JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 32, No. 2, March–April 2009

Three-Dimensional Trajectory Optimization Satisfying Waypoint and No-Fly Zone Constraints

Timothy R. Jorris* and Richard G. Cobb†
Air Force Institute of Technology, Wright–Patterson Air Force Base, Ohio 45433

DOI: 10.2514/1.37030

To support the U.S. Air Force's global reach concept, a Common Aero Vehicle is being designed to support the global strike mission. *Waypoints* are specified for reconnaissance or multiple payload deployments and *no-fly zones* are specified for geopolitical restrictions or threat avoidance. Because of time critical targets and multiple scenario analysis, an autonomous solution is preferred over a time-intensive, manually iterative one. Thus, a real-time or near real-time autonomous trajectory optimization technique is presented to minimize the flight time, satisfy terminal and intermediate constraints, and remain within the specified vehicle heating and control limitations. This research uses the hypersonic cruise vehicle as a simplified two-dimensional platform to compute an optimal analytical solution. An up-and-coming numerical technique is a direct solution method involving discretization and then dualization, with pseudospectral methods and nonlinear programming used to converge to the optimal solution. This numerical technique is first compared to the previously derived 2-D hypersonic cruise vehicle analytical results to demonstrate convergence to the optimal solution. Then, the numerical approach is applied to the 3-D Common Aero Vehicle as the test platform for the flat Earth three-dimensional reentry trajectory optimization problem. The culmination of this research is the verification of the optimality of this proposed numerical technique, as shown for both the two-dimensional and three-dimensional models. Additionally, user implementation strategies are presented to improve accuracy, enhance solution convergence, and facilitate autonomous implementation.

Nomenclature

B	=	vehicle/mission specific constant
C	=	path constraint
C_D	=	coefficient of drag
C_L	=	coefficient of lift
C_L^*	=	coefficient of lift that produces the maximum lift-to-drag ratio
c	=	number of path constraints
c_ℓ	=	fraction of C_L^*
D	=	differential approximation matrix
E^*	=	maximum lift-to-drag ratio
f	=	state vector derivative function
g_0	=	initial gravity
H	=	Hamiltonian
$\mathcal{H}$	=	Hamiltonian (not augmented)
h	=	altitude
i	=	waypoint number
J	=	cost function
$\bar{J}$	=	adjoined cost function
J_a	=	NLP augmented cost function
j	=	no-fly zone number
K	=	heating rate normalization constant
L	=	integrand cost
$\mathcal{L}$	=	Lagrange interpolating polynomials, basis $0, \ldots, N$
L_p	=	number of phases to be linked
m	=	number of controls
N	=	interior-point constraint, number of nodes
n	=	number of states
P	=	number of phases
$\mathbf{P}$	=	linkage constraints
Q	=	heating rate path constraint
q	=	number of endpoint constraints
$\mathbf{q}$	=	static parameter vector
$\dot{q}_s$	=	nondimensional heating rate at stagnation point
$\dot{q}_{s\,\mathrm{dim}}$	=	dimensional heating rate at stagnation point
$\dot{q}_{s\,\max}$	=	dimensional maximum heat flux
$R_\oplus$	=	Earth radius
r_{nose}	=	vehicle nose radius
r_0	=	initial radius
S	=	state inequality constraint function
S_{ref}	=	aerodynamic reference area
$S^{(q)}$	=	qth time derivative
t	=	time
t_d	=	discrete time
t_f	=	final time, terminal time, or time on target
t_i	=	passage time of waypoint i
t_j	=	contact time of no-fly zone j
t_0	=	initial time
$\mathbf{U}(\tau)$	=	approximate control vector, basis $1, \ldots, N$
$\mathbf{u}(t)$	=	control vector
V	=	airspeed
w_k	=	Gauss quadrature weights
$\mathbf{X}(\tau)$	=	approximate state vector, basis $0, \ldots, N$
x	=	coordinate in x direction
$\mathbf{x}(t)$	=	state vector
$\dot{\mathbf{x}}(t)$	=	state vector derivative
$\mathbf{x}_f$	=	final state vector
y	=	coordinate in y direction
β	=	atmospheric constant
γ	=	flight path angle
θ	=	heading angle
$\mathbf{\Lambda}$	=	approximate costate vector, basis $1, \ldots, N+1$
λ	=	costate Lagrange multiplier
μ	=	path constraint Lagrange multiplier
ν	=	terminal constraint Lagrange multiplier
π_n	=	path constraint multiplier
σ	=	bank angle
τ	=	transformed time

Presented at the 2007 AAS/AIAA Astrodynamics Specialist Conference, Mackinac Island, Michigan, 19–23 August 2007; received 6 February 2008; revision received 24 June 2008; accepted for publication 7 July 2008.

*Adjunct Faculty, Department of Aeronautics and Astronautics, 2950 Hobson Way, Building 640; timothy.jorris@us.af.mil. Member AIAA.

†Assistant Professor, Department of Aeronautics and Astronautics, 2950 Hobson Way, Building 640; richard.cobb@afit.edu. Senior Member AIAA.

Φ	=	function of terminal cost and terminal constraint
ϕ	=	cost at a discrete time
ψ	=	terminal (endpoint) constraint

Subscripts

end	=	last in finite sequence
F	=	associated with point $N+1$
k	=	associated with points $1, \ldots, N$
$\mathbf{u}$	=	partial derivative with respect to $\mathbf{u}$
$\mathbf{x}$	=	partial derivative with respect to $\mathbf{x}$

Superscripts

p	=	phase number
(q)	=	lowest time derivative of the path constraint containing a control
$*$	=	optimal or basis $1, \ldots, N$
$\dagger$	=	basis $1, \ldots, N+1$
$\tilde{}$	=	associated with KKT conditions

I. Introduction

GLOBAL strike and global persistent attack are two of the seven United States Air Force concepts of operations [1]. Various hypersonic and reentry vehicle technologies are being pursued to enable such a prompt global reach capability [2]. The government competition for the Force Application Launch from the Continental United States (FALCON) [3] program and the NASA Next Generation Launch Technology Program Office [4] validate both the interest and the need for future research. The Common Aero Vehicle (CAV) is an unmanned aerial vehicle supporting these technologies with the role of "striking from space" [5].

The space shuttle [6] clearly demonstrates that successful techniques for reentry trajectory generation exist. Hanson and Jones [7] perform an extensive comparison among five predominant contenders [8–14]. Flight test results are found in [15,16]. These do an excellent job of rapidly computing a feasible reentry trajectory. Some require an existing nominal trajectory [14], while others implement bank reversal logic [12]. However, one of the research goals herein is to generate a reentry trajectory without a preexisting solution. Another fundamental goal of this research is to avoid geopolitically sensitive regions which is not addressed in previous research. Flying over specified waypoints is yet another research goal. Linear quadratic regulators [17], dynamic optimization [18], and fuzzy logic [19,20] have been shown to handle waypoints. Voronoi diagrams [21] and barrier [22] and interior point [23] methods have been shown to handle threats and restricted airspace. A solution technique was sought to handle all these criteria in a single problem. The Program to Optimize Simulated Trajectories (POST) [24] can suffer from divergent integration when implementing the shooting method. Therefore, a method outside these techniques was investigated. Preliminary studies of pseudospectral methods [25] lead to software implementation [26–29], which culminated in the solution process used in this research.

The overall mission objective is to fly from an initial point to a final/terminal point or target, in minimum time. The generic start and finish point of a trajectory are called endpoints. Waypoints are specified as intermediate coordinates to fly over to satisfy payload delivery or reconnaissance mission requirements. The vehicle must fly directly over each waypoint, also called waypoint passage; however, time, altitude, bank angle, flight path angle, velocity, and heading are not constrained. Some control law solutions are called bang–bang or bang–level–bang, meaning the control takes on the value of minimum, zero, or maximum. For a bang–bang controller for bank angle, a turnpoint is defined as a discrete point where there is a change in the current constant control; therefore, the vehicle rolls into a turn or completes a turn at a turnpoint. There is no predetermined limit on the number of discrete changes in control, that is, turnpoints. If waypoint passage occurs at some point within a constant control turn, the waypoint and turnpoints are not coincident. A no-fly zone is a region with a boundary that the vehicle may contact but must not violate. The entire trajectory can be broken up into sequential legs, segments, or phases; these terms may be used interchangeably. The breaks between phases may occur at waypoints, turnpoints, or other ending criteria such as no-fly zone contact. The problem starts at the initial time t_0. Because there can be multiple waypoints in a mission, each is numbered i, and the time of passage is denoted t_i. Similarly, each no-fly zone is numbered j; however, the time of interest for each no-fly zone is the time of no-fly zone boundary contact t_j. The target is reached at the final or terminal time t_f and is called target intercept.

II. Problem Definitions and Assumptions

This section documents the generic dynamic optimization problem as well as the specific constraints used throughout this research. Specific aerospace vehicles are used to exemplify the dynamic optimization solution techniques. The simpler 2-D (horizontal plane) trajectory solution [30,31] uses the powered hypersonic cruise vehicle (HCV) model and mission [32,33]. This is used to build up to the more complex 3-D trajectory using the CAV model and mission [31,34,35]. The constraints of waypoints and no-fly zones are added to further define an operational mission. With the problem and assumptions clearly defined, the analysis and results follow in Sec. III.

A. Generic Problem Statement

The overall objective of this research is to create a constrained optimal trajectory, for a given vehicle, to strike a final target in minimum time. The cost or objective function defines optimality, which is mission dependent and typically defined by the user. For example, the user may want to arrive in minimum time or maximum energy. The constraints refer to adhering to equations of motion, limitations on allowable control, and mission specific waypoints and/or no-fly zones. The constraints may be implemented as equality and/or inequality constraints. Other path restrictions, such as thermal limits or rate limits, may also exist. The vehicle itself is represented by the dynamics or equations of motion and any heating or control limitations. The mission is represented via the final target, waypoints, and/or no-fly zones. These generic definitions are formulated into equations and used to solve for the optimal control, leading to the optimal trajectory. Therefore, the first step is to introduce the dynamic optimization nomenclature used throughout this research.

1. Dynamic Optimization

The continuous-time, dynamic optimization problem with open final time is a particular problem of interest. The open final time allows time to be in the cost function, which is the intent for a minimum time problem. The advantage to the continuous-time formulation, over discrete time formulation, is it allows the direct incorporation of the differential equations of motion. The equations of motion in Eq. (1) are a function of the states, $\mathbf{x}(t) \in \mathbb{R}^n$, the controls $\mathbf{u}(t) \in \mathbb{R}^m$, and the independent variable $t \in \mathbb{R}$, where n is the number of states and m is the number of controls,

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

The objective or cost function J, in Bolza form [36], has two components as seen in Eq. (2). The first component $\phi(\mathbf{x}(t_d), t_d)$ is a cost defined at a discrete time t_d, or the sum of costs at discrete times, $\phi(\mathbf{x}(t_{d_1}), t_{d_1}) + \phi(\mathbf{x}(t_{d_2}), t_{d_2}) + \cdots$. For this derivation t_d is the final time t_f, and the final state vector $\mathbf{x}(t_f)$ is represented by $\mathbf{x}_f$; thus, the cost may be a function of the final states and/or the final time. The second component is the integrand cost L, taken from the initial time t_0 to final time t_f. The combined cost function is

$$J = \phi(\mathbf{x}_f, t_f) + \int_{t_0}^{t_f} L(\mathbf{x}(t), \mathbf{u}(t), t)\,\mathrm{d}t \tag{2}$$

Additionally, there may be a terminal constraint ψ. The terminal constraint may be used to specify an exact target, impact time, or other final state/time function. The equation form of this constraint is

$$\psi(\mathbf{x}_f, t_f) = 0 \tag{3}$$

Certain path equality constraints C may also exist within the trajectory [37,38]. For example, the trajectory may need to follow along the boundary of a no-fly zone or the control may be maintained at its maximum value. Notice here that the equality constraint C is a function of the states and the controls:

$$C(\mathbf{x}(t), \mathbf{u}(t), t) = 0 \tag{4}$$

Inequality constraints are also important to this research. The inequality may be a function of either the states or the controls [[37], pp. 108–118], for example, maximum allowable heating or control. The inequality constraints only appear in the solution during the times they are active, that is, when the constraint is at its limit which is formulated to be zero. The state inequality constraint is $S(\mathbf{x}(t), t) \leq 0$. When the control constraint is active, the control inequality constraint in Eq. (5) is a special case of Eq. (4), that is, $C(\mathbf{u}(t), t) = 0$. Equation (7) illustrates how active and inactive constraints are handled. The technique formulated herein requires the equality constraint to be a function of the control; therefore, the nomenclature for C has been maintained whereas the distinction is made for S because it is not a function of the control. The separate state and control inequality constraints are

$$S(\mathbf{x}(t), t) \leq 0 \qquad C(\mathbf{u}(t), t) \leq 0 \tag{5}$$

The goal of optimization is to determine a control to minimize the cost while satisfying the dynamics and constraints. To ensure the constraints are always satisfied they are adjoined to the cost function using Lagrange multipliers: ν, λ, and μ. Notice that the following adjoined cost function $\bar{J}$ maintains the same value as Eq. (2) if all the constraints are satisfied. To simplify notation, the dependence of each variable is omitted, for example, $\phi(\mathbf{x}_f, t_f)$ is simply written as ϕ. Thus the optimization problem is to determine the control $\mathbf{u}$ that minimizes the scalar cost $\bar{J}$:

$$\min_{\mathbf{u}} \bar{J} = \phi + \nu^T\psi + \int_{t_0}^{t_f} [L + \lambda^T(f - \dot{\mathbf{x}}) + \mu^T C]\,\mathrm{d}t \tag{6}$$

The constraint C need only enter the problem when it is active, that is, when $C = 0$. Conversely, it can be ignored when the constraint is satisfied or nonactive, that is, when $C < 0$. The components of the multiplier μ are defined to create this functionality:

$$\mu \begin{cases} >0, & C = 0, \\ =0, & C < 0 \end{cases} \tag{7}$$

The state inequality constraint S appears to be missing; however, a method to incorporate this constraint is shown later. The calculus of variation dictates that the variations of the adjoined cost, taken with respect to each variable, must be zero in order for the solution to possibly be a minimum. Thus, the variation of the adjoined cost function must be

$$\delta\bar{J} = 0 \tag{8}$$

Before presenting the results, some intermediate variables are defined. The Hamiltonian H is a scalar defined as

$$H \equiv L + \lambda^T f + \mu^T C \tag{9}$$

and Φ is the scalar defined as

$$\Phi \equiv \phi + \nu^T\psi \tag{10}$$

The time derivative of Φ is defined as

$$\dot{\Phi} \equiv \Phi_{\mathbf{x}}\dot{\mathbf{x}} + \Phi_{t_f} \tag{11}$$

The subscripts represent a partial derivative with respect to that variable. The following summarizes the necessary optimality conditions for the continuous-time, free final time problem; the intermediate steps can be found in [[36], p. 159]. The dependence notation is returned here to clarify the conditions that are functions of time versus those that apply only at the final time. The following shortened notation has been maintained; $\mathbf{x}$ is a simplification for $\mathbf{x}(t)$ and $\mathbf{u}$ is a simplification for $\mathbf{u}(t)$. The costates λ, differential equations are

$$\begin{aligned} \dot{\lambda}^T(t) &= -H_{\mathbf{x}}(\mathbf{x}, \mathbf{u}, t) \\ &= -L_{\mathbf{x}}(\mathbf{x}, \mathbf{u}, t) - \lambda^T(t) f_{\mathbf{x}}(\mathbf{x}, \mathbf{u}, t) - \mu^T(t) C_{\mathbf{x}}(\mathbf{x}, \mathbf{u}, t) \end{aligned} \tag{12}$$

The final conditions on the costates are

$$\lambda^T(t_f) = \Phi_x(\mathbf{x}_f, t_f) = \phi_x(\mathbf{x}_f, t_f) + \nu^T\psi_x(\mathbf{x}_f, t_f) \tag{13}$$

The optimality criterion:

$$\begin{aligned} &H_{\mathbf{u}}(\mathbf{x}, \mathbf{u}, t) = L_{\mathbf{u}}(\mathbf{x}, \mathbf{u}, t) + \lambda^T f_u(\mathbf{x}, \mathbf{u}, t) \\ &\quad + \mu^T(t) C_u(\mathbf{x}, \mathbf{u}, t) = 0 \quad \forall\ t \geq t_0 \end{aligned} \tag{14}$$

and finally the transversality condition:

$$\dot{\Phi}(\mathbf{x}_f, t_f) + L(\mathbf{x}_f, \mathbf{u}_f, t_f) = 0 \tag{15}$$

Notice that S does not explicitly appear in these optimality criteria; however, Sec. II.A.3 covers how S is to be included.

2. *Discontinuous Lagrange Multipliers*

A unique situation occurs when there are interior-point state constraints, that is, constraints that apply at a single unspecified time within the trajectory. A problem specific example is waypoints. This brings about the concept of discontinuous Lagrange multipliers. These discontinuities will also be called "jumps" in the costates [[37], p. 101]. Assume there is only one interior-point constraint, occurring at the unspecified time t_1, as a function of the states and time:

$$N(\mathbf{x}(t_1), t_1) = 0 \tag{16}$$

The time t_1 now represents the final time of the previous segment and the initial time of the next segment; thus, let t_1^- represent just before t_1 and t_1^+ just after t_1. Similar to the final condition on the costates in Eq. (13), there is now a jump to accommodate the intermediate condition in Eq. (16) [37,38]:

$$\lambda^T(t_1^-) = \lambda^T\left(t_1^+\right) + \pi_n^T \frac{\partial N}{\partial \mathbf{x}(t_1)} \tag{17a}$$

$$H(t_1^-) = H\left(t_1^+\right) - \pi_n^T \frac{\partial N}{\partial t_1} \tag{17b}$$

The vector of multipliers π_n must be solved to satisfy Eqs. (17a) and (17b). The finite number of interior-point constraints may be greater than 1; therefore, Eq. (17) can be generalized to apply to any one of the jumps simply by changing t_1 to the applicable time of the interior-point constraint, for example, t_i or t_j.

3. *Path Inequality and Equality Constraints*

The path inequality constraint is introduced in Eq. (5) as a function of the state and time; it is repeated here for convenience:

$$S(\mathbf{x}(t), t) \leq 0 \tag{18}$$

Because the control does not explicitly appear in Eq. (18), the time derivative of S is taken until the control does appear. If it takes q derivatives for the control to appear, this time derivative is written as $S^{(q)}$. For the special case where this is the only constraint, then $C = S^{(q)}$ and Eq. (9) becomes

$$H(\mathbf{x}, \mathbf{u}, t) = L(\mathbf{x}, \mathbf{u}, t) + \lambda^T f(\mathbf{x}, \mathbf{u}, t) + \mu^T S^{(q)}(\mathbf{x}, \mathbf{u}, t) \tag{19}$$

This process adds interior-point constraints for all of the $q-1$ derivatives which each must equal zero. Thus, the interior-point constraint vector in Sec. II.A.2 becomes

$$N(\mathbf{x}(t_1), t_1) = \begin{bmatrix} S(\mathbf{x}(t_1), t_1) \\ S^{(1)}(\mathbf{x}(t_1), t_1) \\ \vdots \\ S^{(q-1)}(\mathbf{x}(t_1), t_1) \end{bmatrix} \tag{20}$$

This interior-point constraint occurs at some time t_1, which for the path constraint, is the point of boundary contact. Boundary contact refers to the instant the edge, or boundary, of the path constraint S is encountered, that is, the first instant that $S = 0$. For this mission, $S = 0$ at the no-fly zone boundary, and $S^{(1)} = 0$ implies the velocity is tangent to the boundary; see Sec. II.C.7. There is a jump in the costates occurring when the constraint becomes active at time t_1. Since $S^{(q)}$ is a function of the control, it may be possible to maintain $S^{(q)} = 0$; this represents being on the constrained arc. The constrained arc is the time the solution follows the boundary of S, for example, $S = 0$ translates to flight along the no-fly zone boundary. It is considered constrained at $S = 0$ because any deviation inward, $S > 0$, would violate the constraint, and any deviation outward, $S < 0$, may increase the cost. Unlike entering (starting along) the constrained arc, there is no jump in the costates upon exiting (leaving) the constrained arc. Also, since $S^{(q)}$ is a function of the states, the partial derivative $S_x^{(q)}$ appears in the propagation of the costate equations, Eq. (12). The optimal control continues to be solved from Eq. (14). If the computed optimal control would lead to violating the constraint, then the control required to remain on the constraint is used instead. Therefore, the optimal trajectory may follow the constraint or simply touch the constraint, based on whichever minimizes the Hamiltonian without violating the constraint.

This section defined the continuous-time dynamic optimization problem with free final time. Equality and inequality constraints were addressed as a function of the controls or the states. The next step is to specify the particular mission requirements in terms of constraints to be incorporated into the constrained dynamic optimization problem for this research. The following sections address the operational mission and the CAV model.

4. *Cost Function*

The objective is to minimize flight time. The cost can therefore be expressed as t_f or as $\int_0^{t_f} 1\,\mathrm{d}t$. The former seems to produce slightly faster numerical results, possibly because the integral of 1 internally requires an additional state. The results should obviously be the same with either expression, and so for this research the cost function is

$$J = t_f \tag{21}$$

B. Mission Assumptions

The overall problem has mission components and vehicle components. The mission components include the target, waypoints, and no-fly zones; whereas the vehicle components include the equations of motion, control limitations, and aerodynamic heating limitations. The following mission assumptions are presented first to begin to scope the problem:

1) The waypoints are specified in the desired sequence.
2) Waypoint passage must be directly overhead.
3) Inner-loop control is available. Only the outer loop or trajectory generation is addressed.
4) Waypoints are sufficiently spaced such that no two are within the turn radius of the vehicle.
5) The altitude for waypoint passage is not specified.
6) No-fly zones are specified as circular exclusion zones with infinite altitude.
7) No-fly zones must not be violated.
8) Target coordinates and final altitude are specified.

Item 1 ensures the user's mission definition is not altered, for example, the vehicle must fly over tracking station 1 before tracking station 2. Item 2 avoids modifying the solution by manipulating the penalty of a near miss. Item 3 scopes this research effort to focus on the primary trajectory optimization objective; it is assumed that the inner-loop problem (tracking a reference trajectory) can be achieved but is not addressed herein. Item 4 characterizes the expected performance of a hypersonic vehicle. Item 5 attempts to avoid overspecifying the problem; furthermore, waypoint altitude does not enhance the mission objectives. Item 6 simplifies the derivations; however, other shapes could alteratively be incorporated. Item 7, similar to item 2, avoids modification of the solution by manipulating the penalty of a slight violation of a no-fly zone. Item 8 ensures the solution does not place the vehicle at such an excessive altitude that spike heating becomes a factor during a rapid final descent, or leave the vehicle with insufficient altitude to perform any required terminal maneuvers.

There are consequences associated with each of these assumptions which characterize the solution. Item 1 implies there is no attempt to minimize the cost by reordering the waypoints; thus, the research as stated is not commensurate with a traveling salesman problem. Item 2 can lead to infeasibilities because there is no leniency for a near miss. One way to deal with a near miss is to specify an acceptable radius, meaning the interior-point constraint would be a region rather than a point. Alternatively, a cost could be assigned to the waypoint miss distance, but this cost must be carefully balanced with the terminal cost such as time or energy. The inner-loop control assumed in item 3 may not be possible if the rate of control in the solution exceeds the vehicle's capability. If this were an area of concern the problem could be modified to manipulate control rate, for example, $\dot{\alpha}$ rather than just α. The control rate would then have a defined limit as well as the magnitude of the control. Item 4 means that cloverleaf maneuvers have not been considered. Two-dimensional solutions to these turn back trajectories are in [39,40]. The altitude in item 6 could be specified to provide an acceptable overflight altitude, or an acceptable risk from a threat. Allowing violation of the no-fly zone in item 7 could be well posed as a cost associated with proximity to a threat. Therefore, these assumptions are considered part of the problem definition rather than limitations to the solution process. The following sections cover the vehicle specific assumptions and individual vehicle descriptions.

C. Common Aero Vehicle

The CAV is an operationally representative vehicle for this 3-D trajectory research. It represents a hypersonic reentry vehicle with cross-range capability necessary to acquire waypoints and avoid no-fly zones. It has no thrust; therefore, the reentry profile must avoid exceeding a specified limit of the heating rate at the stagnation point. The following describes the vehicle model and expresses the CAV mission in terms of dynamic optimization constraints.

1. *CAV: Vehicle Description*

Phillips [35] describes a low-lift CAV and a high-lift CAV. The high-lift CAV, CAV-H, is modeled here to extend the cross-range capability. The full aerodynamic database and characteristic parameters are presented in Appendix A. The control and heating limitations are realistic, but are chosen to force boundary contact in order to verify different forms of the solution.

2. *CAV: Mission Profile*

A target nearly halfway around the world is used to represent a global strike mission. A waypoint in the middle of the Atlantic represents a telemetry station that can validate the vehicle navigation and control to terminate a rogue vehicle if necessary. The radius of the first no-fly zone is intentionally chosen to be much smaller than the turn capability of the CAV, thus forcing the solution to only be able to contact the boundary at a single point. The next waypoint

Table 1 CAV mission description

Descriptor	Latitude	Longitude	Radius
Initial	N 28° 35.286′	W 80° 40.194′	
Waypoint 1	N 34° 2.810′	W 27° 18.430′	
No-fly zone 1	N 20° 15.513′	W 3° 27.588′	960 nmi
Waypoint 2	N 33° 13.298′	E 41° 41.266′	
No-fly zone 2	N 55° 43.849′	E 58° 33.688′	1500 nmi
Target	N 31° 36.653′	E 65° 42.016′	
$h_0 = 122$ km, $V_0 = 24,000$ ft/s $= 7.3152$ km/s, $\gamma_0 = -1.5$ deg, $\theta_0 = 4$ deg			

represents a secondary mission target for reconnaissance or payload delivery. The last no-fly zone has a large enough radius to allow the boundary to be followed if optimal, and is positioned to force full control authority to clear the no-fly zone and still make it to the target. The heating constraint limitation is set low enough such that the unconstrained heating problem exceeds the value; thus the constrained solution must incorporate the limitation. The initial conditions are approximately those presented in [35]. The specified final altitude is the approximate altitude that the space shuttle uses to complete its entry guidance phase [6]. Table 1 provides the mission objectives.

Altitude is h. The flight path angle γ is the angle with respect to the local horizontal, measured positive away from the Earth. Variable V is airspeed, and θ is the heading angle, measured positive counterclockwise from the east.

3. *CAV: Assumptions*

This research follows a buildup in complexity. The intent is to start with a more intuitive analysis and advance to a more operationally representative model. The powered HCV allows for the simpler 2-D (horizontal plane) trajectory analysis as presented in [30,31]. These 2-D results are summarized herein and are used as a foundation for the more complex 3-D trajectory analysis. The 3-D analysis uses an unpowered hypersonic vehicle model, the CAV. A flat Earth assumption is employed to allow for consistent state definitions to facilitate comparison of results. Specifically, position can be consistently characterized using the horizontal components x and y, rather than making the transition to the spherical coordinates of latitude and longitude.

The following assumptions simplify the 3-D equations of motion and maintain the x, y, and θ states presented previously for the 2-D case [30,31]. These simplifications do not indicate limitations to this solution process; they are intended to make the derived analysis more intuitive and easier to follow. The CAV 3-D model assumptions are as follows: 1) flat, nonrotating Earth; 2) gravity is constant; 3) flight path angle is small; 4) drag is the dominant deceleration term; 5) coefficient of lift (C_L) and coefficient of drag (C_D) are only a function of angle of attack (AOA); 6) control is bank angle and angle of attack, both limited; 7) atmospheric density is modeled as a simple exponential; 8) heat flux at the stagnation point is limited; and 9) G loading and total heat are not addressed.

Ignoring the rotation of the Earth is a common assumption seen in [41–43]. The flat Earth approximation is most applicable to lifting bodies with a low flight path angle as seen in [44–46]. The flat Earth model is used as a proof of concept and to provide commonality with the 2-D derivation; however, a spherical Earth model would be more appropriate for computing operationally representative results. The small flight path angle leads to the small angle approximations. For hypersonic vehicles, C_L and C_D are often assumed independent of Mach number, that is, simply a function of angle of attack [47]. The estimated fit to the CAV aerodynamic data is in Appendix A.

4. *CAV: 3-D Equations of Motion*

The equations of motion represent a flat Earth model with zero Earth rotation. A small angle approximation for the flight path angle implies $\cos\gamma \approx 1$ and $\sin\gamma \approx \gamma$. It is also assumed that the drag term is the dominant term in the $\dot{V}$ equation, that is, the component of gravity in the velocity direction is negligible compared to the aerodynamic drag. This last assumption is consistent with the small angle approximation for γ. The nondimensional equations of motion originating from [31,44,45] are

$$\dot{x} = V\cos\theta \tag{22a}$$

$$\dot{y} = V\sin\theta \tag{22b}$$

$$\dot{h} = V\gamma \tag{22c}$$

$$\dot{V} = -\frac{BV^2 e^{-\beta r_0 h}(1 + c_\ell^2)}{2E^*} \tag{22d}$$

$$\dot{\gamma} = BVe^{-\beta r_0 h} c_\ell \cos\sigma - \frac{1}{V} + V \tag{22e}$$

$$\dot{\theta} = BVe^{-\beta r_0 h} c_\ell \sin\sigma \tag{22f}$$

The states x and y are position, h is altitude, V is airspeed, γ is flight path angle, and θ is heading. The constant $B = (\rho_0 r_0 S_{\text{ref}} C_L^*)/(2m)$ is a function of the vehicle parameters, atmospheric constant, and initial conditions. The constant E^* is the maximum lift-to-drag ratio. The controls are bank angle σ, and the fraction of C_L^*, c_ℓ. The coefficient C_L^* is the coefficient of lift that produces the maximum lift-to-drag ratio, as seen in Appendix A.

5. *CAV: Control Constraints*

The bank angle is limited to ± 60 deg to represent stability limitations, and to force the control to reach the imposed limit for this research. The maximum limitation of $c_\ell = 2$ is derived from the aerodynamic data in Appendix A. The minimum $c_\ell = 0$ maintains proper orientation for the thermal protection system. Thus the control limitations are

$$C(\mathbf{u}, t) = \begin{bmatrix} \sigma - \pi/3 \\ -\sigma - \pi/3 \\ c_\ell - 2 \\ -c_\ell \end{bmatrix} \leq 0 \tag{23}$$

6. *CAV: Terminal Constraints*

The terminal constraints for the CAV are the target coordinates and the specified final altitude:

$$\psi(\mathbf{x}(t_f), t_f) = \begin{bmatrix} x(t_f) - x_f \\ y(t_f) - y_f \\ h(t_f) - h_f \end{bmatrix} \tag{24}$$

7. *CAV: Path Inequality Constraints*

The path inequality constraints for the CAV are the no-fly zones and the maximum stagnation point heat rate. The no-fly zone constraints are

$$\begin{array}{c} S(\mathbf{x}(t), t) \\ j = 1, 2, \ldots, j_{\text{end}} \end{array} = \frac{1}{2}(R_j^2 - \Delta x_j^2 - \Delta y_j^2) \leq 0 \tag{25}$$

The radius of no-fly zone number j is R_j, centered at $[x_j, y_j]$. The distances from the center of no-fly zone j are $\Delta x_j = x - x_j$ and $\Delta y_j = y - y_j$. The function for the heating constraint is found in [41], which has its origin from [48], and is also discussed in [49]. The differences within these references are the selections of the second exponent value in Eq. (26). Some of the exponent choices are decimals, for example, 3.15. The integer exponent 3 was chosen to simplify the partial derivatives during analytical derivations. The dimensional form of the equation for $\dot{q}_{s\,\text{dim}}$ is

$$\dot{q}_{s\,\text{dim}} = \frac{k_{\text{dim}}}{\sqrt{r_{\text{nose}}}} \left(\frac{\rho}{\rho_{sl}}\right)^{1/2} \left(\frac{V_{\text{dim}}}{\sqrt{g_0 r_0}}\right)^3 \tag{26}$$

The dimensions are based on the converted units of the constant k_{dim} [48], which is $17,000$ Btu ft$^{-3/2}$ s^{-1}, and vehicle dependent radius of the nose, r_{nose}. The heating limitation is specified as a maximum allowable heat flux, $\dot{q}_{s\,\max}$. The dimensional $\dot{q}_{s\,\max}$ is used to convert the dimensional heating rate in Eq. (26) to the nondimensional form in Eq. (27), making the maximum allowable heat $\dot{q}_s = 1$. The nondimensional altitude is $h = (r_{\text{dim}} - r_0)/r_0$ and the nondimensional velocity is $V = V_{\text{dim}}/\sqrt{g_0 r_0}$ [47]. The nondimensional heating rate at stagnation point $\dot{q}_s$, in terms of the nondimensional states h and V, is

$$\dot{q}_s = K e^{-\beta r_0 h/2} V^3 \tag{27}$$

where

$$K = k_{\text{dim}} e^{-\beta(r_0 - R_\oplus)/2} / (\sqrt{r_{\text{nose}}} \dot{q}_{s\,\max})$$

The heating path inequality constraint will have the same form as S in Eq. (25); however, Q is used for distinction. There is no minimum heating limitation; however, from Eq. (27) $\dot{q}_s$ is always ≥ 0. Thus the heating inequality constraint is

$$Q(\mathbf{x}(t), t) = [\dot{q}_s - 1] = [K e^{-\beta r_0 h/2} V^3 - 1] \leq 0 \tag{28}$$

This Q is not related to total heating which is often represented by this symbol. For this research the total heating is considered a vehicle design parameter and not a trajectory optimization parameter.

8. *CAV: Interior-Point Constraints*

There are interior-point constraints associated with each waypoint i and each no-fly zone j. Each waypoint is specified at $[x_i, y_i]$. Part of the solution is the time of each waypoint passage t_i and the time of no-fly contact t_j.

The waypoint interior-point constraint is

$$\begin{array}{c} N(\mathbf{x}(t_i), t_i) \\ i = 1, 2, \ldots, i_{\text{end}} \end{array} = \begin{bmatrix} x(t_i) - x_i \\ y(t_i) - y_i \end{bmatrix} = 0 \tag{29}$$

Similarly, the derived no-fly zone interior-point constraint is

$$\begin{aligned} \begin{array}{c} M(\mathbf{x}(t_j), t_j) \\ j = 1, 2, \ldots j_{\text{end}} \end{array} &= \begin{bmatrix} S(\mathbf{x}(t_j), t_j) \\ S^{(1)}(\mathbf{x}(t_j), t_j) \end{bmatrix} \\ &= \begin{bmatrix} \frac{1}{2}(R_j^2 - \Delta x_j^2 - \Delta y_j^2) \\ -\Delta x_j V \cos\theta - \Delta y_j V \sin\theta \end{bmatrix} = 0 \end{aligned} \tag{30}$$

It can be seen that the heating constraint, Eq. (28), is not a function of either control. Therefore, as described in Sec. II.A.3, time derivatives must be taken until a control explicitly appears. This only requires the first time derivative; thus the interior-point constraint derived from the heating constraint Q up to the $q - 1$ derivative is simply

$$\begin{array}{c} Q(\mathbf{x}(t_k), t_k) \\ k = 1, 2, \ldots, k_{\text{end}} \end{array} = [\dot{q}_s(t_k) - 1] = [K e^{-\beta r_0 h(t_k)/2} V(t_k)^3 - 1] = 0 \tag{31}$$

This equation only requires satisfaction upon contact with the heating constraint at time t_k, where k represents the number of contacts with the heating constraint.

9. *CAV: Path Equality Constraints*

Path equality constraints arise when a constraint is active, that is, while the inequality constrained *equals* zero. As described in Sec. II.A.3, the control to remain on the constraint arc is used, that is, $S^{(q)} = 0$ will ensure S remains zero. This will be the control in use until the optimal control does not violate the constraint. For this case, the boundary of the no-fly zone is flown until it is no longer in the way. From Sec. II.A.3, $S^{(q)}$ is the lowest time derivative that contains the control and allows the boundary of the path constraint to be followed. Therefore, if on the no-fly zone constraint, $q = 2$, and so the second time derivative of the inequality constraint must equal zero:

$$\begin{aligned} C(\mathbf{x}, \mathbf{u}, t) &= S^{(q)}(\mathbf{x}, \mathbf{u}, t) = S^{(2)}(\mathbf{x}, \mathbf{u}, t) \\ &= -V^2 + (\Delta x_j \sin\theta - \Delta y_j \cos\theta)\dot{\theta} \\ &\quad + (-\Delta x_j \cos\theta - \Delta y_j \sin\theta)\dot{V} = 0 \end{aligned} \tag{32}$$

The term in front of $\dot{V}$ is simply $S^{(1)}/V$, thus it will be zero while Eq. (30) is satisfied; however, the term is retained to take the partial derivative with respect to $\mathbf{x}$ to propagate the costates. From Eq. (7), μ will be nonzero on the constraint, $C = 0$ and the $-\mu^T(t) C_{\mathbf{x}}(\mathbf{x}, \mathbf{u}, t)$ term will appear in Eq. (12) based on C defined in Eq. (32).

Similarly, while on the heating constraint, $q = 1$, and so the first time derivative of the heating inequality constraint must equal zero:

$$Q^{(1)}(\mathbf{x}, t) = \beta r_0 \gamma + \frac{3B}{E^*} e^{-\beta r_0 h}\left(1 + c_\ell^2\right) = 0 \tag{33}$$

Once again, contact with a constraint does not dictate remaining on the constraint. A boundary arc is only traversed if the optimal control would result in violation of the constraint. For example, in the 2-D case [30,31], the optimal control is bang–level–bang; however, a maximum bank turn would violate the no- fly zone. The optimal control is to follow the constrained arc, the boundary of the no-fly zone, until the bang–level–bang control can be resumed toward the next no-fly zone, waypoint, or target.

D. Numerical Methods Implementation Techniques

For the simpler 2-D HCV problem it is possible to derive the complete analytical solution; contrarily, the complexity of the 3-D CAV problem requires numerical techniques to efficiently compute the solution. The numerical technique proposed for the CAV solution is first applied to the HCV problem. This is performed to verify the numerical convergence and to analyze the accuracy of the numerical results as compared to the 2-D analytical solution. Problem setup using one phase or multiple phases, discontinuities in the control and costates, and discretization spacing are also investigated during the 2-D numerical implementation. A numerical technique typically represents a continuous problem via some form of discretization. The discretization method, that is, the spacing of the discretized points, is going to be based on the dynamics. Thus, the discretization may need to be different for different portions of the modeled dynamics. For example, Runge–Kutta integration is often run with a variable step size to allow for an increase in the number of time steps during rapid state changes. This methodology remains true when attempting to model the original states plus the added dimensionality of the costates. Therefore, an increased number of time steps is required for rapid changes in the states as well as rapid changes in the costates. Section II.A.2 discusses discontinuous Lagrange multipliers, that is, jumps in the costates. Thus, to properly model the system, any numerical technique must be able to model these discontinuities accurately. This modeling can be achieved by creating a dense set of time steps, or by actually allowing for a discontinuity. The dense set of time steps is used when the numerical method forces continuity, thus the dense set of continuous points replicates a discontinuity. Alternatively, a model can allow for discontinuities for certain variables while maintaining continuity for others. For example, the controls or costates may have permissible discontinuities whereas the states may be forced to be continuous. These differences in discretization, and discontinuity modeling, are characteristic of some of the software available.

One primary difference between software packages is the implementation of multiple segments or phases. Ross and D'Souza [50] and Fahroo and Ross's [51,52] DIDO uses a unique method to improve the numerical accuracy of its seemingly single phase solution. DIDO increases the density of the time steps, called nodes, by specifying intermediate "knots" at points where increased numerical accuracy is desired. Ross et al. [27] demonstrate a multistage optimization through *pseudospectral knotting*. Conversely, NASA's optimal trajectories by implicit simulation (OTIS) [53–55], Betts's sparse optimal control software (SOCS) [56], von Stryk's direct collocation method for the numerical solution of optimal control problems (DIRCOL) [57], Astos Solutions' graphical environment for simulation and optimization (GESOP) [58], Rutquist and Edvall's PROPT [59], and Rao's Gauss pseudospectral optimal control software (GPOCS), based on Huntington's dissertation [60], allow the user to break the problem up into multiple phases. The multiple-phased approach allows for a multipoint boundary value problem by specifying terminal constraints at the end of each phase, that is, interior-point constraints. Additionally, each phase is discretized individually; therefore, the nonuniformly spaced collocation points [25,61–65], for example, Chebyshev or Gauss points, create the desired situation of higher node density at the beginning and end of each phase. The choice of phase breakpoints dictates where desired higher accuracy occurs or where discontinues are allowed. This nonuniform node spacing can have the adverse effect of leaving sparse spacing in the middle of the phase, even with an increase in the specified number of nodes. If this middle region requires higher accuracy, it may be necessary to force another phase change in this region. In summary, a numerical solution may be broken up into phases for three purposes: to specify interior-point constraints, to allow for discontinuities, and/or to increase the accuracy at a region of high state, control, or costate change.

One of the challenges of a phased trajectory solution or approach is determining the required number of phases. To maximize accuracy, every jump in the costates should occur at a phase breakpoint to allow for such discontinuities. The number of jumps is dictated by the number of interior-point constraints. For this problem, the number of interior-point constraints is the number of waypoints plus the number of times the path constraints are contacted. Unfortunately, this number is typically part of the solution and is not known a priori. It is possible to run a lower number of phases, and then add phase breaks at each path constraint contact. However, it may also be possible to allow the numerical solution to model these discontinuities with continuous costates, that is, without a phase break. For simplified end user implementation, it is therefore desirable to minimize the number of phases, that is, eliminate the need to add more breakpoints. The ability to still converge on the optimal solution even with a minimal number of phases is therefore one of the topics of this research. Section III.E provides a comparison between a trajectory broken up into a higher number of phases, that is, increased accuracy, and a trajectory broken up into a minimal number of phases, that is, more desirable user implementation. Appendix B contains an overview of the pseudospectral method, nonlinear programming implementation, and multiple-phase connectivity which is the foundation of the solution technique used herein.

III. Analysis and Results

The culmination of this research effort lies in the analysis and results presented in this section. Section II developed the fundamental theory; however, the problem specific solutions remain to be solved. For each vehicle and mission, the goal is to *analytically* compute the optimal solution to the maximum extent possible. For the 2-D constant speed HCV a solution is determined using an indirect method by satisfying an analytically derived set of optimality criteria [30,31]. The solution to the HCV problem is recomputed using this new numerical solution technique and the results are compared to those previously computed via the analytical approach. The proposed direct numerical solution technique is used to solve the optimal 3-D CAV reentry trajectory problem. A two-step process is implemented to validate these numerical results. First, the candidate numerical result is tested to verify it matches the derived three-dimensional analytical form of the control. Once this rigorous process of verifying the optimality of the results is complete, a minimal set of user representative steps is outlined for future implementation.

A. 2-D HCV Trajectory: Numerical Solution

This section is an introduction to the numerical technique required to solve the more complex 3-D CAV problem. These numerical results are computed using the software package GPOCS,[‡] and are compared to previously computed analytical results [30,31]. Figure 1 shows that the numerical results are approximately equal to the analytically derived results. It also illustrates the numerical technique of separate phases, where the waypoints and target are specified as terminal boundary conditions for the appropriate phase.

The costates in Fig. 2 match for most of the time history. However, the numerical results have different jump values and derivatives while the trajectory is on the no-fly zone constraint. In comparison to Bryson and Ho's method [37], the GPOCS software package uses an alternate path constraint method that is compatible with the Gauss pseudospectral costate mapping theorem [64,66]. To understand the differences, this alternate method is used to derive another set of analytical equations for the costates; the derivation and comparison are presented next in Sec. III.B.

B. 2-D HCV Trajectory: Alternate Analytical Solution

The Gauss pseudospectral costate mapping theorem is used within GPOCS to convert the discrete Karush–Kuhn–Tucker (KKT) multipliers into the continuous costates. The user supplied path function provided to GPOCS is the no-fly zone constraint S; however, the original analytical result uses $S^{(2)}$ to adjoin to the Hamiltonian. Adjoining S directly, as in Eq. (34), is the D-form of the Lagrangian, and adjoining $S^{(2)}$, as in Sec. II.A.3, is the P-form named after Pontryagin [67]. For the Breakwell problem [37], this distinction is mentioned via footnote in [52], and analytically and numerically solved using both methods in [31],

$$H = L + \lambda^T f + \mu S \tag{34}$$

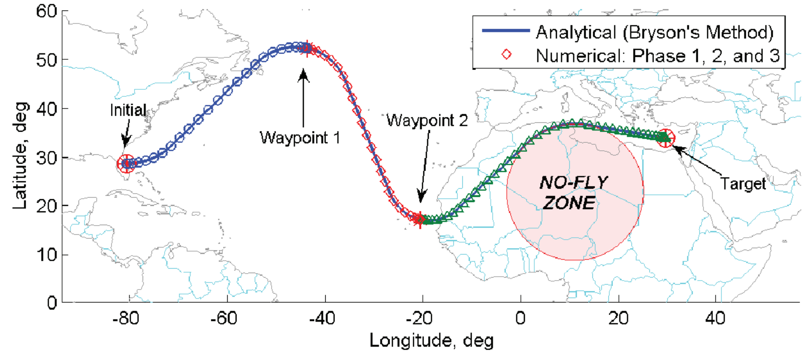
This alternate form of the Hamiltonian implies a change in the costate propagation equations from previous results [30,31]:

$$\dot{\lambda} = [-\lambda^T f_{\mathbf{x}}]^T - \mu[S_{\mathbf{x}}]^T$$

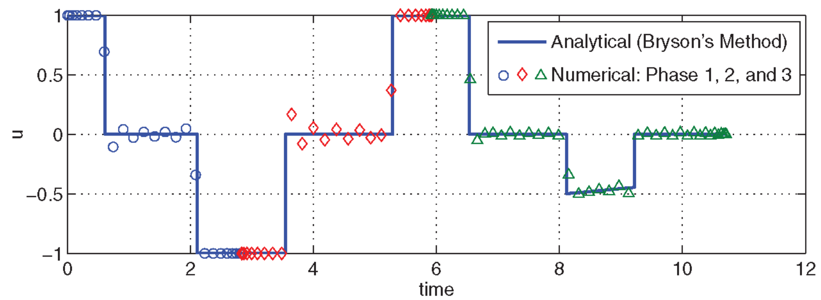
$$\begin{bmatrix} \dot{\lambda}_x \\ \dot{\lambda}_y \\ \dot{\lambda}_\theta \\ \dot{\lambda}_V \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \lambda_x V \sin\theta - \lambda_y V \cos\theta \\ -\lambda_x \cos\theta - \lambda_y \sin\theta + \lambda_\theta \frac{\tan\sigma_{\max}}{V^2} u \end{bmatrix} - \mu \begin{bmatrix} -\Delta x_j \\ -\Delta y_j \\ 0 \\ 0 \end{bmatrix} \tag{35}$$

This change in adjoined path constraint has created nonzero derivatives for λ_x and λ_y. Furthermore, the previously derived interior-point constraint M is not imposed and so there is no jump in the costates. From the new definition of $\dot{\lambda}_\theta$ in Eq. (35) there is zero change in its derivative occurring at no-fly zone contact; therefore, without a change in derivative it must remain its previous value of $\dot{\lambda}_\theta = 0$. This criteria leads to the following relationship along the

[‡]Results were produced using GPOCS Version 1.1.



a)



b)

Fig. 1 a) Trajectory and b) control: analytical (Bryson and Ho's method [37]) versus numerical results.

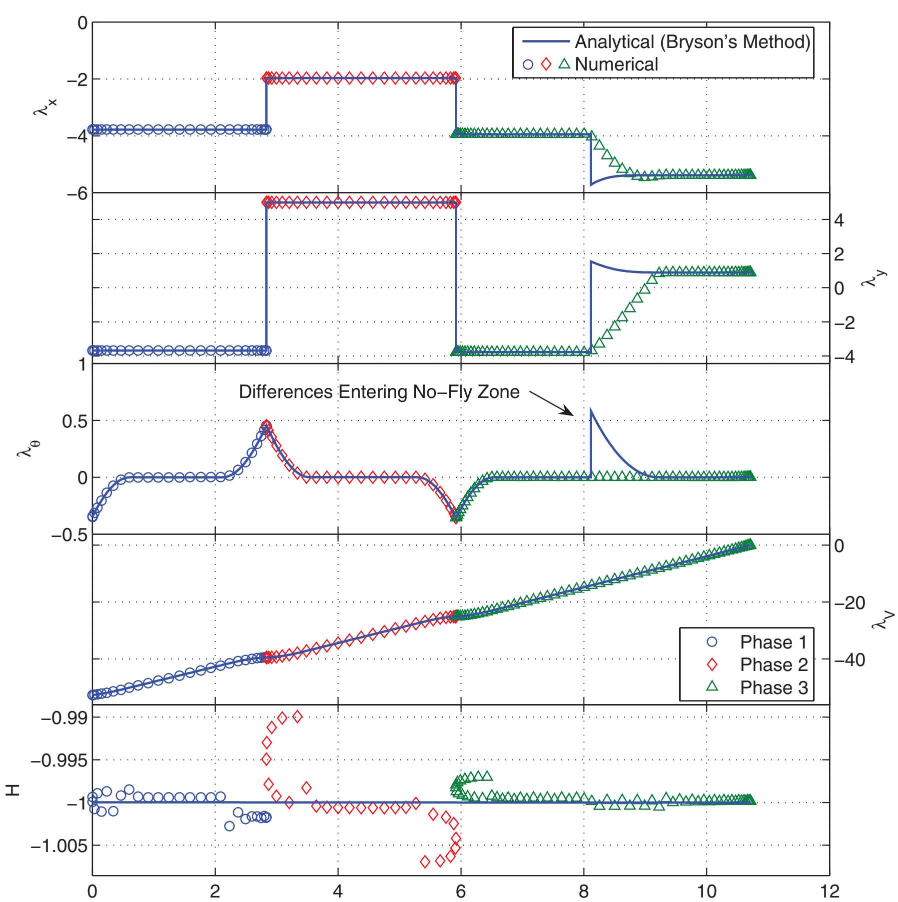


Fig. 2 Costates and Hamiltonian: analytical (Bryson and Ho's method [37]) versus numerical results.

no-fly zone boundary:

$$\lambda_x V \sin\theta - \lambda_y V \cos\theta = 0 \tag{36}$$

Since V is always positive it can be divided out of this equation. The next step is to take the time derivative of both sides of the equation:

$$\sin\theta\dot{\lambda}_x + \lambda_x \cos\theta\dot{\theta} - \cos\theta\dot{\lambda}_y + \lambda_y \sin\theta\dot{\theta} = 0 \tag{37}$$

Substituting in the values of $\dot{\lambda}_x = \mu\Delta x_j$ and $\dot{\lambda}_y = \mu\Delta y_j$, from Eq. (35), leads to an expression for μ:

$$\mu = \frac{-(\lambda_x \cos\theta + \lambda_y \sin\theta)V}{(\Delta x_j \sin\theta - \Delta y_j \cos\theta)^2} \tag{38}$$

Upon no-fly zone contact the costates are propagated using Eqs. (35) and (38) to create the time history presented in Fig. 3. The control time history is unchanged since $u^*(t)$ is

$$S = \tfrac{1}{2}\left(R_j^2 - \Delta x_j^2 - \Delta y_j^2\right) \tag{39}$$

$$S^{(1)} = -\Delta x_j V \cos\theta - \Delta y_j V \sin\theta \tag{40}$$

$$S^{(2)} = -V^2 + u(\Delta x_j \sin\theta - \Delta y_j \cos\theta)\tan\sigma_{\max} + (-\Delta x_j \cos\theta - \Delta y_j \sin\theta)a = 0 \tag{41}$$

$$u^*(t) = \frac{V^2}{(\Delta x_j \sin\theta - \Delta y_j \cos\theta)\tan\sigma_{\max}} \tag{42}$$

The u^* in Eq. (42) is only applicable while on a no-fly zone constraint.

The correlation in the numerical results, as shown in Fig. 3, with the alternate derivation of the costate propagation equations clearly shows that numerical computation is conforming to the optimality criteria. This understanding of what path constraint is being adjoined to the Hamiltonian to create the time history of the costates is used in the CAV 3-D analysis.

C. 3-D CAV Trajectory: Analytical Analysis

The analytical solution is derived first to determine the form of the optimal control. The first step is to derive the solution with no path constraints. Then, with path constraints, determine how the propagation of the costates changes with the inclusion of the path multipliers, μ^T.

1. 3-D CAV: Optimal Unconstrained Control

First assume the Hamiltonian is only a function of the integrand of the cost and the costates, that is, $\mu = 0$ because the path constraints are assumed inactive. From the cost definition in Eq. (21), the integrand is $L = 0$ so the Hamiltonian becomes

$$H = L + \lambda^T f = \lambda_x V \cos + \lambda_y V \sin\theta + \lambda_h V_\gamma - \lambda_V \frac{BV^2 e^{-\beta r_0 h}(1 + c_\ell^2)}{2E^*} + \lambda_\gamma \left[BVe^{-\beta r_0 h} c_\ell \cos\sigma - \frac{1}{V} + V\right] + \lambda_\theta BVe^{-\beta r_0 h} c_\ell \sin\sigma \tag{43}$$

The first-order optimality condition requires $H_{\mathbf{u}} = 0$ and the second-order criteria requires $H_{\mathbf{uu}} > 0$ [36,68], that is, positive definite. These represent the following matrices:

$$\mathbf{u} = [\sigma \quad c_\ell]^T \tag{44}$$

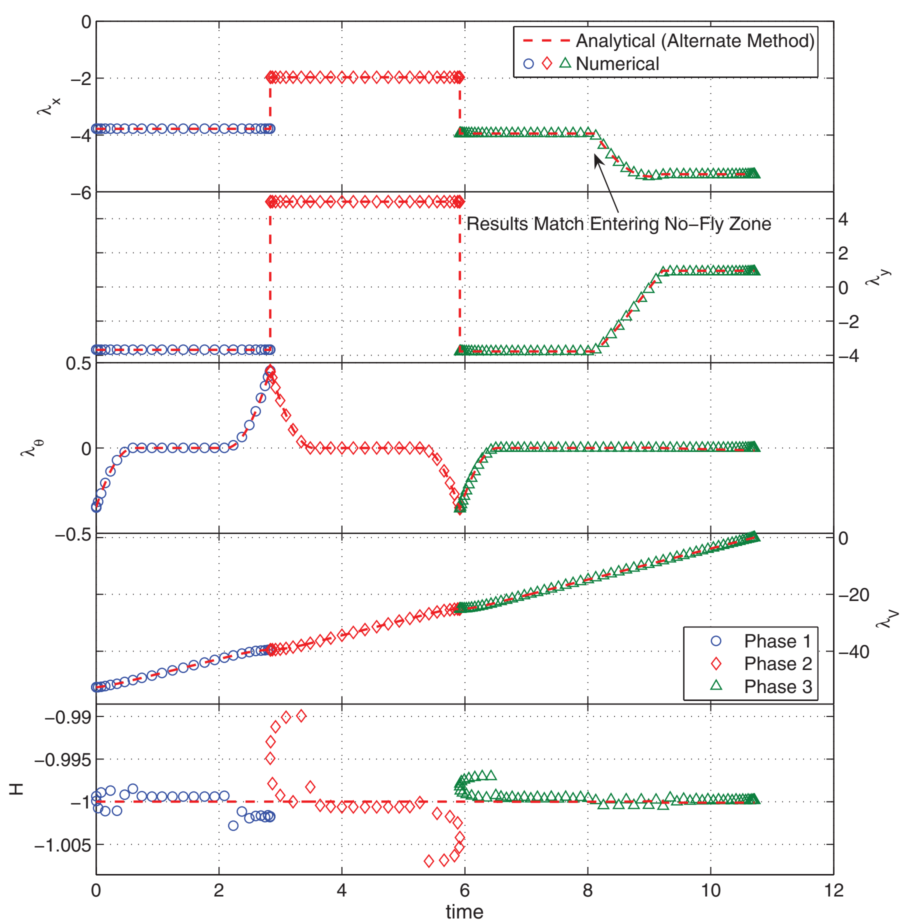


Fig. 3 Costates and Hamiltonian: analytical (alternate method) versus numerical results.

$$H_{\mathbf{u}} = [\,H_\sigma \quad H_{c_\ell}\,] = 0 \tag{45}$$

$$H_{\mathbf{uu}} = \begin{bmatrix} H_{\sigma\sigma} & H_{\sigma c_\ell} \\ H_{c_\ell \sigma} & H_{c_\ell c_\ell} \end{bmatrix} > 0 \tag{46}$$

Starting with Eq. (43) the partial derivatives of the Hamiltonian with respect to the controls are

$$H_\sigma = -\lambda_\gamma BVe^{-\beta r_0 h} c_\ell \sin\sigma + \lambda_\theta BVe^{-\beta r_0 h} c_\ell \cos\sigma = 0 \tag{47a}$$

$$H_{c_\ell} = -\lambda_V \frac{BV^2 e^{-\beta r_0 h}}{E*} c_\ell + \lambda_\gamma BVe^{-\beta r_0 h}\cos\sigma + \lambda_\theta BVe^{-\beta r_0 h}\sin\sigma = 0 \tag{47b}$$

The second partial derivative is taken to ensure $H_{\mathbf{uu}}$ is positive definite:

$$H_{\mathbf{uu}} = \begin{bmatrix} -c_\ell BVe^{-\beta r_0 h}(\lambda_\gamma \cos\sigma + \lambda_\theta \sin\sigma) & BVe^{-\beta r_0 h}(-\lambda_\gamma \sin\sigma + \lambda_\theta \cos\sigma) \\ BVe^{-\beta r_0 h}(-\lambda_\gamma \sin\sigma + \lambda_\theta \cos\sigma) & -\lambda_V \frac{BV^2 e^{-\beta r_0 h}}{E*} \end{bmatrix} > 0 \tag{48}$$

The off diagonals of $H_{\mathbf{uu}}$ are zero because they are simply H_σ/c_ℓ, where $H_\sigma = 0$ from Eq. (47a). Thus, the eigenvalues of $H_{\mathbf{uu}}$, which must be >0 [[36], p. 66], are equal to the terms on the main diagonal:

$$-c_\ell BVe^{-\beta r_0 h}(\lambda_\gamma \cos\sigma + \lambda_\theta \sin\sigma) > 0 \tag{49a}$$

$$-\lambda_V \frac{BV^2 e^{-\beta r_0 h}}{E*} > 0 \tag{49b}$$

In analyzing Eq. (49b), the terms B, V^2, $e^{-\beta r_0 h}$, and E^* are all positive, therefore λ_V must be negative:

$$\lambda_V < 0 \tag{50}$$

To satisfy the conditions in Eq. (47), two equations must be solved simultaneously:

$$-\lambda_\gamma \sin\sigma + \lambda_\theta \cos\sigma = 0 \tag{51a}$$

$$-V\lambda_V c_\ell / E^* + \lambda_\gamma \cos\sigma + \lambda_\theta \sin\sigma = 0 \tag{51b}$$

To solve Eq. (51a), assume $\lambda_\gamma = k\cos\sigma$ and $\lambda_\theta = k\sin\sigma$ where k is a constant to be determined. From the assumed values for λ_γ and λ_θ, and the trigonometric identity $\sin^2\sigma + \cos^2\sigma = 1$ [[69], p. A52], the following must also be true:

$$\lambda_\gamma^2 + \lambda_\theta^2 = k^2 \tag{52}$$

Thus, an intermediate solution is

$$k = \pm\sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \tag{53a}$$

$$\lambda_\gamma = \pm\cos\sigma\sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \quad \text{implies} \quad \cos\sigma = \pm\lambda_\gamma \big/ \sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \tag{53b}$$

$$\lambda_\theta = \pm\sin\sigma\sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \quad \text{implies} \quad \sin\sigma = \pm\lambda_\theta \big/ \sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \tag{53c}$$

The sign of k must still be determined from Eq. (49a); enforcing the control limitation $c_\ell \geq 0$, implies $\lambda_\gamma \cos\sigma + \lambda_\theta \sin\sigma < 0$. Using the solution from Eq. (53), or the assumed definitions of λ_γ and λ_θ, leads to

$$\left(\pm\lambda_\gamma^2 \pm \lambda_\theta^2\right) \big/ \sqrt{\lambda_\gamma^2 + \lambda_\theta^2} < 0 \tag{54a}$$

$$(k\cos\sigma)\cos\sigma + (k\sin\sigma)\sin\sigma < 0 \tag{54b}$$

To satisfy Eq. (54), k must be negative; therefore, the solution for the control σ is

$$\cos\sigma = -\lambda_\gamma \big/ \sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \tag{55a}$$

$$\sin\sigma = -\lambda_\theta \big/ \sqrt{\lambda_\gamma^2 + \lambda_\theta^2} \tag{55b}$$

$$\sigma = \mathrm{atan2}(\sin\sigma, \cos\sigma) \tag{55c}$$

Equation (55) is now plugged into Eq. (51b) to solve for c_ℓ as

$$c_\ell = \frac{-E^*\sqrt{\lambda_\gamma^2 + \lambda_\theta^2}}{V\lambda_V} \tag{56}$$

The atan2 function in Eq. (55c) solves for the correct quadrant for the angle σ. This may seem unnecessary since σ is confined to $-60 \leq \sigma \leq 60$ deg; however, the sign of σ is still determined by the unbounded solution. Thus, using this methodology maintains the correct sign for σ and avoids incorrectly jumping from one bound to the next when the unbounded solution exceeds ± 90 deg.

The numerically solved values, from the next section, for λ_V, λ_γ, and λ_θ will be input into Eqs. (55) and (56) to integrate the states forward to verify the numerical results are adhering to the optimality criteria enforced in this derivation.

2. 3-D CAV: Optimal Constrained Control

The CAV model allows for two controls, σ and c_ℓ, which have specified limitations. It will be shown that when one control is restricted to its limit, the other control law must be modified to maintain optimality. In other words, if a control is at its limit, the control laws derived in Eqs. (55) and (56) may no longer be valid.

The control limitations, Eq. (23), are adjoined to the Hamiltonian via the multiplier μ, as defined in Eq. (9), and repeated here:

$$H = L + \lambda^T f + \mu^T C \tag{57}$$

As an example, assume the bank angle has reached its maximum positive value, and thus the bank angle constraint is active. For optimality, H_σ and H_{c_ℓ} must still equal zero; however, since μ is now nonzero the solutions in Eqs. (55) and (56) are no longer valid. The new equations for H_σ and H_{c_ℓ} are

$$H_\sigma = -BVe^{-\beta r_0 h}\lambda_\gamma c_\ell \sin\sigma + BVe^{-\beta r_0 h}\lambda_\theta c_\ell \cos\sigma + \mu = 0 \tag{58a}$$

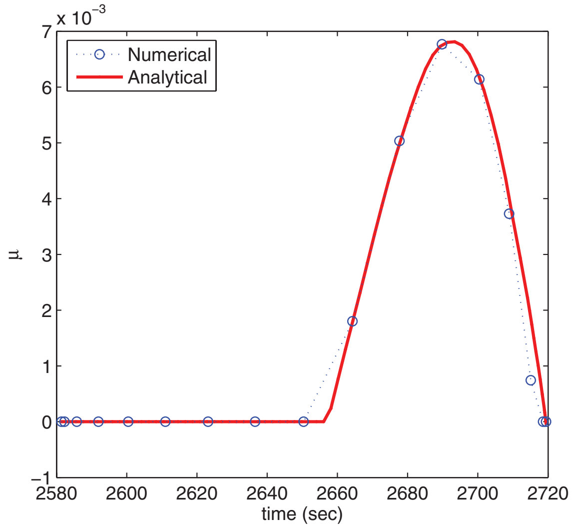


Fig. 4 Control multiplier μ for constrained bank angle ($\sigma = 60$ deg).

$$H_{c_\ell} = -\lambda_V \frac{BV^2 e^{-\beta r_0 h} c_\ell}{E^*} + \lambda_\gamma BV e^{-\beta r_0 h} \cos\sigma + \lambda_\theta BV e^{-\beta r_0 h} \sin\sigma = 0 \tag{58b}$$

Unlike the previous section, σ is now known to be at its maximum. Thus, when the bank angle is at its maximum positive value the optimal c_ℓ is

$$c_\ell = \frac{(\lambda_\gamma \cos\sigma + \lambda_\theta \sin\sigma)E^*}{V\lambda_V} \tag{59}$$

To compare this derivation to the numerical output, Eq. (59) is used in Eq. (58a) to define μ:

$$\mu = BV e^{-\beta r_0 h}(\lambda_\gamma \sin\sigma - \lambda_\theta \cos\sigma) c_\ell \tag{60}$$

The comparison between the numerically generated results and this analytical form is presented in Fig. 4. The time span in Fig. 4 is at the end of the trajectory where the maximum bank angle is required to achieve the optimal solution. The matching results demonstrate that the numerical results are satisfying the optimality criteria. For the case of maximum c_ℓ, the optimality criteria in Eqs. (55) and (56) remain valid because Eq. (47a) is not influenced by c_ℓ (c_ℓ is present but it is divided out, thus leaving the originally derived solution).

D. 3-D CAV: Comparison of Analytical Versus Numerical Results

The 3-D CAV mission is described in Sec. II.C.2. It is anticipated that there will be a large spike in heating during the initial descent into denser atmosphere. From Table 1, the trajectory should first encounter waypoint 1, followed by contact with no-fly zone 1, then onto waypoint 2, contact with no-fly zone 2, and lastly hit the target. Any interior-point constraint may result in a discontinuity in one or more Lagrange multipliers. Furthermore, each path constraint translates to an interior-point constraint upon boundary contact, see Sec. II.A.3. By the evolution described below, the problem is eventually broken up into seven phases to account for every discontinuity. The first ends when touching the heating constraint, the second ends at waypoint 1, the third ends upon touching no-fly zone 1, the fourth ends at waypoint 2, the fifth ends upon touching no-fly zone 2, the sixth ends at the second contact of no-fly 2, and the seventh ends at the target. The test of results is to determine how well the numerical solution matches the analytically optimal control derived in Sec. III.C.

To maximize the accuracy of the numerical results the trajectory is broken up such that every jump in the costates occurs at the end of a phase. This requires an iterative approach to produce these more accurate results. However, the more accurate results are used to verify optimality and to test the solution convergence with fewer phases, that is, less iterations. The initial guess is a straight line trajectory to the target, that is, "point to target."

The costate jumps occur at contact with the heating and no-fly zone constraints. Preliminary numerical results, with fewer phases, are used to identify the number of path constraint contacts. Next, additional phases are specified to more accurately capture these events and verify the original results. Thus, this trajectory is broken into seven phases: t_0 to heating constraint contact, heating constraint contact to waypoint 1, waypoint 1 to no-fly zone 1, no-fly zone 1 to waypoint 2, waypoint 2 to first contact of no-fly zone 2, first contact of no-fly zone 2 to second contact of no-fly zone 2, and second contact of no-fly zone 2 to target, as seen on the map in Fig. 5.

In Figs. 5–9 the numerical results are compared to the form of the solution analytically derived in Secs. III.C.1 and III.C.2. The states and constraints are nearly a perfect match. For the controls in Fig. 7 the analytical results appear smoother, especially at the beginning of the trajectory. This is attributed to costate interpolation used to compute the controls. The derivations in Sec. III.C.2 discussed the change in the optimal control law when a control constraint is reached. The correlation in the results at the end of the trajectory in Fig. 6 demonstrates the successful implementation of the revised control law. The data point jump at the last point in Fig. 7 is caused from the singularities that arise as the costates λ_V, λ_γ, and λ_θ approach the analytically derived values of zero. Figure 8 shows the path constraints; the constraint is inactive while negative, active when equal to zero, and violated if it becomes positive. Therefore, boundary contact first occurs with the heating constraint at about 250 s, followed by no-fly zone 1 at about 1235 s, and then onto no-fly zone 2 at about 2460 s. The scalings on the no-fly zone plots are misleading because they have been converted back to dimensional units; however, the same S plots in Fig. 9 in nondimensional units have an order of magnitude of 10^{-4}. Lasty, in the expanded view in Fig. 9 there are differences in constraint contact points. Herein lies the instability with the integration in the shooting method: the points of contact and departure must be exact or the propagated costates may diverge.

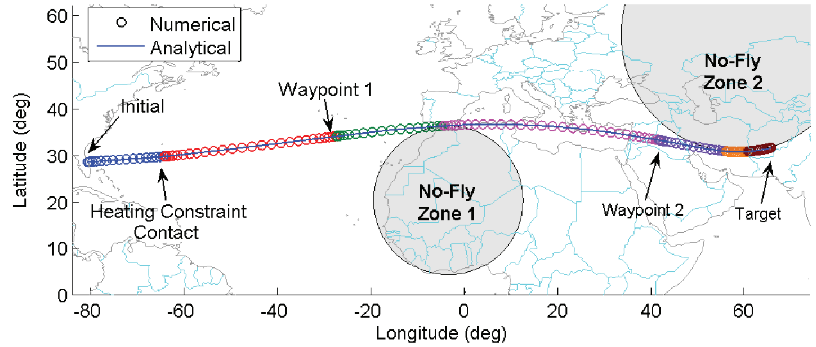


Fig. 5 Trajectory: seven phase numerical and analytical comparison.

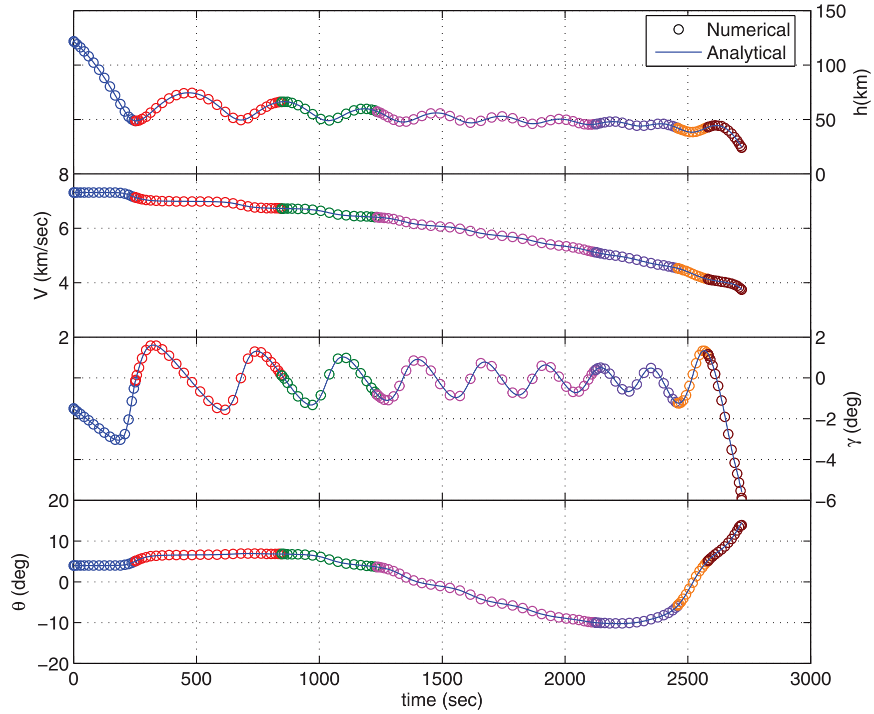


Fig. 6 States: seven phase numerical and analytical comparison.

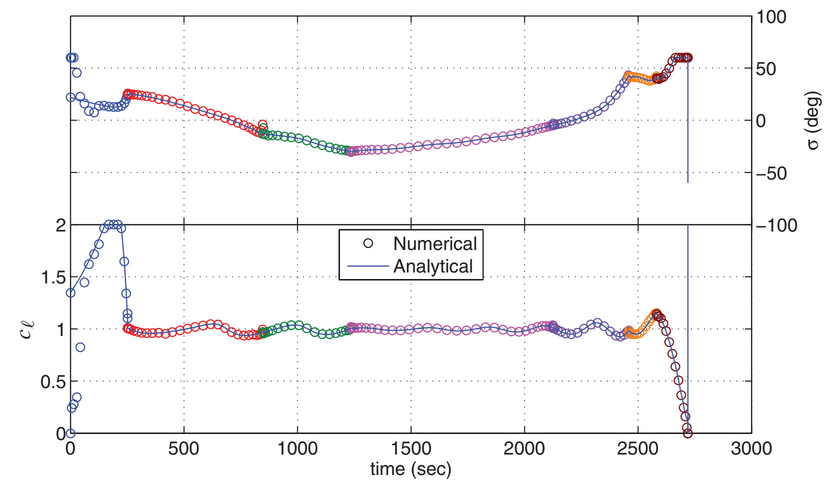


Fig. 7 Controls: seven phase numerical and analytical comparison.

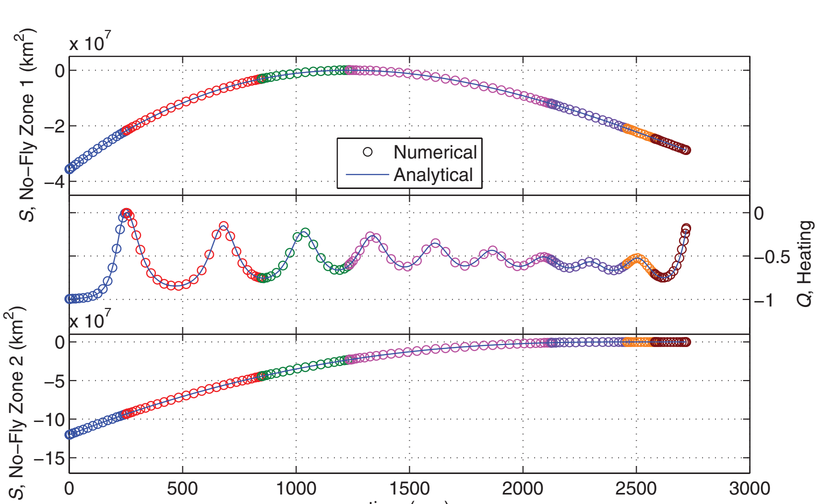
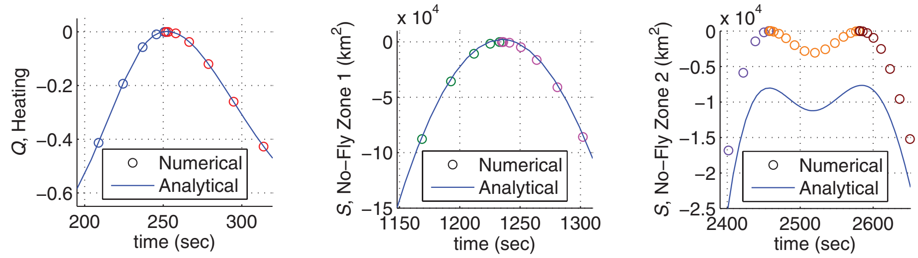


Fig. 8 Path constraints: seven phase numerical and analytical comparison.



a) b) c)

Fig. 9 Path constraints expanded: seven phase numerical and analytical comparison.

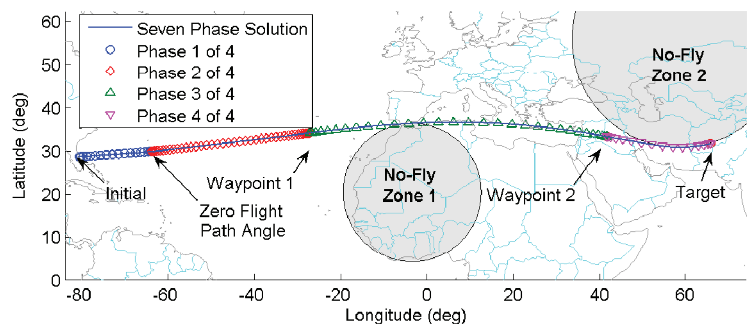


Fig. 10 Trajectory: four versus seven phase comparison.

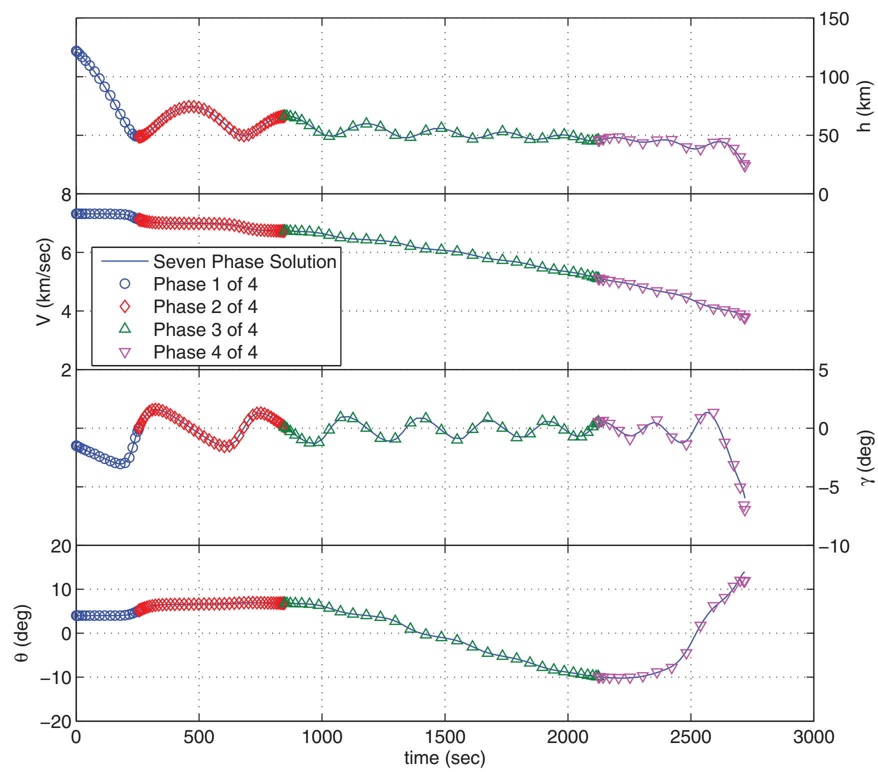


Fig. 11 States: four versus seven phase comparison.

E. 3-D CAV: Numerical Comparison of Phase Breakpoints

This rigorous analysis above forces this problem to be broken up into many phases, several requiring advanced knowledge of the solution, that is, the number of boundary contacts. The next objective is to evaluate the solution effectiveness using fewer phases, as seen in Fig. 10.

For this mission, there are two known intermediate waypoints. Therefore, there are at least three phases: t_0 to $t_{i=1}$, $t_{i=1}$ to $t_{i=2}$, and $t_{i=2}$ to t_f. Making some initial assumptions about the solution, but without confining the solution, a fourth phase is added. This phase ends at the first point the flight path angle passes through zero. The rational that is the first assault on the heating limit requires the most accuracy. Therefore, by placing a phase transition near the potential heating constraint contact, additional solution accuracy is achieved. The rigorous seven phase results are compared to the much simpler four phase results in Figs. 10–15.

Figure 13 shows some notable differences in the controls, especially at the beginning and end of the trajectory. At the beginning

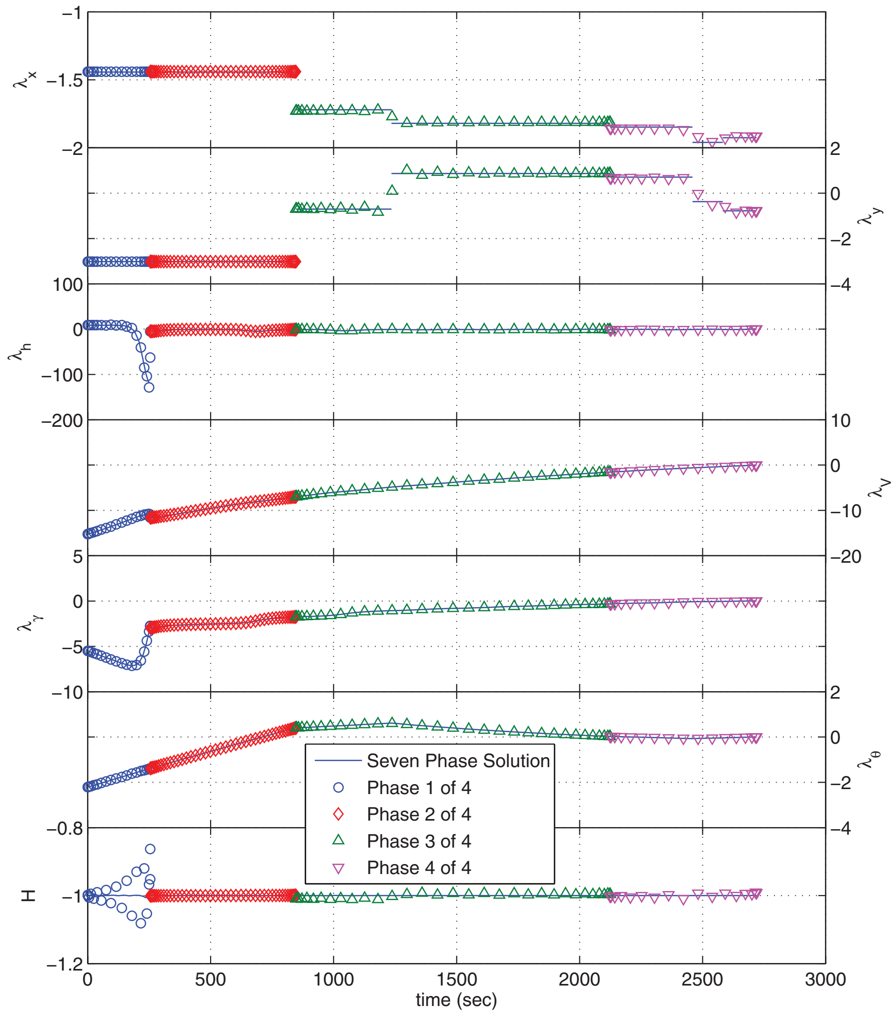


Fig. 12 Costates and Hamiltonian: four versus seven phase comparison.

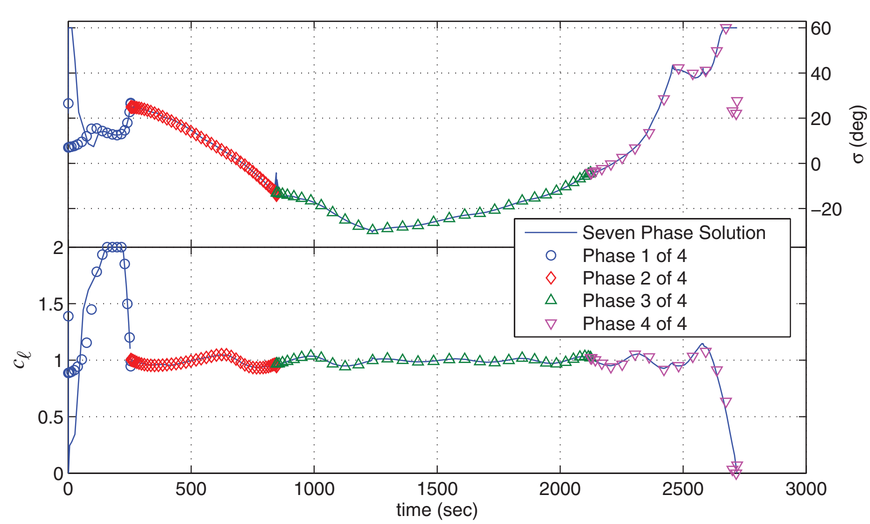


Fig. 13 Controls: four versus seven phase comparison.

of the trajectory, due to the extremely low atmospheric density, the controls are rather ineffective; therefore, seemingly large differences in the control should have very little impact on the trajectory. For the bank control at the end, there seems to be a timing issue as to when the maximum bank will be released. This is likely a result of the terminal costates going to zero. The state matching in Fig. 11 is arguably a better indication of a successful comparison. This shows the solutions have converged to the same trajectory.

The closeness in results demonstrates that the jump in costates due to path constraint contact can be estimated as continuous. Quantitatively, the final times for the seven phase and four phase solutions are 2719.42 and 2719.18 s, respectively, and thus, well within the accuracy of the model. Future analysis is required to generalize this small error to higher numbers of waypoints, no-fly zones, and heating constraint contacts. The setup for the four phase problem is much simpler and it does not require advanced knowledge

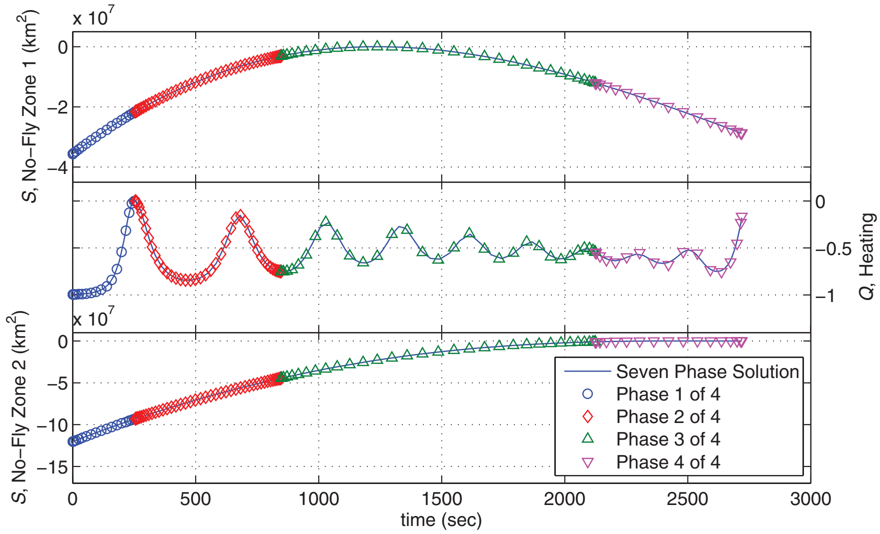
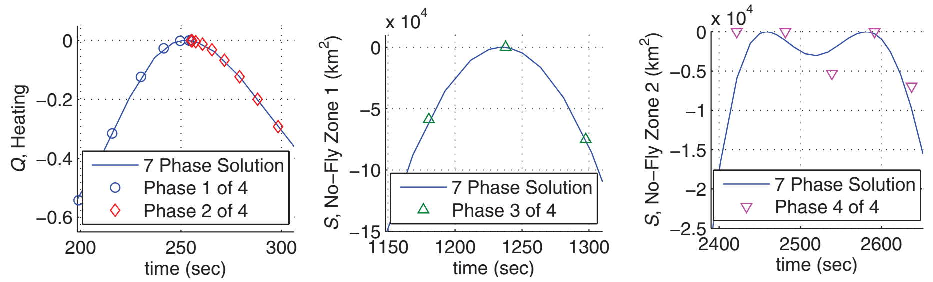


Fig. 14 Path constraints: four versus seven phase comparison.



a) b) c)

Fig. 15 Path constraints expanded: four versus seven phase comparison.

of the number of times the path constraints are contacted. The results herein show that solution accuracy can be maintained even with unpredicted heating and no-fly zone boundary contact. Therefore, the problem setup can be automated because the number of phases is only dependent on the number of waypoints.

F. Summary of Analysis

The development herein built upon itself to eventually determine and verify a viable optimal reentry trajectory solution technique. The simpler 2-D HCV case is used as a foundation to compare the analytical and numerical techniques. Bryson and Ho's path constraint method [37] is seen to differ from that used in the numerical collocation software. Understanding this difference provides insight into which costates are expected to have discontinuities, therefore, providing a means of replicating the results using analytical methods. The more complex 3-D CAV model demonstrates the ability to incorporate all of the vehicle and mission constraints. Using dynamic optimal control theory, an analytical solution is derived to validate these numerical results. After verifying that the numerical results are indeed satisfying the optimality criteria, the next step is to simplify implementation. By replicating the numerical results using just four phases, instead of the full seven phases, demonstrates that a priori knowledge of path constraint contact is unnecessary. Therefore, an automated problem setup technique can be developed primarily dependent on the number of waypoints.

IV. Conclusions

This research successfully determined and verified a numerical technique capable of generating an optimal reentry trajectory satisfying all of the specified vehicle and mission constraints. The suitability for satisfying the global strike mission also includes the autonomy and user implementation of the solution technique. The solution is considered autonomous due to several factors: manual manipulation of variables is not required, the problem has a repeatable setup structure, and the solution can converge even with a very simple point-to-target initial guess. The repeatability is the structured setup based on the number of waypoints and convergence independent of the number of no-fly zones or heating constraint contacts. The desirable user implementation is commensurate with the autonomy, namely, repeatability and no a priori solution requirements. The next measure of mission suitability is the adaptability to alternate missions. Because of the ability to set up independent phases, the number of waypoints should not hinder solution performance. Also, because the knowledge of the number of no-fly zone and heating constraint contacts is not required in the problem setup, this technique dynamically determines the optimal solution. And last, user implementation does not require extensive training or excessive intervention to converge to a solution. Knowledge of optimal control theory is not required, and the described problem setup minimized user intervention or iterations. Thus, once the cost and dynamics are defined, the phases can be automatically separated based on the given waypoints and if a feasible solution exists, convergence has been extremely fast.

Future applications research may include convergence and automation. Research could investigate increasing the likelihood, and speed, of convergence to a solution. Some candidate investigations could include the feasibility of the initial guess, node buildup, and implementing soft constraints instead of hard constraints. Currently, the initial guess can be any control and any time history of states; whereas, it may be more advantageous to simply start with a guess of the initial control and integrate that forward to become the initial guess of the states. The solution may not be optimal, but it will immediately satisfy the differential

Table A1 CAV-H aerodynamic data[a]

AOA	Mach 3.5	Mach 5	Mach 8	Mach 10	Mach 15	Mach 20	Mach 23
Lift-to-drag ratio, L/D							
10 deg	2.2000	2.5000	3.1000	3.5000	3.3846	3.2692	3.2000
15 deg	2.5000	2.6616	2.9846	3.2000	3.0846	2.9692	2.9000
20 deg	2.2000	2.3616	2.6846	2.9000	2.7846	2.6692	2.6000
Coefficient of lift, C_L							
AOA	Mach 3.5	Mach 5	Mach 8	Mach 10	Mach 15	Mach 20	Mach 23
10 deg	0.4500	0.4250	0.4000	0.3800	0.3700	0.3600	0.3500
15 deg	0.7400	0.7000	0.6700	0.6300	0.6000	0.5700	0.5570
20 deg	1.0500	1.0000	0.9500	0.9000	0.8500	0.8000	0.7800
Coefficient of drag, C_D							
AOA	Mach 3.5	Mach 5	Mach 8	Mach 10	Mach 15	Mach 20	Mach 23
10 deg	0.2045	0.1700	0.1290	0.1090	0.1090	0.1090	0.1090
15 deg	0.2960	0.2630	0.2240	0.1970	0.1950	0.1920	0.1920
20 deg	0.4770	0.4230	0.3540	0.3100	0.3050	0.3000	0.3000

[a]CAV-H reference area $S_{\mathrm{ref}} = 750$ in.^{2}; CAV-H mass $m = 2000$ lbs $= 907.186$ kg (assuming g $= 32.174$ ft/s^2).

equations. Increasing the number of nodes typically increases the accuracy of the solution. Initial results have shown that is it faster to start with a lower number of nodes to get a crude solution, and then feed that back in as an initial guess to a higher node solution. Simply starting with a bad guess, with a high number of nodes, led to an erratic control time history. The current implementation has the waypoints and no-fly zones as hard constraints, meaning they cannot be violated. The consequence is an increased potential for an infeasible solution, even if the vehicle only ran out of energy a very short distance from the target. Alternatively, the waypoints, no-fly zones, and target could be set up as costs that are proportional to proximity. The initial conditions could also be cost driven. For example, if the range makes the problem infeasible, a potential solution would be to increase the initial altitude and/or velocity. This could be achieved by putting a cost on the initial condition, rather than mandating a specified value. Other research could investigate alternate objectives, such as maximizing the terminal energy or terminal velocity. The final flight path angle could also be specified to vertical, or within a tolerance of vertical.

The benefit of increased feasibility is its compatibility with automation. The more autonomous the solution process the less user intervention is required. The results herein were run on a 2 GHz personal computer with solutions converging in less than a minute when starting from a simple point-to-target guess. The computational efficiency and automation potential makes these multiple-phase pseudospectral methods ideal candidates for real time, or near real time, on-vehicle implementation.

Appendix A: CAV Aerodynamic Data

The following is taken from [35] and is used to create a Mach independent model for use in this research.

The coefficient of drag C_D is assumed to be a parabolic function of the coefficient of lift C_L [47]:

$$C_D = C_{D_0} + K C_L^2 \tag{A1}$$

Taking this definition of C_D leads to an expression for lift over drag, L/D:

$$\frac{L}{D} = \frac{C_L}{C_D} = \frac{C_L}{C_{D_0} + K C_L^2} \tag{A2}$$

To find the C_L that will produce the maximum L/D, take the partial derivative of Eq. (A2) with respect to C_L:

$$\frac{\partial(L/D)}{\partial C_L} = \frac{1}{C_{D_0} + K C_L^2} - \frac{2 K C_L^2}{(C_{D_0} + K C_L^2)^2} \tag{A3}$$

Set Eq. (A3) equal to zero and solve for the optimal C_L, C_L^*:

$$C_{D_0} + K C_L^2 - 2 K C_L^2 = 0 \tag{A4}$$

Thus,

$$C_L^* = \sqrt{\frac{C_{D_0}}{K}} \tag{A5}$$

Input this equation into Eq. (A1) to find the corresponding C_D, C_D^*:

$$C_D^* = C_{D_0} + K\left(\sqrt{\frac{C_{D_0}}{K}}\right)^2 \tag{A6}$$

$$C_D^* = 2 C_{D_0} \tag{A7}$$

Then the maximum L/D is

$$E^* = \frac{C_L^*}{C_D^*} \tag{A8}$$

The normalized coefficient of lift is defined as

$$c_\ell = \frac{C_L}{C_L^*} \tag{A9}$$

The values in Table A1 are plotted in Fig. A1 with the curve fit using

$$C_L^* = 0.45 \tag{A10a}$$

$$E^* = 3.24 \tag{A10b}$$

Assuming $C_{L_{\max}} = 0.9$ from Fig. A1

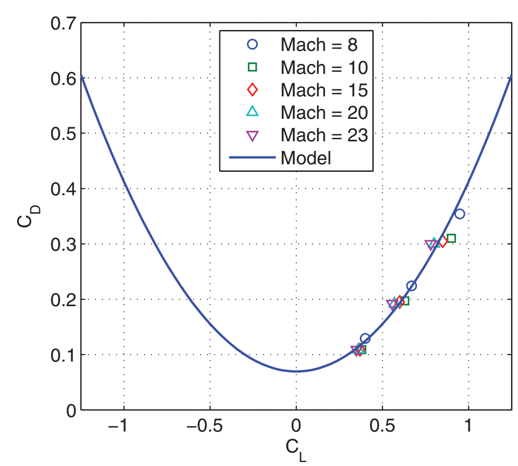


Fig. A1 CAV-H aerodynamic data and model.

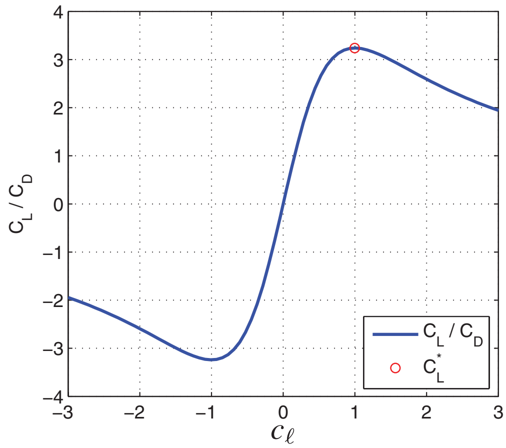


Fig. A2 CAV-H nondimensional variable c_ℓ.

$$c_{\ell_{\max}} = \frac{C_{L_{\max}}}{C_L^*} = 2.0 \tag{A10c}$$

From Fig. A1 the Mach independence appears valid; therefore, for this research Eq. (A10) is used to represent the CAV model. In Fig. A2 the new nondimensional variable c_ℓ is plotted to verify the $c_\ell = 1$ corresponding to the maximum L/D.

Appendix B: Gauss Pseudospectral Method Exposition

For completeness, an introduction to the Gauss pseudospectral method (GPM), which is the basis for the software package used in this research, is presented next. The GPM is an orthogonal collocation method where the collocation points are the Legendre–Gauss (LG) points. The description herein is a compilation from [29,64,65,70–72] which is based on [60,73]. Additional implementation methods are in [28,65,74,75].

I. Continuous Bolza Problem

The dynamic optimization problem of Sec. II.A.1 is restated here with the transformation of the independent variable t:

$$t = \frac{t_f - t_0}{2}\tau + \frac{t_f + t_0}{2} \tag{B1}$$

The optimal control problem is to determine the state $\mathbf{x}(\tau) \in \mathbb{R}^n$, control $\mathbf{u}(\tau) \in \mathbb{R}^m$, initial time t_0, and final time t_f that minimizes the cost functional:

$$J = \phi(\mathbf{x}(-1), t_0, \mathbf{x}(1), t_f) + \frac{t_f - t_0}{2}\int_{-1}^{1} L(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau; t_0, t_f)\,\mathrm{d}\tau \tag{B2}$$

subject to the constraints

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} = \frac{t_f - t_0}{2} f(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau; t_o, t_f) \tag{B3}$$

$$\psi(\mathbf{x}(-1), t_0, \mathbf{x}(1), t_f) = 0 \tag{B4}$$

$$C(\mathbf{x}(\tau), \mathbf{u}(\tau), \tau; t_o, t_f) \le 0 \tag{B5}$$

Herein, the optimal control problem of Eqs. (B2–B5) is called the continuous Bolza problem.

II. Gauss Pseudospectral Discretization of Continuous Bolza Problem

The direct approach to solving the continuous Bolza optimal control problem of Sec. B.I is to discretize and transcribe Eqs. (B2–B5) to a nonlinear programming problem (NLP). The Gauss pseudospectral method, similar to Legendre and Chebyshev methods, is based on approximating the state and control trajectories using interpolating polynomials. The state is approximated using a basis of $N+1$ Lagrange interpolating polynomials, $\mathcal{L}$,

$$\mathbf{x}(\tau) \approx \mathbf{X}(\tau) = \sum_{i=0}^{N} \mathbf{X}(\tau_i)\mathcal{L}_i(\tau) \tag{B6}$$

where $\mathcal{L}_i(\tau)$ $(i = 0, \ldots, N)$ are defined as

$$\mathcal{L}_i(\tau) = \prod_{j=0, j\ne i}^{N} \frac{\tau - \tau_j}{\tau_i - \tau_j} \tag{B7}$$

Additionally, the control is approximated using a basis of N Lagrange interpolating polynomials $\mathcal{L}_i^*(\tau)$ $(i = 1, \ldots, N)$ as

$$\mathbf{u}(\tau) \approx \mathbf{U}(\tau) = \sum_{i=1}^{N} \mathbf{U}(\tau_i)\mathcal{L}_i^*(\tau) \tag{B8}$$

where

$$\mathcal{L}_i^*(\tau) = \prod_{j=1, j\ne i}^{N} \frac{\tau - \tau_j}{\tau_i - \tau_j} \tag{B9}$$

It can be seen from Eqs. (B7) and (B9) that $\mathcal{L}_i(\tau)$ $(i = 0, \ldots, N)$ and $\mathcal{L}_i^*(\tau)$ $(i = 1, \ldots, N)$ satisfy the properties

$$\mathcal{L}_i(\tau_j) = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases} \quad \text{and} \quad \mathcal{L}_i^*(\tau_j) = \begin{cases} 1, & i = j \\ 0, & i \ne j \end{cases} \tag{B10}$$

Differentiating the expression in Eq. (B6) produces

$$\dot{\mathbf{x}}(\tau) \approx \dot{\mathbf{X}}(\tau) = \sum_{i=0}^{N} \mathbf{X}(\tau_i)\dot{\mathcal{L}}_i(\tau) \tag{B11}$$

The derivative of each Lagrange polynomial at the LG points can be represented in a differential approximation matrix, $D \in \mathbb{R}^{\mathcal{N}\times\mathcal{N}+1}$. The elements of the differential approximation matrix are determined offline as follows:

$$D_{ki} = \dot{\mathcal{L}}_i(\tau_k) = \sum_{l=0}^{N} \frac{\prod_{j=0, j\ne i,l}^{N}(\tau_k - \tau_j)}{\prod_{j=0, j\ne i}^{N}(\tau_i - \tau_j)} \tag{B12}$$

where $k = 1, \ldots, N$ and $i = 0, \ldots, N$. The dynamic constraint is transcribed into algebraic constraints via the differential approximation matrix as follows:

$$\sum_{i=0}^{N} D_{ki}\mathbf{X}_i - \frac{t_f - t_0}{2} f(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) = 0 \qquad (k = 1, \ldots, N) \tag{B13}$$

where $\mathbf{X}_k \equiv \mathbf{X}(\tau_k) \in \mathbb{R}^n$ and $\mathbf{U}_k \equiv \mathbf{U}(\tau_k) \in \mathbb{R}^m$ $(k = 1, \ldots, N)$. Note that the dynamic constraint is collocated only at the LG points and *not* at the boundary points (this form of collocation differs from other well-known pseudospectral methods [75,76]). Additional variables in the discretization are defined as follows: $\mathbf{X}_0 \equiv \mathbf{X}(-1)$, and $\mathbf{X}_f \equiv \mathbf{X}(1)$, where $\mathbf{X}_f$ is defined in terms of $\mathbf{X}_k$ $(k = 0, \ldots, N)$ and $\mathbf{U}(\tau_k)$ $(k = 1, \ldots, N)$ via the Gauss quadrature [77]:

$$\mathbf{X}_f = \mathbf{X}_0 + \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k f(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) \tag{B14}$$

The continuous cost function of Eq. (B2) is approximated using a Gauss quadrature [77] as

$$J = \phi(\mathbf{X}_0, t_0, \mathbf{X}_f, t_f) + \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k L(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) \tag{B15}$$

where w_k are the Gauss weights. The boundary constraint of Eq. (B4) is expressed as

$$\psi(\mathbf{X}_0, t_0, \mathbf{X}_f, t_f) = 0 \tag{B16}$$

Furthermore, the path constraint of Eq. (B5) is evaluated at the LG points as

$$C(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) \le 0 \qquad (k = 1, \ldots, N) \tag{B17}$$

The cost function of Eq. (B15) and the algebraic constraints of Eqs. (B13), (B14), (B16), and (B17) define an NLP whose solution is an approximate solution to the continuous Bolza problem. Finally, it is noted that the above discretization can be employed in multiple-phase problems by transcribing the problem in each phase using the above discretization and connecting the phases by *linkage* constraints:

$$\begin{gathered}\mathbf{P}^{(s)}\Big(\mathbf{x}^{(p_l^s)}(t_f), t_f^{(p_l^s)}; \mathbf{q}^{(p_l^s)}, \mathbf{x}^{(p_u^s)}(t_0), t_0^{(p_u^s)}; \mathbf{q}^{(p_u^s)}\Big) = 0 \\ (p_l, p_u \in [1, \ldots, P], s = 1, \ldots, L_p)\end{gathered} \tag{B18}$$

where $\mathbf{x}^{(p)}(t) \in \mathbb{R}^{n_p}$, $\mathbf{u}^{(p)}(t) \in \mathbb{R}^{m_p}$, $\mathbf{q}^{(p)} \in \mathbb{R}^{q_p}$, and $t \in \mathbb{R}$ are, respectively, the state, control, static parameters, and time in phase $p \in [1, \ldots, P]$, L_p is the number of phases to be linked, $p_l^s \in [1, \ldots, P]$ $(s = 1, \ldots, L_p)$ are the "left" phase numbers, and $p_u^s \in [1, \ldots, P]$ $(s = 1, \ldots, L_p)$ are the "right" phase numbers, where the sum of the cost for each phase:

$$J = \sum_{p=1}^{P} J^{(p)} \tag{B19}$$

becomes the final cost to be minimized.

III. Karush–Kuhn–Tucker Conditions of the Nonlinear Programming Problem

The first-order optimality conditions (i.e., the KKT conditions) of the NLP can be obtained using the augmented cost function or Lagrangian. The augmented cost function is formed using the Lagrange multipliers $\tilde{\mathbf{\Lambda}}_k \in \mathbb{R}^n$, $\tilde{\mu}_k \in \mathbb{R}^c$, $k = 1, \ldots, N$, $\tilde{\mathbf{\Lambda}}_F \in \mathbb{R}^n$, and $\tilde{\nu} \in \mathbb{R}^q$ as

$$\begin{aligned}J_a = {} & \phi(\mathbf{X}_0, t_0, \mathbf{X}_f, t_f) + \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k L(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f) \\ & - \tilde{\nu}^T \psi(\mathbf{X}_0, t_0, \mathbf{X}_f, t_f) - \sum_{k=1}^{N} \tilde{\mu}_k^T C(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_o, t_f) \\ & - \sum_{k=1}^{N} \tilde{\mathbf{\Lambda}}_k^T \Big(\sum_{i=1}^{N} D_k \mathbf{X}_i - \frac{t_f - t_0}{2} f(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f)\Big) \\ & - \tilde{\mathbf{\Lambda}}_F^T \Big(\mathbf{X}_f - \mathbf{X}_0 - \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k f(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_o, t_f)\Big)\end{aligned} \tag{B20}$$

The KKT conditions are found by setting the derivatives of the Lagrangian equal to zero with respect to $\mathbf{X}_0, \mathbf{X}_k, \mathbf{X}_f, \mathbf{U}_k, \tilde{\mathbf{\Lambda}}_k, \tilde{\mu}_k, \tilde{\mathbf{\Lambda}}_F$, $\tilde{\nu}$, t_0, and t_f. The solution to the NLP of Sec. B.II must satisfy the following KKT conditions:

$$\sum_{i=0}^{N} \mathbf{X}_i D_{ki} = \frac{t_f - t_0}{2} f_k \tag{B21}$$

$$\begin{aligned}&\sum_{i=1}^{N}\Big(\frac{\tilde{\mathbf{\Lambda}}_i^T}{w_i} + \tilde{\mathbf{\Lambda}}_F^T\Big) D_{ki}^{\dagger} + \tilde{\mathbf{\Lambda}}_F^T D_{kN+1}^{\dagger} \\ &= \frac{t_f - t_0}{2}\Big(\frac{\partial L_k}{\partial \mathbf{X}_k} - \Big(\frac{\tilde{\mathbf{\Lambda}}_k^T}{w_k} + \tilde{\mathbf{\Lambda}}_F^T\Big)\frac{\partial f_k}{\partial \mathbf{X}_k} + \frac{2}{t_f - t_0}\frac{\tilde{\mu}_k^T}{w_k}\frac{\partial C_k}{\partial \mathbf{X}_k}\Big)\end{aligned} \tag{B22}$$

$$0 = \frac{\partial L_k}{\partial \mathbf{U}_k} + \Big(\frac{\tilde{\mathbf{\Lambda}}_k^T}{w_k} + \tilde{\mathbf{\Lambda}}_F^T\Big)\frac{\partial f_k}{\partial \mathbf{U}_k} - \frac{2}{t_f - t_0}\frac{\tilde{\mu}_k^T}{w_k}\frac{\partial C_k}{\partial \mathbf{U}_k} \tag{B23}$$

$$\psi(\mathbf{X}_o, t_0, \mathbf{X}_f, t_f) = 0 \tag{B24}$$

$$\tilde{\mathbf{\Lambda}}_0^T = -\frac{\partial \phi}{\partial \mathbf{X}_0} + \tilde{\nu}^T \frac{\partial \psi}{\partial \mathbf{X}_0} \tag{B25}$$

$$\tilde{\mathbf{\Lambda}}_F^T = \frac{\partial \phi}{\mathbf{X}_f} - \tilde{\nu}^T \frac{\partial \psi}{\mathbf{X}_f} \tag{B26}$$

$$-\frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k \frac{\partial \tilde{H}_k}{\partial t_0} + \frac{1}{2}\sum_{k=1}^{N} w_k \tilde{H}_k = \frac{\partial \phi}{\partial t_0} - \tilde{\nu}^T \frac{\partial \psi}{\partial t_0} \tag{B27}$$

$$\frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k \frac{\partial \tilde{H}_k}{\partial t_f} + \frac{1}{2}\sum_{k=1}^{N} w_k \tilde{H}_k = -\frac{\partial \phi}{\partial t_f} - \tilde{\nu}^T \frac{\partial \psi}{\partial t_f} \tag{B28}$$

$$C_k \le 0 \tag{B29}$$

$$\tilde{\mu}_{jk} = 0, \quad \text{when } C_{jk} < 0 \tag{B30}$$

$$\tilde{\mu}_{jk} \le 0, \quad \text{when } C_{jk} = 0 \tag{B31}$$

$$\mathbf{X}_f = \mathbf{X}_0 + \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k f_k \tag{B32}$$

$$\begin{aligned}\tilde{\mathbf{\Lambda}}_F = \tilde{\mathbf{\Lambda}}_0 + \frac{t_f - t_0}{2}\sum_{i=1}^{N} w_k \Big(&-\frac{\partial L_k}{\partial \mathbf{X}_k} \\ &- \Big(\frac{\tilde{\mathbf{\Lambda}}_k^T}{w_k} + \tilde{\mathbf{\Lambda}}_F^T\Big)\frac{\partial f_k}{\partial \mathbf{X}_k} + \frac{2}{t_f - t_0}\frac{\tilde{\mu}_k^T}{w_k}\frac{\partial C_k}{\partial \mathbf{X}_k}\Big)\end{aligned} \tag{B33}$$

where the shorthand notation

$$L_k \equiv L(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f)$$

$$f_k \equiv f(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_o, t_f)$$

$$\tilde{H}_k \equiv \tilde{H}_k(\mathbf{X}_k, \tilde{\mathbf{\Lambda}}_k, \tilde{\mu}_k, \mathbf{U}_k, \tau_k; t_0, t_f)$$

and

$$C_{jk} \equiv C_j(\mathbf{X}_k, \mathbf{U}_k, \tau_k; t_0, t_f)$$

is used. The augmented Hamiltonian $\tilde{H}_k$ is defined as

$$\tilde{H}_k \equiv L_k + \Big(\frac{\tilde{\mathbf{\Lambda}}_k^T}{w_k} + \tilde{\mathbf{\Lambda}}_F^T\Big) f_k - \frac{2}{t_f - t_0}\frac{\tilde{\mu}_k^T}{w_k} C_k \tag{B34}$$

and $\tilde{\mathbf{\Lambda}}_0$ is defined as

$$\tilde{\mathbf{\Lambda}}_0^T = -\frac{\partial \phi}{\partial \mathbf{X}_0} + \tilde{\nu}^T \frac{\partial \psi}{\partial \mathbf{X}_0} \tag{B35}$$

Note that the sign of the $\tilde{\mu}_k^T$ term is opposite to the μ^T term in Eq. (9); thus, the software output requires a sign change to match the analytical results presented before this Appendix.

IV. First-Order Optimality Conditions of Continuous Bolza Problem

The indirect approach to solving the continuous Bolza problem of Eqs. (B2–B5) in Sec. B.I is to apply the calculus of variations and Pontryagin et al.'s maximum principle [78] to obtain the first-order necessary conditions for optimality [79]. These variational conditions are typically derived using the augmented Hamiltonian H defined as

$$H(\mathbf{x},\lambda,\mu,\mathbf{u},\tau;t_0,t_f) = L(\mathbf{x},\mathbf{u},\tau;t_0,t_f) + \lambda^T(\tau)f(\mathbf{x},\mathbf{u},\tau;t_o,t_f) - \mu^T(\tau)C(\mathbf{x},\mathbf{u},\tau;t_0,t_f) \tag{B36}$$

where $\lambda(\tau)\in\mathbb{R}^n$ is the costate and $\mu(\tau)\in\mathbb{R}^c$ is the Lagrange multiplier associated with the path constraint. The continuous-time first-order optimality conditions can be shown to be

$$\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}\tau} = \frac{t_f-t_0}{2}f(\mathbf{x},\mathbf{u},\tau;t_0,t_f) = \frac{t_f-t_0}{2}\frac{\partial H}{\partial\lambda} \tag{B37a}$$

$$\frac{\mathrm{d}\lambda}{\mathrm{d}\tau} = \frac{t_f-t_0}{2}\left(-\frac{\partial L}{\partial\mathbf{x}} - \lambda^T\frac{\partial f}{\partial\mathbf{x}} + \mu^T\frac{\partial C}{\partial\mathbf{x}}\right) = -\frac{t_f-t_0}{2}\frac{\partial H}{\partial\mathbf{x}} \tag{B37b}$$

$$0 = \frac{\partial L}{\partial\mathbf{u}} + \lambda^T\frac{\partial f}{\partial\mathbf{u}} - \mu^T\frac{\partial C}{\partial\mathbf{u}} = \frac{\partial H}{\partial\mathbf{u}} \tag{B37c}$$

$$\psi(\mathbf{x}(\tau_0),t_0,\mathbf{x}(\tau_f),t_f) = 0 \tag{B37d}$$

$$\lambda(\tau_0) = -\frac{\partial\phi}{\partial\mathbf{x}(\tau_0)} + \nu^T\frac{\partial\psi}{\partial\mathbf{x}(\tau_0)}, \qquad \lambda(\tau_f) = \frac{\partial\phi}{\partial\mathbf{x}(\tau_f)} - \nu^T\frac{\partial\psi}{\partial\mathbf{x}(\tau_f)} \tag{B37e}$$

$$H(t_0) = \frac{\partial\phi}{\partial t_0} - \nu^T\frac{\partial\phi}{\partial t_0}, \qquad H(t_f) = -\frac{\partial\phi}{\partial t_f} + \nu^T\frac{\partial\phi}{\partial t_f} \tag{B37f}$$

$$\mu_j(\tau) = 0, \quad \text{when } C_j(\mathbf{x},\mathbf{u},\tau;t_0,t_f) < 0, \qquad j=1,\dots,c \tag{B37g}$$

$$\mu_j(\tau) \le 0, \quad \text{when } C_j(\mathbf{x},\mathbf{u},\tau;t_0,t_f) = 0, \qquad j=1,\dots,c \tag{B37h}$$

where $\nu\in\mathbb{R}^q$ is the Lagrange multiplier associated with the boundary condition ψ. It can be shown that the augmented Hamiltonian at the initial and final times can be written, respectively, as

$$H(t_0) = -\frac{t_f-t_0}{2}\int_{-1}^{1}\frac{\partial H}{\partial t_o}\,\mathrm{d}\tau + \frac{1}{2}\int_{-1}^{1}H\,\mathrm{d}\tau \tag{B38}$$

$$H(t_f) = \frac{t_f-t_0}{2}\int_{-1}^{1}\frac{\partial H}{\partial t_f}\,\mathrm{d}\tau + \frac{1}{2}\int_{-1}^{1}H\,\mathrm{d}\tau \tag{B39}$$

V. Gauss Pseudospectral Discretization Necessary Conditions

To discretize the variational conditions of Sec. B.IV using the Gauss pseudospectral discretization, it is necessary to form a suitable approximation for the costates. In this method, the costate $\lambda(\tau)$ is approximated as follows:

$$\lambda(\tau) \approx \Lambda(\tau) = \sum_{i=1}^{N+1}\lambda(\tau_i)\mathcal{L}_i^\dagger(\tau) \tag{B40}$$

where $\mathcal{L}_i^\dagger(\tau)$ $(i=1,\dots,N+1)$ are defined as

$$\mathcal{L}_i^\dagger(\tau) = \prod_{j=1,j\ne i}^{N+1}\frac{\tau-\tau_j}{\tau_i-\tau_j} \tag{B41}$$

It is emphasized that the costate approximation is *different* from the state approximation. In particular, the basis of the $N+1$ Lagrange interpolating polynomials $\mathcal{L}_i^\dagger(\tau)$ $(i=1,\dots,N+1)$ includes the costate at the *final* time (as opposed to the initial time which is used in the state approximation). This (nonintuitive) costate approximation is necessary to provide a complete mapping between the KKT conditions and the variational conditions.

Using the costate approximation of Eq. (B40), the first-order necessary conditions of the continuous Bolza problem in Eq. (B37) are discretized as follows. First, the state and control are approximated using Eqs. (B6) and (B8), respectively. Next, the costate is approximated using the basis of $N+1$ Lagrange interpolating polynomials as defined in Eq. (B40). The continuous-time first-order optimality conditions of Eq. (B37) are discretized using the variables $\mathbf{X}_0\equiv\mathbf{X}(-1)$, $\mathbf{X}_k\equiv\mathbf{X}(\tau_k)\in\mathbb{R}^n$, and $\mathbf{X}_f\equiv\mathbf{X}(1)$ for the state; $\mathbf{U}_k\equiv\mathbf{U}(\tau_k)\in\mathbb{R}^m$ for the control; $\mathbf{\Lambda}_0\equiv\mathbf{\Lambda}(-1)$, $\mathbf{\Lambda}_k\equiv\mathbf{\Lambda}(\tau_k)\in\mathbb{R}^n$, and $\mathbf{\Lambda}_F\equiv\mathbf{\Lambda}(1)$ for the costates; and $\mu_k\equiv\mu(\tau_k)\in\mathbb{R}^c$ for the Lagrange multiplier associated with the path constraints at the LG points $k=1,\dots,N$. The other unknown variables in the problem are the initial and final times, $t_0\in\mathbb{R}$, $t_f\in\mathbb{R}$, and the Lagrange multiplier, $\nu\in\mathbb{R}^q$. The total number of variables is then given as $(2n+m+c)N+4n+q+2$. These variables are used to discretize the continuous necessary conditions of Eq. (B47) via the Gauss pseudospectral discretization. Note that the derivative of the state is approximated using Lagrange polynomials based on $N+1$ points consisting of the N LG points and the initial time τ_0, while the derivative of the costate is approximated using Lagrange polynomials based on $N+1$ points consisting of the N LG points and the final time, τ_f. The resulting algebraic equations that approximate the continuous necessary conditions at the LG points are given as

$$\sum_{i=1}^{N}\mathbf{X}_iD_{ki} = \frac{t_f-t_0}{2}f_k \tag{B42}$$

$$\sum_{i=1}^{N}\mathbf{\Lambda}_iD_{ki}^\dagger + \mathbf{\Lambda}_FD_{kN+1}^\dagger = \frac{t_f-t_0}{2}\left(-\frac{\partial L_k}{\partial\mathbf{X}_k} - \mathbf{\Lambda}_k^T\frac{\partial f_k}{\partial\mathbf{X}_k} + \mu_k^T\frac{\partial C_k}{\partial\mathbf{X}_k}\right) \tag{B43}$$

$$0 = \frac{\partial L_k}{\partial\mathbf{U}_k} + \mathbf{\Lambda}_k^T\frac{\partial f_k}{\partial\mathbf{U}_k} - \mu_k^T\frac{\partial C_k}{\partial\mathbf{U}_k} \tag{B44}$$

$$\psi(\mathbf{X}_0,t_0,\mathbf{X}_f,t_f) = 0 \tag{B45}$$

$$\mathbf{\Lambda}_0 = -\frac{\partial\phi}{\partial\mathbf{X}_0} + \nu^T\frac{\partial\psi}{\partial\mathbf{X}_0} \tag{B46}$$

$$\mathbf{\Lambda}_F = \frac{\partial\phi}{\partial\mathbf{X}_f} - \nu^T\frac{\partial\psi}{\partial\mathbf{X}_f} \tag{B47}$$

$$-\frac{t_f-t_0}{2}\sum_{k=1}^{N}w_k\frac{\partial H_k}{\partial t_0} + \frac{1}{2}\sum_{k=1}^{N}w_kH_k = \frac{\partial\phi}{\partial t_0} - \nu^T\frac{\partial\psi}{\partial t_0} \tag{B48}$$

$$\frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k \frac{\partial H_k}{\partial t_f} + \frac{1}{2}\sum_{k=1}^{N} w_k H_k = -\frac{\partial \phi}{\partial t_f} + \nu^T \frac{\partial \psi}{\partial t_f} \tag{B49}$$

$$\mu_{jk} = 0, \quad \text{when } C_{jk} < 0 \tag{B50}$$

$$\mu_{jk} \leq 0, \quad \text{when } C_{jk} = 0 \tag{B51}$$

for $k = 1, \ldots, N$ and $j = 1, \ldots, c$. The final two equations that are required (in order to link the initial and final state and costate, respectively) are

$$\mathbf{X}_f = \mathbf{X}_0 + \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k f_k \tag{B52}$$

$$\mathbf{\Lambda}_F = \mathbf{\Lambda}_0 + \frac{t_f - t_0}{2}\sum_{k=1}^{N} w_k \left(-\frac{\partial L_k}{\partial \mathbf{X}_k} - \mathbf{\Lambda}_k^T \frac{\partial f_k}{\partial \mathbf{X}_k} + \mu_k^T \frac{\partial C_k}{\partial \mathbf{X}_k} \right) \tag{B53}$$

The total number of equations in the set of discrete necessary conditions of Eqs. (B42–B53) is $(2n + m + c)N + 4n + q + 2$ (the same number of unknown variables). Solving these nonlinear algebraic equations would be an indirect solution to the optimal control problem.

VI. Costate Estimate

One of the key features of the Gauss pseudospectral method is the ability to map the KKT multipliers of the NLP to the costates of the continuous-time optimal control problem. In particular, using the results of Secs. B.III and B.V a costate estimate for the continuous Bolza problem can be obtained at the Legendre–Gauss points and the boundary points. This costate estimate is taken from [73] and is summarized here via the Gauss pseudospectral costate mapping theorem:

Theorem B.1. Gauss Pseudospectral Costate Mapping Theorem: The KKT conditions of the NLP are exactly equivalent to the discretized form of the continuous first-order necessary conditions of the continuous Bolza problem when using the Gauss pseudospectral discretization. Furthermore, a costate estimate at the initial time, final time, and the Legendre–Gauss points can be found from the KKT multipliers, $\tilde{\mathbf{\Lambda}}_k$, $\tilde{\mu}_k$, $\tilde{\mathbf{\Lambda}}_F$, and $\tilde{\nu}$,

$$\begin{gathered}\mathbf{\Lambda}_k = \frac{\tilde{\mathbf{\Lambda}}_k}{w_k} + \tilde{\mathbf{\Lambda}}_F, \qquad \mu_k = \frac{2}{t_f - t_0}\frac{\tilde{\mu}_k}{w_k}, \qquad \nu = \tilde{\nu} \\ \mathbf{\Lambda}(t_0) = \tilde{\mathbf{\Lambda}}_o, \qquad \mathbf{\Lambda}(t_f) = \tilde{\mathbf{\Lambda}}_F\end{gathered} \tag{B54}$$

Using the substitutions of Eq. (B54), it can be seen that Eqs. (B21–B33) are exactly the same as Eqs. (B42–B53).

VII. Computation of Boundary Controls

It is seen in the GPM that the control is discretized only at the LG points and is not discretized at either the initial or the terminal point. Consequently, the solution of the NLP defined by Eqs. (B13–B17) does not produce values of the controls at the boundaries. The ability to obtain accurate initial and terminal controls can be important in many applications, particularly in guidance where real-time computation of the initial control is of vital interest.

At first glance, it may seem that the lack of control information at the boundaries can be overcome simply via extrapolation of the control at the LG points. However, multiple reasons exist as to why this is not the best approach. First, no particular functional form for the control is assumed in the GPM discretization. As a result, the best function to use for extrapolation is ambiguous. Second, any reasonable extrapolation of the control (e.g., linear, quadratic, cubic, or spline) may violate a path constraint which, in general, will render the extrapolated control infeasible. Third, even if the extrapolated control is feasible, it will not satisfy the required optimality conditions at the boundaries (i.e., the control will be suboptimal with respect to the NLP). Consequently, it is both practical and most rigorous to develop a systematic procedure to compute the boundary controls. The primal (state) and dual (costate) solutions of the NLP arising from the Gauss pseudospectral method are used to compute the boundary controls [72].

Computation of the initial control is done first because the approach for computing the terminal control is identical. First, recalling the augmented Hamiltonian H for the continuous-time optimal control problem in Eq. (B46) is

$$H(\mathbf{x}, \mathbf{u}, \lambda, \mu) \equiv L + \lambda^T f - \mu^T C \tag{B55}$$

where shorthand notation is used. Recall from the principle of Pontryagin, at every instant of time the optimal control is the control $\mathbf{u}^*(\tau) \in \mathcal{U}$ that satisfies the condition

$$H(\mathbf{x}^*, \mathbf{u}^*, \lambda^*, \mu^*) \leq H(\mathbf{x}^*, \mathbf{u}, \lambda^*, \mu^*) \tag{B56}$$

where $\mathcal{U}$ is the feasible control set. Consequently, for a given instant of time τ where $\mathbf{x}^*(\tau)$, $\lambda^*(\tau)$, and $\mu^*(\tau)$ are known, Eq. (B56) is a constrained optimization problem in $\mathbf{u}(\tau) \in \mathbb{R}^m$. To solve this constrained optimization problem at the *initial* time, it is necessary to know $\mathbf{x}^*(\tau_0)$, $\lambda^*(\tau_0)$, and $\mu^*(\tau_0)$.

Consider the information that can be obtained by solving the NLP associated with the GPM. In particular, the primal solution to the NLP produces $\mathbf{X}(\tau_0)$ while the dual solution to the NLP can be manipulated algebraically to obtain the initial costate, $\mathbf{\Lambda}(\tau_0)$. However, because the NLP does not evaluate the path constraint at the boundaries, there is no associated Lagrange multiplier $\mu^*(\tau_0)$. This apparent impediment can be overcome by applying the minimum principle in a manner somewhat different from that given in Eq. (B56). In particular, suppose we let $\mathcal{H}$ be the *Hamiltonian* (not the augmented Hamiltonian), where $\mathcal{H}$ is defined as

$$\mathcal{H}(\mathbf{x}, \mathbf{u}, \lambda) \equiv L + \lambda^T f \tag{B57}$$

It is seen in Eq. (B57) that the term involving the path constraint is not included. The path constraint is instead incorporated into the feasible control set. In particular, suppose we let $\mathcal{V}_0$ be

$$\mathcal{V}_0 = \mathcal{U} \bigcap \mathcal{C}_0 \tag{B58}$$

where $\mathcal{V}_0$ is the intersection of the original set of feasible controls at time τ_0, denoted $\mathcal{U}$, with the set of all controls at time τ_0 that satisfy the inequality constraint of Eq. (B27), denoted $\mathcal{C}_0$. Then, using the values of $\mathbf{X}(\tau_0)$ and $\mathbf{\Lambda}(\tau_0)$, the following modified optimization problem in m variables $\mathbf{U}(\tau_0) \in \mathbb{R}^m$ can be solved to obtain the initial control, $\mathbf{U}(\tau_0)$:

$$\min_{\mathbf{U}(\tau_0) \in \mathcal{V}_0} \mathcal{H}(\mathbf{X}(\tau_0), \mathbf{U}(\tau_0), \mathbf{\Lambda}(\tau_0), \tau_0; t_0, t_f) \tag{B59}$$

It is noted that, because $\mathcal{V}_0$ is restricted by the inequality path constraint at τ_0, the solution of $\mathbf{U}(\tau_0)$ is equivalent to the solution of the following problem:

$$\begin{gathered}\min_{\mathbf{U}(\tau_0) \in \mathcal{U}} \mathcal{H}(\mathbf{X}(\tau_0), \mathbf{U}(\tau_0), \mathbf{\Lambda}(\tau_0), \tau_0; t_0, t_f) \\ \text{subject to } C(\mathbf{X}(\tau_0), \mathbf{U}(\tau_0), \tau_0; t_0, t_f) \leq 0\end{gathered} \tag{B60}$$

Interestingly, if the constraint is *active*, then the initial path constraint multiplier $\mu^*(\tau_0)$ will also be determined by the minimization problem of Eq. (B60). Finally, similar to the initial time, the control at the terminal time $\mathbf{U}(\tau_f)$ can be obtained by solving the minimization problem of Eq. (B60) at $\tau = \tau_f$, that is,

$$\min_{\mathbf{U}(\tau_f)\in\mathcal{U}} \mathcal{H}(\mathbf{X}(\tau_f), \mathbf{U}(\tau_f), \mathbf{\Lambda}(\tau_f), \tau_f; t_0, t_f)$$
$$\text{subject to } C(\mathbf{X}(\tau_f), \mathbf{U}(\tau_f), \tau_f; t_0, t_f) \leq 0 \tag{B61}$$

The GPOCS is a software program written in MATLAB®[§] for solving multiple-phase optimal control problems, which implements the algorithm as described above.

References

[1] USAF, "Air Force Handbook—109th Congress," Department of the Air Force, Washington, D.C., 2005.
[2] Wolf, A. F., "ConventionalWarheads for Long-Range Ballistic Missiles: Background and Issues for Congress," CRS Report for Congress, 19 June 2007, Order Code RL33067.
[3] DARPA, FALCON Force Application and Launch from CONUS Task 1 Small Launch Vehicle (SLV)—Phase II, 2004, Program Solicitation Number 04-05.
[4] Hueter, U., and Hutt, J. J., "NASA's Next Generation Launch Technology Program-Next Generation Space Access Roadmap," AIAA Paper 2003-6941, Dec. 2003.
[5] Richie, G., and Haaren, J., *Striking from Space—The Future of Space Force Applications*, AIAA, Reston, VA, 1998.
[6] Harpold, J. C., and Graves, C. A., Jr., "Shuttle Entry Guidance," *Journal of the Astronautical Sciences*, Vol. 27, No. 3, July–Sept. 1979, pp. 239–268.
[7] Hanson, J. M., and Jones, R. E., "Test Results for Entry Guidance Methods for Space Vehicles," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 6, Nov.–Dec. 2004, pp. 960–966. doi:10.2514/1.10886
[8] Hanson, J. M., Coughlin, D. J., Dukeman, G. A., Mulqueen, J. A., and McCarter, J. W., "Ascent, Transition, Entry, and Abort Guidance Algorithm Design for the X-33 Vehicle," AIAA Paper AIAA-98-4409, 1998.
[9] Hanson, J., "Advanced Guidance and Control Project for Reusable Launch Vehicles," AIAA Paper 2000-3957, Aug. 2000.
[10] Dukeman, G. A., "Profile-Following Entry Guidance Using Linear Quadratic Regulator Theory," AIAA Paper 2002-4457, 2002.
[11] Zimmerman, C., Dukeman, G., and Hanson, J., "Automated Method to Compute Orbital Reentry Trajectories with Heating Constraints," *Journal of Guidance, Control, and Dynamics*, Vol. 26, No. 4, July–Aug. 2003, pp. 523–529. doi:10.2514/2.5096
[12] Saraf, A., Leavitt, J. A., Chen, D. T., and Mease, K. D., "Design and Evaluation of an Acceleration Guidance Algorithm for Entry," *Journal of Spacecraft and Rockets*, Vol. 41, No. 6, Nov.–Dec. 2004, pp. 986–996. doi:10.2514/1.11015.
[13] Shen, Z., and Lu, P.,"Onboard Generation of Three-Dimensional Constrained Entry Trajectories," *Journal of Guidance, Control, and Dynamics*, Vol. 26, No. 1, Jan.–Feb. 2003, pp. 111–121. doi:10.2514/2.5021
[14] Shen, Z., and Lu, P.,"On-Board Entry Trajectory Planning Expanded to Sub-Orbital Flight," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA, Reston, VA, 11–14 Aug. 2003.
[15] Schierman, J. D., Ward, D. G., Hull, J. R., Gandhi, N., Oppenheimer, M. W., and Doman, D. B., "Integrated Adaptive Guidance and Control for Re-Entry Vehicles with Flight Test Results," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 6, Nov.–Dec. 2004, pp. 975–988. doi:10.2514/1.10344
[16] Schierman, J. D., and Hull, J. R., "In-Flight Entry Trajectory Optimization for Reusable Launch Vehicles," AIAA Paper 2005-6434, 2005.
[17] Whang, I. H., and Whang, T. W., "Horizontal Waypoint Guidance Design Using Optimal Control," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 38, No. 3, July 2002, pp. 1116–1120. doi:10.1109/TAES.2002.1039430
[18] Yang, H., and Zhao, Y. J., "Efficient Trajectory Synthesis Through Specified Waypoints," AIAA Paper 2004-6525, 2004.
[19] Baralli, F., Pollini, L., and Innocenti, M., "Waypoint-Based Fuzzy Guidance for Unmanned Aircraft a New Approach," AIAA Paper 2002-4993, 2002.
[20] Innocenti, M., Pollini, L., and Turra, D., "Guidance of Unmanned Air Vehicles Based on Fuzzy Sets and Fixed Waypoints," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 4, 2004, pp. 715–720. doi:10.2514/1.2651
[21] Judd, K. B., and McLain, T. W., "Spline Based Path Planning for Unmanned Air Vehicles," AIAA Paper 2001-4238, 2001.
[22] Twigg, S., Calise, A., and Johnson, E., "On-Line Trajectory Optimization Including Moving Threats and Targets," AIAA Paper 2004-5139, 2004.
[23] Raghunathan, A. U., Gopal, V., Subramanian, D., Biegler, L. T., and Samad, T., "3D Conflict Resolution of Multiple Aircraft via Dynamic Optimization," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA, Reston, VA, 11–14 Aug. 2003.
[24] Raiszadeh, B., and Queen, E. M., "Partial Validation of Multibody Program to Optimize Simulated Trajectories II (POST II) Parachute Simulation with Interacting Forces," NASA Langley Research Center TR TM-2002-211634, 2002.
[25] Reddien, G. W., "Collocation at Gauss Points as a Discretization in Optimal Control," *SIAM Journal on Control and Optimization*, Vol. 17, No. 2, March 1979, pp. 298–306. doi:10.1137/0317023
[26] Michael, I., and Fahroo, F., "A Perspective on Methods for Trajectory Optimization," AIAA Paper 2002-4727, 2002.
[27] Ross, I., D'Souza, C., Fahroo, F., and Ross, J., "A Fast Approach to Multi-Stage Launch Vehicle Trajectory Optimization," AIAA Paper 2003-5639, 2003.
[28] Gong, Q., Kang, W., and Ross, I., "A Pseudospectral Method for the Optimal Control of Constrained Feedback Linearizable Systems," *IEEE Transactions on Automatic Control*, Vol. 51, No. 7, July 2006, pp. 1115–1129. doi:10.1109/TAC.2006.878570
[29] Rao, A. V., "Extension of a Pseudospectral Legendre Method to Non-Sequential Multiple-Phase Optimal Control Problems," AIAA Paper 2003-5634, 2003.
[30] Jorris, T. R., and Cobb, R. G., "2-D Trajectory Optimization Satisfying Waypoints and No-Fly Zone Constraints," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 3, May–June 2008, pp. 543–553. doi:10.2514/1.32354
[31] Jorris, T. R., "Common Aero Vehicle Autonomous Reentry Trajectory Optimization Satisfying Waypoint and No-Fly Zone Constraints," Ph.D. Thesis, Air Force Institute of Technology, Wright–Patterson Air Force Base, OH, Sept. 2007, http://handle.dtic.mil/100.2/ADA472301.
[32] Bourdreau, A. H., "Air Force Technology Roadmaps for Trans-Atmospheric Air Vehicles," AIAA Paper 99-4924, Nov. 1999.
[33] Canon, J. W., "Breathing New Hope into Hypersonics," *Aerospace America*, Nov. 2007, pp. 26–31, http://www.aiaa.org/aerospace/.
[34] Richie, G., "The Common Aero Vehicle: Space Delivery System of the Future," AIAA Paper A99-42026, 1999.
[35] Phillips, T. H., "A Common Aero Vehicle (CAV) Model, Description, and Employment Guide," Schafer Corporation for AFRL and AFSPC, TR 27 Jan. 2003.
[36] Bryson, A. E., Jr., *Dynamic Optimization*, Addison Wesley Longman, Reading, MA, 1999.
[37] Bryson, A. E., Jr., and Ho, Y.-C., *Applied Optimal Control*, Taylor and Francis, London, 1975.
[38] Hull, D. G., *Optimal Control Theory for Applications*, Springer–Verlag, Berlin, 2003.
[39] Erzberger, H., and Lee, H. Q., "Optimum Horizontal Guidance Techniques for Aircraft," *Journal of Aircraft*, Vol. 8, No. 2, Feb. 1971, pp. 95–101. doi:10.2514/3.44235
[40] Shapira, I., and Ben-Asher, J. Z., "Near-Optimal Horizontal Trajectories for Autonomous Air Vehicles," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 4, July–Aug. 1997, pp. 735–741. doi:10.2514/2.4105
[41] Vinh, N. X., Busemann, A., and Culp, R. D., *Hypersonic and Planetary Entry Flight Mechanics*, University of Michigan Press, Ann Arbor, MI, 1980.
[42] Vinh, N. X., *Optimal Trajectories in Atmospheric Flight*, Elsevier, New York, 1981.
[43] Tewari, A., *Atmospheric and Space Flight Dynamics*, Birkhäuser Boston, Cambridge, MA, 2007.
[44] Wiesel, W. E., *Spaceflight Dynamics*, McGraw–Hill, New York, 2nd ed., 1997.
[45] Shaffer, P. J., Ross, I. M., Oppenheimer, M. W., and Doman, D. B., "Optimal Trajectory Reconfiguration and Retargeting for a Reusable Launch Vehicle," *AIAA Guidance, Navigation, and Control Conference and Exhibit*, AIAA, Reston, VA, 15–18 Aug. 2005.
[46] Regan, F. J., and Anandakrishnan, S. M., *Dynamics of Atmospheric Re-Entry*, AIAA, Washington, D.C., 1993.
[47] Vinh, N. X., and Ma, D.-M., "Optimal Multiple-Pass Aeroassisted

[§]MATLAB is a registered trademark of The Mathworks, Inc., 3 Apple Hill Drive, Natick, Massachusetts.

Plane Change," *Acta Astronautica*, Vol. 21, Nos. 11–12, 1990, pp. 749–758.
doi:10.1016/0094-5765(90)90117-4

[48] Chapman, D. R., "An Approximate Analytical Method for Studying Entry Into Planetary Atmospheres," National Advisory Committee for Aeronautics TN 4276, 1958.

[49] Hankey, W. L., *Re-Entry Aerodynamics*, AIAA, Washington, D.C., Jan. 1988.

[50] Ross, I. M., and D'Souza, C. N., "Hybrid Optimal Control Framework for Mission Planning," *Journal of Guidance, Control, and Dynamics*, Vol. 28, No. 4, July–Aug. 2005, pp. 686–697.
doi:10.2514/1.8285.

[51] Fahroo, F., and Ross, I. M., "On Discrete-Time Optimality Conditions for Pseudospectral Methods," *AIAA/AAS Astro-Dynamics Specialist Conference and Exhibit*, AIAA, Reston, VA, 21–24 Aug. 2006.

[52] Ross, I. M., and Fahroo, F., "Legendre Pseudospectral Approximations of Optimal Control Problems," *Lecture Notes in Control and Information Sciences*, Vol. 295, Springer–Verlag, New York, 2003.

[53] Paris, S., and Hargraves, S. R., *Optimal Trajectories by Implicit Simulation (OTIS) Vol. II—User's Manual*, Boeing Defense and Space Group, 1996.

[54] Youssef, J., and Chowdhry, R., "Hypersonic Global Reach Trajectory Optimization," AIAA 2004-5167, 2004.

[55] Nelson, D., "Qualitative and Quantitative Assessment of Optimal Trajectories by Implicit Simulation (OTIS) and Program to Optimize Simulated Trajectories (POST)," Georgia Institute of Technology, TR AE 8900, 2001, Individual Research Project.

[56] Betts, J. T., "Survey of Numerical Methods for Trajectory Optimization," *Journal of Guidance, Control, and Dynamics*, Vol. 21, No. 2, March–April 1998, pp. 193–207.
doi:10.2514/2.4231

[57] von Stryk, O., *User's Guide for DIRCOL: A Direct Collocation Method for the Numerical Solution of Optimal Control Problems*, Technische Universitat Darmstadt, Germany, 1999 (updated April 2002).

[58] Well, K., *Graphical Environment for Simulation and Optimization*," Department of Optimization, Guidance, and Control, Stuttgart, Germany, 2002.

[59] Rutquist, P. E., and Edvall, M. M., PROPT—Matlab Optimal Control Software, TOMLAB Optimization, 15 June 2008.

[60] Huntington, G. T., Advancement and Analysis of a Gauss Pseudospectral Transcription for Optimal Control Problems, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, June 2007.

[61] von Stryk, O., "Numerical Solution of Optimal Control Problems by Direct Collocation," *Optimal Control*, Vol. 111, International Series in Numerical Mathematics, edited by R. Bulirsch, A. Miele, J. Stoer, and K. H. Well, Birkhauser, Basel, Germany, 1993, pp. 129–143.

[62] Fornberg, B., *A Practical Guide to Pseudospectral Methods*, Cambridge University Press, Cambridge, England, U.K., 1998.

[63] Trefethen, L. N., *Spectral Methods in MATLAB*, Society for Industrial and Applied Mathematics, 2000.

[64] Benson, D. A., Huntington, G. T., Thorvaldsen, T. P., and Rao, A. V., "Direct Trajectory Optimization and Costate Estimation via an Orthogonal Collocation Method," *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 6, Nov.–Dec. 2006, pp. 1435–1440.
doi:10.2514/1.20478

[65] Huntington, G. T., and Rao, A. V., "A Comparison of Global and Local Collocation Methods for Optimal Control," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 2, March–April 2008, pp. 432–436.
doi:10.2514/1.30915

[66] Huntington, G. T., Benson, D. A., How, J. P., Kanizay, N., Darby, C., and Rao, A. V., "Computation of Boundary Controls Using a Gauss Pseudospectral Method," *AAS/AIAA Astrodynamics Specialist Conference*, AIAA, Reston, VA, 19–23 Aug. 2007.

[67] Hartl, R. F., Sethi, S. P., and Vickson, R. G., "A Survey of the Maximum Principles for Optimal Control Problems with State Constraints," *SIAM Review*, Vol. 37, No. 2, June 1995, pp. 181–218.
doi:10.1137/1037043

[68] Kreim, H., Kugelmann, B., Pesch, H. J., and Breitner, M. H., "Minimizing the Maximum Heating of a Re-Entering Space Shuttle: An Optimal Control Problem with Multiple Control Constraints," *Optimal Control Applications and Methods*, Vol. 17, No. 1, 1996, pp. 45–69.
doi:10.1002/(SICI)1099-1514(199601/03)17:1<45::AID-OCA564>3.0.CO;2-X

[69] Kreyszig, E., *Advanced Engineering Mathematics*, 8th ed., Wiley, New York, 1999.

[70] Huntington, G. T., and Rao, A. V., "Optimal Reconfiguration of Tetrahedral Spacecraft Formations via a Gauss Pseudospectral Method," AAS 05-338, 2005.

[71] Huntington, G. T., Benson, D. A., and Rao, A. V., "Post-Optimality Analysis and Evaluation of a Formation Flying Problem via a Gauss Pseudospectral Method," AAS Paper 05-339, Aug. 2005.

[72] Huntington, G. T., Benson, D. A., How, J. P., Kanizay, N., Darby, C., and Rao, A. V., "Computation of Endpoint Controls Using a Gauss Pseudospectral Method," AAS Paper 07-381, 2007.

[73] Benson, D., A Gauss Pseudospectral Transcription for Optimal Control, Ph.D. Thesis, Massachusetts Institute of Technology, Cambridge, MA, Feb. 2005.

[74] Ross, I. M., Sekhavaty, P., Flemingz, A., and Gongx, Q., "Pseudospectral Feedback Control: Foundations, Examples and Experimental Results," AIAA Paper 2006-6354, 2006.

[75] Elnagar, G., and Kazemi, M., "Pseudospectral Chebyshev Optimal Control of Constrained Nonlinear Dynamical Systems," *Computational Optimization and Applications*, Vol. 11, No. 2, Nov. 1998, pp. 195–217.
doi:10.1023/A:1018694111831

[76] Elnagar, G., Kazemi, M., and Razzaghi, M., "The Pseudospectral Legendre Method for Discretizing Optimal Control Problems," *IEEE Transactions on Automatic Control*, Vol. 40, No. 10, Oct. 1995, pp. 1793–1796.

[77] Davis, P., *Interpolation and Approximation*, Dover Publications, New York, 1975.

[78] Pontryagin, L., Boltyanskii, V., Gamkrelidze, R., and Mischenko, E., *The Mathematical Theory of Optimal Processes*, Interscience, New York, 1962.

[79] Kirk, D. E., *Optimal Control Theory*, Dover Publications, New York, 1970.