system. The methods for solving optimal control problems are usually classified as direct or indirect. Indirect methods involve determining extremals by solving the Hamiltonian boundary-value problem (HBVP) posed by the first-order optimality conditions. Hamiltonian systems have the property that a bounded, open set of initial conditions in the state-costate space maintains a constant volume when propagated forward in time according to the Hamiltonian differential equations. If the set contracts in some directions, it expands in an equal number of other directions. We say an optimal control problem and its corresponding HBVP are *completely hyper-sensitive*, if the time interval of interest is very long relative to the minimum rate of contraction and expansion of the Hamiltonian system in the neighborhood of the optimal solution to the HBVP. Completely hyper-sensitive HBVPs are a degenerate class of two time-scale HBVPs. For a non-degenerate two time-scale problem, the rates of contraction and expansion are only fast in some directions and the problem is only *partially hyper-sensitive*. For two time-scale problems of the boundary-layer type, the generic solution is composed of a short duration initial boundary-layer segment dominated by fast contracting behavior, a long duration intermediate segment dominated by slow behavior, and a short duration terminal boundary-layer segment dominated by fast expanding behavior.

For a two time-scale HBVP, the terminal boundary-layer segment depends hyper-sensitively on the initial conditions. Regarding numerical computation by single shooting with a specified accuracy on the terminal boundary-layer solution, this hyper-sensitivity can

---

require initial condition accuracy that is difficult to achieve or may even exceed the available precision. Using a non-uniform node distribution with high density in the boundary-layers, indirect multiple shooting or direct methods may make numerical solution to the specified accuracy feasible. However, even if these methods are capable of producing a solution with the specified accuracy, they do not produce information regarding the structure of the two time-scale optimally controlled system. In the special case where the two time-scale Hamiltonian dynamics are represented in the *standard* singularly perturbed form, there is a systematic procedure, called the singular perturbation method, to decompose the dynamical model into reduced-order models for the slow behavior and the fast behavior. The solution to the HBVP is then constructed by matching asymptotic expansions (see Kelley, 1973; Ardema, 1983; Kokotovic et al. 1986). The reduced-order models also facilitate the analysis and design of control systems and increase the likelihood of developing high performance control laws in feedback form (e.g., Calise, 1976; Kokotovic et al., 1986).

Two time-scale optimal control problems are not usually formulated so that the resulting HBVP defining extremals is in standard singularly perturbed form, since doing so requires significant *a priori* knowledge of the time-scale structure. Some procedures, specific to flight mechanics, for transforming state equations to standard form have been proposed (see Ardema and Rajan, 1985). Marino and Kokotovic (1988) propose a method of transforming a general two time-scale control system into singularly perturbed form, but their approach depends on identifying a priori a small parameter that characterizes the time-scale separation.

Our objective is to develop an indirect solution method for nonlinear two time-scale optimal control problems, given in general form — a solution method based on the underlying geometric structure of the trajectories of the Hamiltonian system in the state–costate space — but not requiring a priori knowledge of this structure. Two capabilities are required: (i) splitting the Hamiltonian system into slow and fast parts and (ii) splitting the fast part into contracting and expanding parts. If the system model is given in standard form, the singular perturbation method provides the first capability and reduces the second requirement to splitting the fast part into stable and unstable parts in the neighborhood of an equilibrium point. When the fast (boundary-layer) dynamics are linear, or can be approximated as such, eigenvalues and eigenvectors can be used to split the stable and unstable parts. A dichotomic transformation serves the same purpose (see Wilde and Kokotovic, 1972; Chow, 1979).

In this paper, we present a method for solving completely hyper-sensitive HBVPs. The presentation is primarily conceptual. A specific algorithm is given and illustrated on a simple example, but it is not analysed. Additional development of and computational experience with the dichotomic basis method is described in Rao (1996) and Rao and Mease (1998). The method uses a dichotomic transformation to decouple the contracting and expanding behavior. The solution to the HBVP is constructed by forward integration of the contracting part of the dynamics and backward integration of the expanding part of the dynamics. In this way, error amplification is avoided. The method is inspired by the computational singular perturbation (CSP) method of Lam (1993) and Lam and Goussis (1994). A geometric interpretation of CSP consistent with the perspective of the present paper is given by Mease (1995a). The application of CSP to optimal control problems was previously considered by Ardema (1990). The use of dichotomic transformations for the solution of linear boundary-value problems is well known in the numerical analysis literature (e.g. O'Malley and Anderson, 1982; Ascher et al., 1995). Wilde and Kokotovic (1972) applied dichotomic transformations to linear optimal control problems. Chow (1979) used dichotomic transformations for nonlinear two time-scale HBVPs in standard form with linear boundary-layer dynamics. We use the geometric characterization of two time-scale systems by Fenichel (1979) and the stable and unstable sub-bundles of Sacker and Sell (1980) to extend the use of dichotomic transformations to nonlinear HBVPs. Although we only present a method for completely hyper-sensitive problems here, our approach has the potential for extension to the general two time-scale case as outlined by Mease (1995b). Another method for solving a completely hyper-sensitive optimal control problem is given by Anderson and Kokotovic (1987). They construct an approximate solution from the positive and negative definite solutions to the steady-state Hamilton–Jacobi–Bellman equation. Their method does not appear to be extendible to the two time-scale case, unless the system model is given in standard form.

2. Hamiltonian boundary value problem

Consider the optimal control problem of finding the control u that steers the state x from $x(0) = x_0$ to $x(t_f) = x_f$ in accord with the system evolution equation

$$\dot{x} = f(x, u) \tag{1}$$

and minimizes the scalar cost

$$J = \int_0^{t_f} L(x, u) \, dt. \tag{2}$$

Let $x(t) \in \mathbb{R}^n$ and $u(t) \in \mathbb{R}^m$. The first-order necessary conditions for a local minimum lead to a *Hamiltonian boundary-value problem* (HBVP) for the extremal

trajectories. The HBVP is composed of the differential equations

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} H_\lambda^{\mathrm{T}}(x, \lambda) \\ -H_x^{\mathrm{T}}(x, \lambda) \end{bmatrix} \tag{3}$$

and the boundary conditions

$$x(0) = x_0, \quad x(t_f) = x_f, \tag{4}$$

where $\lambda(t) \in \mathbb{R}^n$ is the costate, $H(x, \lambda) = L(x, u(x, \lambda)) + \lambda^{\mathrm{T}} f(x, u(x, \lambda))$ is the optimal Hamiltonian with $u(x, \lambda)$ the value of the optimal control function at the point (x, λ), and H_λ and H_x are partial derivatives of H with respect to the subscripted variable. We assume that H is a C^2 function of x and λ.

The x-space and the (x, λ)-space are called the *state space* and the *phase space*, respectively. The vector field of Eq. (3) is a *Hamiltonian vector field*. For a function of time, such as the state x, we use the notation $x(\cdot)$ to denote the function on the time interval of interest and $x(t)$ to denote its value at time t.

Our attention is focused on the HBVP, its solution that is the minimizing solution of the optimal control problem, and the structure of this solution trajectory and its neighbors in the phase space. We assume that there exists a unique solution to the optimal control problem (see Anderson and Kokotovic (1987) for existence theory paralleling that for time-invariant linear quadratic problems for a certain class of completely hyper-sensitive optimal control problems). Although the specifics of the approach to be discussed are given for the HBVP corresponding to the above form of optimal control problem, the approach is applicable to completely hyper-sensitive HBVPs corresponding to other forms of optimal control problems as well.

3. Supporting theory

In this section we present the terminology, concepts and theory that serve as the basis of our solution method. To simplify notation, we sometimes use $p = (x, \lambda)$, $\delta p = (\delta x, \delta \lambda)$, and $v_H = (H_\lambda, -H_x)$, and we use P to represent the Euclidean phase space $\mathbb{R}^{2n}$. Because we want to consider the structure beyond that associated with an individual trajectory, we need to introduce some geometric objects (see Boothby, 1986). The *tangent space* at a point p of the phase space P is the space of tangent vectors to all possible C^1 trajectories through p. The tangent space at $p \in P$ is a linear vector space isomorphic to $\mathbb{R}^{2n}$ and is denoted by $T_p P$. The *tangent bundle* is $TP = \bigcup_{p \in P} T_p P$, i.e., the collection of all the tangent spaces associated with points in P. A point in the tangent bundle is denoted by $(p, \delta p)$ where δp is the $2n$-dimensional vector in the tangent space with base point p, i.e.,

$\delta p \in T_p P$. Since the phase space is $\mathbb{R}^{2n}$, we commit to a particular coordinate representation of the state and costate and thus of p. We use δp to denote the coordinate representation of a tangent vector in the basis induced on $T_p P$ by the coordinate basis for P.

3.1. Hyper-sensitivity in a nonlinear Hamiltonian system

A consequence of the Hamiltonian form of the vector field in Eq. (3) is that

$$\mathrm{div}(v_H) = \mathrm{div} \begin{bmatrix} H_\lambda^{\mathrm{T}} \\ -H_x^{\mathrm{T}} \end{bmatrix} = \sum_{i=1}^{n} \left[\frac{\partial}{\partial x_i} \frac{\partial H}{\partial \lambda_i} - \frac{\partial}{\partial \lambda_i} \frac{\partial H}{\partial x_i} \right] = 0. \tag{5}$$

From Liouville's theorem (e.g., Arnol'd, 1992), it follows that the flow corresponding to the vector field of Eq. (3) preserves volume. We can write the solution to Eq. (3) as $p(t) = \phi(t, p(0))$, where the function ϕ is the *flow* corresponding to the Hamiltonian vector field and is a time-dependent mapping of points in $\mathbb{R}^{2n}$ to points in $\mathbb{R}^{2n}$. Consider a bounded, open, connected set $S(0) \subset \mathbb{R}^{2n}$ of initial phase points. If we propagate each point in $S(0)$ ahead t units of time according to the flow, we obtain a new set $S(t)$ given explicitly by

$$S(t) = \{ p \in \mathbb{R}^{2n} : p = \phi(t, p_0) \text{ for some } p_0 \in S(0) \}. \tag{6}$$

The set $S(t)$ will have the same volume as $S(0)$, but may have a much different shape. A property of a Hamiltonian flow is that, if the flow contracts S in certain directions, then it expands S in an equal number of other directions. It is this last property that can make a HBVP difficult to solve numerically.

Consider the numerical computation of a solution to a particular HBVP. All the elements of the state and costate are not known at either the initial time or the final time. Since we are solving the problem numerically, it only makes sense to ask that initial and final phase vectors be determined to within some error. For example we may require that the initial and final phase vectors each lie within a ball of radius TOL of the exact value. The difficulty of achieving such an error tolerance increases with the degree of contraction and expansion over the time interval for the problem. If we attempt to determine the unknown elements of the phase vector at the initial time, the unknown elements corresponding to expanding directions must be determined more accurately than the TOL-ball at the initial time would suggest, so that when they expand during the mapping to the final time, the phase vector will lie within the TOL-ball at the final time. The greater the rate of expansion is, the greater the required accuracy in the unknown initial conditions corresponding to the expanding directions. This characterization, albeit qualitative, identifies the potential for hyper-sensitivity in nonlinear HBVPs.

3.2. Dichotomy and stable/unstable decomposition of the induced linear flow

By considering the linearized motion with respect to a given solution of Eq. (3), one can characterize the contracting and expanding directions of the Hamiltonian system by applying the theory of Sacker and Sell (1980). The Hamiltonian vector field v_H on the phase space induces a linear Hamiltonian flow on the tangent bundle. A tangent vector $\delta p \in T_p P$ is mapped during time interval t to the tangent vector $\Phi(t, p)\delta p \in T_{\phi(t,p)}P$, where $\phi(t, p)$ and $\Phi(t, p)$ comprise the solution to the initial value problem

$$\dot{\phi} = v_H(\phi), \qquad \phi(0, p) = p, \tag{7}$$

$$\dot{\Phi} = \frac{\partial v_H}{\partial p}(\phi)\Phi, \qquad \Phi(0, p) = I. \tag{8}$$

Definition 1. A mapping $\hat{\Pi} : TP \to TP$ is a projector if

(i) $\hat{\Pi}$ is continuous, and
(ii) for each $p \in P$, there is a linear projection $\Pi(p)$ (i.e., $\Pi^2(p) = \Pi(p)$) in $T_p P$ such that $\hat{\Pi}(p, \delta p) = (p, \Pi(p)\delta p)$ for all $(p, \delta p) \in TP$.

If $\hat{\Pi}$ is a projector on TP, then

$$\begin{aligned} \text{range}(\hat{\Pi}) &= \mathfrak{R} = \{(p, \delta p) \in TP : \Pi(p)\delta p = \delta p\}, \\ \text{nullspace}(\hat{\Pi}) &= \aleph = \{(p, \delta p) \in TP : \Pi(p)\delta p = 0\}, \end{aligned} \tag{9}$$

are complementary sub-bundles of TP, i.e., $\mathfrak{R}(p) \cap \aleph(p) = \{0\}$ and $T_p P = \mathfrak{R}(p) + \aleph(p)$, $\forall p \in P$, where $\mathfrak{R}(p)$ and $\aleph(p)$ denote fibers of the sub-bundles. Conversely, if $\mathscr{V}_1$ and $\mathscr{V}_2$ are complementary sub-bundles in TP, then there is a unique projector $\hat{\Pi}$ on TP such that $\mathfrak{R} = \mathscr{V}_1$ and $\aleph = \mathscr{V}_2$.

Let $Q \subset P$ be a smooth, compact, connected sub-manifold of P, and let Q be invariant with respect to the Hamiltonian flow ϕ.

Definition 2. The induced linear Hamiltonian flow on the tangent bundle TQ admits an *exponential dichotomy* if there is a projector $\hat{\Pi}_D$ on TQ, where $\Pi_D(p)$ has rank n on Q, and positive constants K and α such that

$$\begin{aligned} \|\Phi(t, p)\Pi_D(p)\Phi^{-1}(\tau, p)\| &\leq Ke^{-\alpha(t-\tau)}, \quad t \geq \tau, \\ \|\Phi(t, p)(I - \Pi_D(p))\Phi^{-1}(\tau, p)\| &\leq Ke^{-\alpha(\tau-t)}, \quad t \leq \tau \end{aligned} \tag{10}$$

for all $p \in Q$ (Sacker and Sell, 1980). The notation $\|\cdot\|$ denotes the norm induced by the Euclidean norm on P.

A dichotomic projector $\hat{\Pi}_D$ allows the splitting of TQ into invariant, complementary stable and unstable sub-bundles

$$\text{range}(\hat{\Pi}_D) = \mathscr{V}_s = \{(p, \delta p) \in TQ : \|\Phi(t, p)\delta p\| \to 0 \\ \text{as } t \to +\infty\},$$

$$\text{nullspace}(\hat{\Pi}_D) = \mathscr{V}_u = \{(p, \delta p) \in TQ : \|\Phi(t, p)\delta p\| \to 0 \text{ as} \\ t \to -\infty\}. \tag{11}$$

The condition that the base space is compact and invariant is too restrictive for our needs. We want to apply the above characterization to a finite neighborhood of a saddle-type equilibrium, yet the set consisting of the equilibrium point is the only compact, invariant set in this neighborhood. The above results would only serve to characterize the tangent space at the equilibrium point as the direct sum of an n-dimensional stable eigenspace and an n-dimensional unstable eigenspace. By allowing the base space Q to be an "overflowing invariant" manifold with a boundary, a segment of the stable manifold (with time reversed) or the unstable manifold, that includes the equilibrium point, can be taken as the base space (see Fenichel, 1979). We go a step further and modify the dichotomy definition such that the base space can be a non-invariant, bounded, open set. In particular, let $\Omega \subset P$ be a neighborhood of the optimal phase trajectory, composed of extremals of duration t_f or longer.

Definition 3. The linear Hamiltonian flow on the tangent bundle $T\Omega$ induced by v_H admits a *finite-time exponential dichotomy* if

(i) there are positive constants K and α and a projector $\hat{\Pi}_D$, such that, for all $p \in \Omega$, $\text{rank}\Pi_D(p) = n$ and the bounds in Eq. (10) are satisfied for all values of τ and t in the time interval $I(p)$ – the longest time interval such that $\phi(s, p) \in \Omega$ for all $s \in I(p)$.
(ii) the exponential bounds are reasonably tight uniformly on $T\Omega$.
(iii) t_f is larger than several times the reciprocal of the average minimum rate of contraction and expansion, whose estimate is $1/\alpha$.

The notion of a *finite time* exponential dichotomy is used in numerical analysis for linear time-varying systems Ascher et al., 1995). Definition 3 is our modification for an induced linear flow on a tangent bundle. On a finite time interval, any bounded behavior, whether contracting or expanding, can be exponentially bounded with either a decreasing or increasing exponential by simply choosing a sufficiently large value of K. Stipulation (ii) ensures that the local minimum rate of contraction and expansion is not too different from the average minimum rate of contraction and expansion α. (This is important for our application even in the case of a compact, invariant base space for which Definition 2 is appropriate.) Stipulation (iii) ensures that the minimum rate of contraction and expansion is large enough that a dichotomic splitting is effectively a stable/unstable splitting on $T\Omega$. Consequently, we will use the terms stable and unstable sub-bundles in connection

with a finite-time exponential dichotomy, even though finite-time intervals are involved.

Consider a linear coordinate change $(p, \delta p) \mapsto (p, v)$ on $T\Omega$ with $\delta p = D(p)v$, where v is the new coordinate vector and D is a nonsingular and continuously differentiable $2n \times 2n$ matrix-valued function of p on Ω. The columns of $D(p)$ are basis vectors for $T_p\Omega$ for each $p \in \Omega$. In terms of v, the variational equation of Eq. (8) is

$$\dot{v} = [D^{-1}JD - D^{-1}\dot{D}]v = \Lambda(\phi(t, p))v, \qquad (12)$$

where $J = \partial v_H/\partial p$ is the local Jacobian matrix, $\dot{D}$ is obtained by taking the Lie derivative, in the v_H direction, of each element of D, and Λ is used to denote the transformed system matrix more concisely.

Definition 4. Concerning the case of a finite-time exponential dichotomy for the linear flow on $T\Omega$, a basis D is a *dichotomic basis*, if it satisfies the following two requirements.

(i) The corresponding system matrix Λ is block-triangular of the form

$$\Lambda = \begin{bmatrix} \Lambda_s(p) & \Lambda_{su}(p) \\ 0 & \Lambda_u(p) \end{bmatrix} \qquad (13)$$

for all $p \in \Omega$, and the matrices Λ_s and Λ_u are both $(n \times n)$ dimensional.

(ii) The transition matrices $\Phi_s(t, p)$ and $\Phi_u(t, p)$ corresponding to Λ_s and Λ_u, defined such that $\Phi_s(0, p) = I$ and $\Phi_u(0, p) = I$ satisfy the inequalities

$$\|\Phi_s(t, p)\Phi_s^{-1}(\tau, p)\| \leq K_1 \|D(\phi(\tau, p))\| \|D^{-1}(\phi(t, p))\| e^{-\alpha(t-\tau)},$$
$$t \geq \tau,$$

$$\|\Phi_u(t, p)\Phi_u^{-1}(\tau, p)\| \leq K_1 \|D(\phi(\tau, p))\| \|D^{-1}(\phi(t, p))\| e^{-\alpha(\tau-t)},$$
$$t \leq \tau, \qquad (14)$$

where τ and t are in $I(p)$ as defined in Definition 3 and K_1 is a positive constant that can be related (Ascher et al., 1995) to K in Definition 3.

The subscripts "s" and "u" indicate the stable and unstable nature of the two subsystems. A dichotomic basis can be split in the form

$$D(p) = [D_s(p) \ D_u(p)], \qquad (15)$$

where

$$D_s(p) = [d_1(p) \ldots d_n(p)] \text{ and } D_u(p) = [d_{n+1}(p) \ldots d_{2n}(p)].$$

Correspondingly, the new coordinate vector splits as

$$v = \begin{bmatrix} v_s \\ v_u \end{bmatrix}. \qquad (16)$$

A point in the stable sub-bundle $\mathscr{V}_s$ has the form $(p, (v_s, 0))$. The fiber $\mathscr{V}_s(p)$ of the stable sub-bundle is given by the column span of $D_s(p)$. Note that the stable sub-bundle is invariant under the dynamics, since an initial

point of the form $(p, (v_s(0), 0))$ maps to a point $(\phi(t, p), (v_s(t), 0)) = (\phi(t, p), (\Phi_s(t)v_s(0), 0))$ and remains in the stable sub-bundle. We could go a step further and require a dichotomic basis to yield $\Lambda_{su} = 0$, i.e., Λ block diagonal. Then $\mathscr{V}_u(p)$ would be given by span $D_u(p)$. But this property is not needed.

The Hamiltonian vector field in Eq. (3) assigns to each phase point $p = (x, \lambda)$ a vector $v_H(p) = (H_\lambda(x, \lambda), -H_x(x, \lambda))$ in the tangent space at that point. At each phase point, we can express this particular tangent vector in the dichotomic basis as

$$v_H(p) = \begin{bmatrix} H_\lambda^T(x, \lambda) \\ -H_x^T(x, \lambda) \end{bmatrix} = D_s(x, \lambda)v_s + D_u(x, \lambda)v_u, \qquad (17)$$

where the tangent vector coordinates v_s and v_u are determined by

$$v_s = h_s(x, \lambda) = D_s^{-1}(x, \lambda) \begin{bmatrix} H_\lambda^T(x, \lambda) \\ -H_x^T(x, \lambda) \end{bmatrix},$$

$$v_u = h_u(x, \lambda) = D_u^{-1}(x, \lambda) \begin{bmatrix} H_\lambda^T(x, \lambda) \\ -H_x^T(x, \lambda) \end{bmatrix}. \qquad (18)$$

and D_s^{-s} and D_u^{-1} denote the first and second n rows of D^{-1}.

Two alternatives to integrating Eq. (3) can be identified to compute the extremal for a given initial condition (x_0, λ_0). One is to replace the right-hand side of Eq. (3) with the dichotomic basis representation given in Eqs. (17) and (18). The second is to integrate the system

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = D_s(x, \lambda)v_s + D_u(x, \lambda)v_u$$

$$\begin{bmatrix} \dot{v}_s \\ \dot{v}_u \end{bmatrix} = \begin{bmatrix} \Lambda_s(x, \lambda) & \Lambda_{su}(x, \lambda) \\ 0 & \Lambda_u(x, \lambda) \end{bmatrix} \begin{bmatrix} v_s \\ v_u \end{bmatrix} \qquad (19)$$

with the initial conditions

$$x(0) = x_0, \qquad \lambda(0) = \lambda_0,$$

$$v_s(0) = h_s(x_0, \lambda_0), \qquad v_u(0) = h_u(x_0, \lambda_0). \qquad (20)$$

The second alternative makes clear that a property of a dichotomic basis is that, for a given initial condition on the state, $x(0) = x_0$, the unstable component of the Hamiltonian vector field can be suppressed at $t = 0$ by choosing $\lambda(0) = \lambda_0$ such that $h_u(x_0, \lambda_0) = 0$; and that the unstable component will remain suppressed, due to the lack of coupling from v_s to v_u. In other words, we choose λ_0 such that we are at a point in the phase space on the $x = x_0$ hyper-plane where the Hamiltonian vector field lies in the fiber $\mathscr{V}_s(x_0, \lambda_0)$ of the stable sub-bundle.

3.3. Geometric structure associated with a completely hyper-sensitive HBVP

In the phase space the solution to a two time-scale HBVP and its neighbors are organized by a normally

hyperbolic slow invariant manifold (see Fenichel, 1979). Each trajectory on the slow manifold is the transverse intersection of fast-stable and fast-unstable manifolds. Except in the initial and terminal layers, the solution of interest is well approximated by a trajectory on the slow manifold referred to as the reduced-order trajectory. The complete solution trajectory begins slightly off the fast-stable manifold of the reduced-order trajectory and follows it quickly towards the slow manifold. The trajectory then progresses slowly alongside the reduced-order trajectory. Near the final time, the trajectory quickly follows the unstable manifold of the reduced-order trajectory to the terminal condition slightly off the unstable manifold. For fixed boundary conditions on the state, as the length of the time interval increases, the initial and terminal boundary-layer segments of the trajectory lie closer and closer to the fast-stable and fast-unstable manifolds of the reduced-order trajectory, respectively.

The completely hyper-sensitive class of two time-scale HBVPs is degenerate in that the slow manifold is an isolated saddle-type equilibrium point. Moreover, the fast-stable and fast-unstable manifolds are simply the n-dimensional stable and unstable manifolds of the saddle point. The stable manifold is composed of all the phase trajectories that approach the equilibrium asymptotically in forward time, while the unstable manifold is composed of all the phase trajectories that approach the equilibrium asymptotically in backward time. The stable and unstable eigenspaces of the linearized Hamiltonian system at the saddle point are n-dimensional hyperplanes and are tangent to the stable and unstable manifolds at the equilibrium point. (See Guckenheimer and Holmes, 1990 for precise definitions of stable and unstable manifolds and additional information.) It follows from this geometric characterization of a completely hyper-sensitive optimal control problem that there exists a neighborhood $\Omega \subset P$ of the solution of interest that contains portions of the stable and unstable manifolds, including their intersection point at the equilibrium, that admits a finite-time exponential dichotomy. Consequently, the tangent bundle $T\Omega$ can be split into a stable sub-bundle $\mathcal{V}_s$ and a complementary sub-bundle. We assume that Ω is small enough that complications that could arise from the presence of homoclinic or heteroclinic orbits can be ignored.

The stable manifold is positively invariant with respect to the Hamiltonian flow. In other words, a trajectory that begins on the stable manifold will remain on the stable manifold as time progresses. Let $\mathcal{S}$ denote the set of phase points in Ω that are on the stable manifold. The invariance of $\mathcal{S}$ implies that for every $p \in \mathcal{S}$, the Hamiltonian vector field must lie in the tangent space to the stable manifold, denoted $v_H(p) \in T_p\mathcal{S}$. For all $p \in \mathcal{S}$, $T_p\mathcal{S}$ and $\mathcal{V}_s(p)$ are the same subspace of $T_p\Omega$. Thus we have $v_H(p) \in \mathcal{V}_s(p)$ for all $p \in \mathcal{S}$. Representing $v_H(p)$ in a dichotomic basis as in Eq. (17), points $p \in \mathcal{S}$ are identified

as those for which the unstable component of the Hamiltonian vector field is zero, i.e., $v_u = h_u(x, \lambda) = 0$. This can be viewed as a partial equilibrium condition. On the stable manifold the unstable component is zero, whereas at the equilibrium point, both stable and unstable components are zero. It is clear from the block triangular form of the variational equations in Eq. (19) that if the unstable component is zero at some initial point, it will remain zero along the subsequent trajectory.

4. Solution of completely hyper-sensitive HBVPs

We are now in a position to describe a method for solving completely hyper-sensitive nonlinear HBVPs. An approximate solution is constructed by piecing together the extremal on the stable manifold that satisfies the condition on the initial state, the equilibrium solution, and the extremal on the unstable manifold that satisfies the condition on the final state. This composite, approximate solution is essentially identical to that of Anderson and Kokotovic (1987). The composite solution can be continuous by introducing bridging segments, or discontinuities within a specified tolerance can be allowed at the junctions. Our method differs from that of Anderson and Kokotovic in the computation of the boundary-layer segments. We only describe the computation of the initial boundary-layer segment, since the computation of the terminal boundary-layer segment is essentially the same with time reversed. We present two methods: one in which a dichotomic basis is known and one in which an approximate dichotomic basis is known.

4.1. Dichotomic basis method

Assume that we have a dichotomic basis D. We can use it to split the Hamiltonian vector field into its stable and unstable components. The stable manifold is composed of phase points for which the unstable component of the vector field is zero. Thus, we determine λ_0 such that

$$v_u(0) = h_u(x_0, \lambda_0) = 0 \qquad (21)$$

is satisfied for the initial state x_0. A solution exists, but may not be unique. For example, one may find two solutions corresponding to the stable manifolds of two equilibria, only one of which is the correct one. Other considerations must guide the selection of the appropriate solution, if there is more than one. A rigorous treatment of the existence and uniqueness of this solution is beyond the scope of this paper.

The initial boundary-layer solution is then computed by integrating Eqs. (19) to ensure that the trajectory remains on the stable manifold. The integration is continued until the trajectory is close enough to the equilibrium to connect it to the equilibrium solution to within

the specified tolerance (discontinuity is allowed). We denote the duration of the initial boundary-layer by t_{ibl}.

4.2. Approximate dichotomic basis method

Even if a dichotomic basis is not available, it may be possible to solve a completely hyper-sensitive HBVP using an approximate dichotomic basis. When an approximate dichotomic basis is used, the problem must be solved iteratively, but the strategy of piecing together the extremal on the stable manifold that satisfies the condition on the initial state, the equilibrium solution, and the extremal on the unstable manifold that satisfies the condition on the final state, remains the same. As before, only the computation of the initial boundary-layer is described.

An approximate dichotomic basis is denoted by

$$A(p) = [A_s(p) \ A_u(p)] \tag{22}$$

where

$$A_s(p) = [a_1(p) \ldots a_n(p)]$$

and

$$A_u(p) = [a_{n+1}(p) \ldots a_{2n}(p)].$$

We express the Hamiltonian vector field in the approximate dichotomic basis as

$$v_H = \begin{bmatrix} H_\lambda^T \\ -H_x^T \end{bmatrix} = A_s(x, \lambda)v_s + A_u(x, \lambda)v_u, \tag{23}$$

where the coordinates v_s and v_u are determined by

$$
\begin{aligned}
v_s &= q_s(x, \lambda) = A_s^{-1}(x, \lambda) \begin{bmatrix} H_\lambda^T(x, \lambda) \\ -H_x^T(x, \lambda) \end{bmatrix} \\
v_u &= q_u(x, \lambda) = A_u^{-1}(x, \lambda) \begin{bmatrix} H_\lambda^T(x, \lambda) \\ -H_x^T(x, \lambda) \end{bmatrix}
\end{aligned}
\tag{24}
$$

and A_s and A_u approximate D_s and D_u, respectively.

The following iterative procedure is proposed for computing the initial boundary-layer solution using an approximate dichotomic basis.

(i) Initialization: Choose t_{ibl} and $v_u(\cdot)$ (e.g., set $v_u(t) \equiv 0$).
(ii) Generate initial conditions at $t = 0$: given $v_u(\cdot)$ from the initialization or the previous backward integration once available
    (a) Solve $q_u(x_0, \lambda(0)) = v_u(0)$ for $\lambda(0) = \lambda_0$.
    (b) Compute $v_s(0) = v_{s0} = q_s(x_0, \lambda_0)$.
(iii) Forward integration: Given the initial conditions x_0, λ_0, and v_{s0}, integrate from $t = 0$ to $t = t_{ibl}$

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = A_s(x, \lambda)v_s + A_u(x, \lambda)v_u,$$

$$\dot{v}_s = [A_s^{-1}(x, \lambda)J(x, \lambda)A(x, \lambda) - A_s^{-1}(x, \lambda)\dot{A}(x, \lambda)] \begin{bmatrix} v_s \\ v_u \end{bmatrix}.$$

(iv) Generate the conditions at $t = t_{ibl}$ for the backward integration: using $x(t_{ibl}) = x_{ibl}$ and $v_s(t_{ibl})$ from the forward integration
    (a) Solve $q_s(x_{ibl}, \lambda(t_{ibl})) = v_s(t_{ibl})$ for $\lambda(t_{ibl}) = \lambda_{ibl}$.
    (b) Compute $v_u(t_{ibl}) = q_u(x_{ibl}, \lambda_{ibl})$.
(v) Backward integration: Given $v_s(\cdot)$ from forward integration and final conditions $(x_{ibl}, \lambda_{ibl}, v_u(t_{ibl}))$, integrate from $t = t_{ibl}$ to $t = 0$

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = A_s(x, \lambda)v_s + A_u(x, \lambda)v_u,$$

$$\dot{v}_u = [A_u^{-1}(x, \lambda)J(x, \lambda)A(x, \lambda) - A_u^{-1}(x, \lambda)\dot{A}(x, \lambda)] \begin{bmatrix} v_s \\ v_u \end{bmatrix}.$$

(vi) Convergence check: If the difference between the forward and backward trajectories is sufficiently small, then stop. Otherwise, repeat (ii)–(v).

With an approximate dichotomic basis, v_u only *approximates* the unstable component of the Hamiltonian vector field. While v_u is *not* identically zero on the stable manifold, it does decay to zero on the stable manifold, as does v_s. Consequently, both v_s and v_u must be computed. The objective of the iterative procedure is to determine v_s and v_u that steer the phase trajectory in the stable manifold from the specified initial state to the vicinity of the equilibrium point.

To reduce error amplification, the predominately stable component of the Hamiltonian vector field, v_s, is determined by forward integration with the predominately unstable component v_u a fixed function of time; v_u is determined by backward integration with v_s a fixed function of time. The other means of controlling error amplification is the choice of t_{ibl}. On one hand, t_{ibl} must be large enough so that the trajectory on the stable manifold reaches a point close enough to the equilibrium to satisfy the matching tolerance. However, the error amplification grows with increasing t_{ibl}. Large error amplification is particularly problematic in the first iterations because the unstable component of the Hamiltonian vector field may not be well suppressed. One could use a relatively small value of t_{ibl} initially until v_u is better known, and then increase t_{ibl} for the subsequent iterations. The larger t_{ibl} is, the more accurately A_s must approximate D_s. Accuracy can be measured by the angle between the subspaces spanned by the columns of the matrices (see Rao, 1996).

5. Nonlinear spring-mass example

Consider a mass connected to a nonlinear spring as modeled by the equations

$$\dot{x}_1 = x_2, \qquad \dot{x}_2 = -x_1 - x_1^3 + u(t), \tag{25}$$

where x_1 is the position of the mass relative to the position of zero spring force, x_2 is the velocity, and $u(t)$ is

the controllable specific force on the mass. The optimal control problem is to determine the control that drives the system from the initial condition $(x_1(0), x_2(0)) = (1, 0)$ to the final condition $(x_1(t_f), x_2(t_f)) = (0.75, 0)$ and minimizes the cost function

$$J = \frac{1}{2} \int_0^{t_f} (x_1^2 + x_2^2 + u^2) \, dt. \tag{26}$$

The optimal Hamiltonian is given by

$$H = \tfrac{1}{2}(x_1^2 + x_2^2 - \lambda_2^2) + \lambda_1 x_2 + \lambda_2(-x_1 - x_1^3), \tag{27}$$

The final time is $t_f = 40$. The first-order necessary conditions for optimality lead to the HBVP

$$\begin{aligned}
\dot{x}_1 &= x_2, & x_1(0) &= 1, \\
\dot{x}_2 &= -x_1 - x_1^3 - \lambda_2, & x_2(0) &= 0, \\
\dot{\lambda}_1 &= -x_1 + \lambda_2(1 + 3x_1^2), & x_1(t_f) &= 0.75, \\
\dot{\lambda}_2 &= -x_2 - \lambda_1, & x_2(t_f) &= 0.
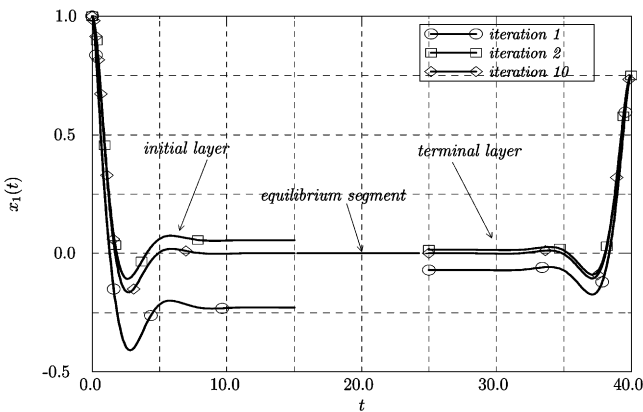\end{aligned} \tag{28}$$

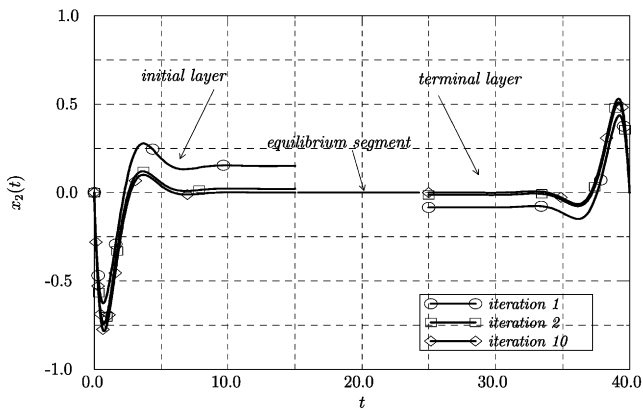Fig. 1. Iterates of $x_1(t)$ versus t for initial and terminal boundary-layers of Eq. (28).

Fig. 2. Iterates of $x_2(t)$ versus t for initial and terminal boundary-layers of Eq. (28).

For this problem, the eigenvectors of the linearized Hamiltonian vector field at the equilibrium phase point $(x_1, x_2, \lambda_1, \lambda_2)_{eq} = (0, 0, 0, 0)$ are used as basis vectors. The initial layer problem is solved on the time interval $t \in [0, 15]$ (i.e., $t_{ibl} = 15$), while the terminal layer is solved in backward time on the interval $t \in [25, 40]$; these time intervals are much larger than the minimum rate of contraction and expansion which is approximately 2. After transforming the time variable to $\sigma = t_f - t$, the terminal layer problem has the same form as the initial layer problem, and the same algorithm is applicable. We use $v_u \equiv 0$ to initialize the algorithm. Figs. 1 and 2 show initial and terminal layer iterations 1, 2, and 10 from the forward integration for $x_1(t)$ versus t and $x_2(t)$ vs. t, respectively. The first important feature is that, even in the first iteration, the unstable behavior has been sufficiently suppressed. Moreover, with increasing iteration, both the initial and terminal boundary-layer solutions approach the equilibrium value in forward and backward time, respectively. After ten iterations, the initial and terminal layer solutions have converged in the sense that the trajectories from the forward and backward integrations differ by less than 10^{-4} and the initial and terminal layer solutions can be matched to the equilibrium solution to within 10^{-4}.

6. Dichotomic basis determination

While a detailed study of determining a dichotomic basis (or a sufficiently close approximation to one) is beyond the scope of this paper, some discussion is warranted to indicate feasibility. Consider the transformation

$$\begin{bmatrix} x \\ \lambda \end{bmatrix} = \Psi(z_s, z_u), \tag{29}$$

where Ψ is a diffeomorphism on Ω (Ω as defined earlier). The corresponding tangent vector coordinates are related by

$$\begin{bmatrix} \delta x \\ \delta \lambda \end{bmatrix} = \frac{\partial \Psi}{\partial z_s} \delta z_s + \frac{\partial \Psi}{\partial z_u} \delta z_u. \tag{30}$$

If z_s and z_u are stable and unstable phase coordinates, then $v_s = \delta z_s$ and $v_u = \delta z_u$ are a particular set of stable and unstable phase rate coordinates corresponding to the dichotomic basis composed of $D_s = \partial \Psi / \partial z_s$ and $D_u = \partial \Psi / \partial z_u$.

We are proposing to determine a dichotomic basis directly, without first determining a transformation to stable and unstable phase coordinates. Then, in principle, the corresponding transformation to stable and unstable phase coordinates can be determined by solving the appropriate partial differential equations. However, if the

basis vectors are determined in numerical form, this step is not straightforward. The most general means of determining a dichotomic basis may be to compute regional Lyapunov exponents and their associated direction vectors (see Wiesel, 1994). This is a computationally intensive means and may not always be necessary. Some alternatives are described briefly in the remainder of this section.

For the class of HBVPs we have addressed in this paper, the stable and unstable eigenspaces at the equilibrium point provide useful information for determining a dichotomic basis. At the equilibrium point, a dichotomic basis $D = [D_s \ D_u]$ must be such that the column span of D_s is the stable eigenspace and the column span of D_u is a complementary subspace. Thus the eigenvectors associated with the linearized system at the equilibrium point can be used to construct a dichotomic basis at the equilibrium point. In the nonlinear spring-mass example, the eigenvectors at the equilibrium point provided a sufficiently accurate approximate dichotomic basis along and in the neighborhood of the trajectory of interest. The adequacy of the equilibrium eigenvectors depends on (i) the distance of the initial (and final) phase point from equilibrium relative to the degree of curvature in the stable manifold and (ii) the length of the time interval. For the case when the equilibrium eigenvectors are inadequate, we now suggest a means of propagating a dichotomic basis to other phase points on the stable manifold.

Consider a basis of the form

$$D = \begin{bmatrix} I & 0 \\ R & I \end{bmatrix}. \tag{31}$$

Applying Eq. (12), the dynamics of the corresponding phase rate coordinates along a trajectory of the Hamiltonian vector field in Eq. (3) are

$$\begin{bmatrix} \dot{v}_s \\ \dot{v}_u \end{bmatrix} = \begin{bmatrix} H_{\lambda x} + H_{\lambda\lambda}R & H_{\lambda\lambda} \\ \Gamma(R) & -(H_{x\lambda} + RH_{\lambda\lambda}) \end{bmatrix} \begin{bmatrix} v_s \\ v_u \end{bmatrix}, \tag{32}$$

where $\Gamma(R) = -\dot{R} - RH_{\lambda x} - RH_{\lambda\lambda}R - H_{xx} - H_{x\lambda}R$. By choosing R to be a solution of the Riccati differential equation $\Gamma(R) = 0$, v_u is decoupled from v_s. Provided the correct boundary condition for the Riccati differential equation is used, it can be shown that

$$\text{span} \begin{bmatrix} I \\ R \end{bmatrix} \tag{33}$$

is the tangent space to the stable manifold at any point along an extremal trajectory lying on the stable manifold of a saddle-point equilibrium. It is known that the stable eigenspace is obtained from Eq. (33) using $R = R_\infty$ where R_∞ is the positive definite solution to the algebraic Riccati equation with matrices evaluated at the equilibrium point. Consequently, as $t_{ibl} \to \infty$, $R = R_\infty$ is the correct

boundary condition and should be quite accurate for sufficiently large t_{ibl}. Along an extremal on the stable manifold, a dichotomic basis can be constructed by integrating the Riccati differential equation in backward time from R_∞. While this connection to Riccati equation based solution methods (e.g. see the backward sweep method described by Bryson and Ho (1975) and the dichotomic transformations in Kokotovic et al. (1986)) is important, it should not obscure the more general perspective of our approach which leaves open the possibilities of generating bases by other means and exploiting the structure of two time-scale systems.

7. Conclusions

As a step toward developing a general method for solving and uncovering the structure of two time-scale optimally controlled nonlinear systems, we have defined a degenerate class of two time-scale optimal control problems, called completely hyper-sensitive problems, and have proposed a solution method for this class of problems. The solution method is indirect and as such seeks a solution to the Hamiltonian boundary-value problem posed by the first-order optimality conditions. The method uses a dichotomic basis to split the Hamiltonian vector field into its stable and unstable components. With these components identified, an accurate approximation to the optimal solution can be constructed by matching initial and terminal boundary-layer segments, on the stable and unstable manifolds, respectively, with the equilibrium segment. A variation of the method for the case of an approximate dichotomic basis was also developed, and was applied to a nonlinear spring-mass problem. The challenging problem of determining a dichotomic basis or a sufficiently accurate approximation to one has been discussed only briefly, but some potential approaches have been identified.

References

Anderson, B. D. O., & Kokotovic, P. V. (1987). Optimal control problems over large time intervals. *Automatica, 23*, 355–363.

Ardema, M. D. (1983). An introduction to singular perturbations in nonlinear control. In M. D. Ardema (Ed.), *Singular perturbations in systems and control*, pp. 1–92.

Ardema, M. D., & Rajan, N. (1985). Slow and fast state variables for three-dimensional flight dynamics. *J. Guidance, 8*, 532–535.

Ardema, M. D. (1990). Computational singular perturbation methods for optimal control. *Proc. of the American Control Conf.*, San Diego.

Arnol'd, V. I. (1992). *Ordinary differential equations*. Berlin: Springer-Verlag.

Ascher, U. M., Mattheij, R. M., & Russell, R. D. (1995). *Numerical solution of boundary value problems for ordinary differential equations*. Philadelphia: SIAM Press, .

Boothby, W. M. (1986). *An introduction to differentiable manifolds and Riemannian geometry*. San Diego: Academic Press.

Bryson, A. E., & Ho, Y. C. (1975). *Applied optimal control*. New York: Hemisphere Publishing.

Calise, A. J. (1976). Singular perturbation methods for variational problems in aircraft flight. *IEEE Trans. Automat. Control*, *21*, 345–353.

Chow, J. H. (1979). A class of singularly perturbed nonlinear, fixed endpoint control problems. *J. Opt. Theory Appl.*, *29*, 231–251.

Fenichel, N. (1979). Geometric singular perturbation theory for ordinary differential equations. *Journal of Differential Equations*, *31*, 53–98.

Guckenheimer, J., & Holmes, P. (1990). *Nonlinear oscillations, dynamical systems, and bifurcations of vector fields*. New York: Springer-Verlag.

Kelley, H. J. (1973). Aircraft maneuver optimization by reduced-order approximation. In Leondes, C. T. (Ed.), *Control and Dynamic Systems* 131–178.

Kokotovic, P. V., Khalil, H. K., & O'Reilly, J. (1986). *Singular perturbation methods in control*: *analysis and design*. New York: Academic Press.

Lam, S. H. (1993). Using CSP to understand complex chemical kinetics. *Comb. Science Tech.*, *89*, 375–404.

Lam, S. H., & Goussis, D. A. (1994). The CSP method of simplifying kinetics. *Int. J. Chem. Kin.*, *26*, 461–486.

Marino, R., & Kokotovic, P. V. (1988). A geometric approach to nonlinearly singularly perturbed control systems. *Automatica*, *24*, 31–41.

Mease, K. D. (1995). Geometry of computational singular perturbations. *Proc. IFAC Nonlinear Control Design Symposium*, Tahoe City.

Mease, K. D. (1995). An approach to solving two time-scale trajectory optimization problems. *IFAC Workshop on Control Applications of Optimization*, Haifa, Israel.

O'Malley, R., & Anderson, L. (1982). Time-scale decoupling and order reduction for linear time-varying systems. *Opt. Control Appl. Meth.*, *3*, 133–153.

Rao, A. V. (1996). *Extension of computational singular perturbation methodology to optimal control problems*. Ph.D. Thesis, Princeton University.

Rao, A. V., & Mease, K. D. (1998). Eigenvector approximate dichotomic basis method for solving hyper-sensitive optimal control problems. *Opt. Contr. Appl. Meth.* (to appear).

Sacker, R. J., & Sell, G. R. (1980). The spectrum of an invariant submanifold. *Journal of Differential Equations*, *38*, 135–160.

Wiesel, W. E. (1994). Modal feedback control on chaotic trajectories. *Phys. Rev. E*, *49*, 1990–1996.

Wilde, R. R., & Kokotovic, P. V. (1972). A dichotomy in linear control theory. *IEEE Trans. Automat. Control*, *17*, 382–383.

Anil V. Rao received his A.B. in mathematics and B.S. in mechanical engineering from Cornell University, his M.S.E. in aerospace engineering from The University of Michigan, and his M.A. and Ph.D. from Princeton University. Since completing his Ph.D., he has been employed by the Aerospace Corporation in El Segundo, California, where he is currently Senior Member of the Technical Staff. His research interests include guidance, navigation, and control of aircraft and spacecraft, nonlinear dynamical systems, and optimal control. He is a member of the American Institute of Aeronautics and Astronautics.

Kenneth D. Mease is Associate Professor of Mechanical and Aerospace Engineering at the University of California, Irvine. He received his B.S. from the University of Michigan and his M.S. and Ph.D. from the University of Southern California. He began his career at the Jet Propulsion Laboratory as a Member of the Technical Staff of the Navigation Systems Section. He then served as Assistant Professor of Mechanical and Aerospace Engineering at Princeton University, before joining the faculty of UC Irvine. His research has contributed to the guidance, navigation and control of aircraft and spacecraft, time-scale decomposition for nonlinear dynamical systems, and molecular control, and has been sponsored by NASA, NSF, McDonnell Douglas Corporation, and Rockwell International Corporation. He also serves as a consultant to several aerospace companies. He has served as Associate Editor for the AIAA Journal of Guidance, Control, and Dynamics and the AAS Journal of the Astronautical Sciences. He is an Associate Fellow of the AIAA, and a Member of SIAM and IEEE.