Costate Estimation Using Multiple-Interval Pseudospectral Methods

Christopher L. Darby[*]
Divya Garg[†]
Anil V. Rao[‡]

University of Florida
Gainesville, FL 32611

Abstract

A method is presented for costate estimation in nonlinear optimal control problems using multiple-interval collocation at Legendre-Gauss (LG) or Legendre-Gauss-Radau (LGR) points. Transformations from the Lagrange multipliers of the nonlinear programming problem to the costate of the continuous-time optimal control problem are given. When the optimal costate is continuous, the transformed adjoint systems of the nonlinear programming problems are discrete representations of the continuous-time first-order optimality conditions. If, however, the optimal costate is discontinuous, then the transformed adjoint systems are not discrete representations of the continuous-time first-order optimality conditions. In the case where the costate is discontinuous, the accuracy of the costate approximation depends on the locations of the mesh points. In particular, the accuracy of the costate approximation is found to be significantly higher when mesh points are located at discontinuities in the costate. Two numerical examples are studied and demonstrate the effectiveness of using the multiple-interval collocation approach for estimating costate in continuous-time nonlinear optimal control problems.

Nomenclature

$\mathbf{C}$	=	Path Constraint Function
$\mathbf{D}$	=	Gauss or Radau Pseudospectral State Differentiation Matrix
$\mathbf{D}^\dagger$	=	Gauss or Radau Pseudospectral Costate Differentiation Matrix
$\mathbf{f}$	=	State Dynamics Function
g	=	Integrand of Cost Functional
H	=	Augmented Hamiltonian
J	=	Continuous-Time Cost Functional

---

[*]PhD Student. Department of Mechanical and Aerospace Engineering. E-mail: cdarby@ufl.edu.

[†]PhD Student. Department of Mechanical and Aerospace Engineering. E-mail: divyagarg2002@ufl.edu. It is noted that the first two authors on this paper share first authorship.

[‡]Assistant Professor. Department of Mechanical and Aerospace Engineering. E-mail: anilvrao@ufl.edu. Corresponding Author.

K	$=$	Number of Mesh Intervals
$L(\tau)$	$=$	Lagrange Polynomial on Time Domain $\tau \in [-1, +1]$
$\mathcal{L}$	$=$	Lagrangian of the Nonlinear Programming Problem
m	$=$	Dimension of Continuous-Time Control
N_k	$=$	Number of Collocation Points in Mesh Interval k
n	$=$	Dimension of Continuous-Time State
q	$=$	Dimension of Boundary Condition Function
s	$=$	Dimension of Path Constraint Function
t_0	$=$	Initial Time
t_f	$=$	Final Time
t	$=$	Time in Interval $[t_0, t_f]$
$\mathbf{U}_j$	$=$	Control Approximation at Time Point τ_j
$\mathbf{u}(t)$	$=$	Control on Time Domain $t \in [t_0, t_f]$
$\mathbf{u}(\tau)$	$=$	Control on Time Domain $\tau \in [-1, +1]$
w_j	$=$	j^{th} Legendre-Gauss or Legendre-Gauss-Radau Quadrature Weight
$\mathbf{X}(\tau)$	$=$	State Approximation on Time Domain $\tau \in [-1, +1]$
$\mathbf{X}_j$	$=$	State Approximation at Time Point τ_j
$\mathbf{x}(t)$	$=$	State on Time Domain $t \in [t_0, t_f]$
$\mathbf{x}(\tau)$	$=$	State on Time Domain $\tau \in [-1, +1]$
$\boldsymbol{\Gamma}$	$=$	Lagrange Multiplier Associated with Discretized Inequality Path Constraint
$\boldsymbol{\gamma}$	$=$	Lagrange Multiplier Associated with Inequality Path Constraint
δ_{ij}	$=$	Kronecker Delta Function
ϵ	$=$	Accuracy Tolerance
$\boldsymbol{\Lambda}_j$	$=$	Lagrange Multiplier of Discretized Dynamic Constraint at τ_j
$\boldsymbol{\lambda}(\tau)$	$=$	Continuous Costate on the Time Domain $\tau \in [-1, +1]$
$\boldsymbol{\lambda}_j$	$=$	Costate Approximation at Time Point τ_j
$\boldsymbol{\pi}$	$=$	Lagrange Multiplier Associated with Quadrature Constraint
$\boldsymbol{\rho}$	$=$	Jump in the Costate at Mesh Point
τ	$=$	Transformed Time Domain $[-1, +1]$
Φ	$=$	Mayer Cost
$\boldsymbol{\phi}$	$=$	Boundary Condition Function
$\boldsymbol{\Psi}$	$=$	Lagrange Multiplier Associated with the Discretized Boundary Condition
$\boldsymbol{\psi}$	$=$	Lagrange Multiplier Associated with the Boundary Condition

1 Introduction

In the past two decades, direct collocation methods have become the preferred approach for solving optimal control problems. In a direct collocation method, the state and/or the control is approximated using trial (basis) functions and the continuous problem is transcribed to a finite-dimensional nonlinear programming problem (NLP). A particular class of direct collocation methods that has become popular in the last decade is that of *pseudospectral methods.*[1–18] In a pseudospectral method, the state is approximated using a basis of Lagrange polynomials and the differential-algebraic constraint equations are enforced

at a finite set of collocation points. The three most commonly used sets of collocation points for pseudospectral methods are the *Legendre-Gauss*[10–12,14,18] (LG), *Legendre-Gauss-Radau*[13–15,17,18] (LGR), and *Legendre-Gauss-Lobatto*[1,3,4,9,19] (LGL) points. These points are the roots of the linear combinations of a Legendre polynomial and/or its derivatives and correspond to the three different types of Gaussian quadrature. All three sets of points are defined on the interval $[-1, 1]$, but differ in how the endpoints are incorporated. LG points do not include the points -1 or $+1$, the LGR points include one of the endpoints, and the LGL points include both the endpoints.

Pseudospectral methods have typically been employed in three different ways: h methods, p methods, and hp methods. In an h method, the optimal control problem is divided into many mesh intervals and a fixed low-degree polynomial is used in each mesh interval. Convergence in an h method is then achieved by refining the mesh without changing the degree of the polynomial approximation in a mesh interval. In a p method, a small number of approximating mesh intervals (often a single mesh interval) are used and convergence is achieved by increasing the degree of the polynomial approximation in each mesh interval. If the solution in each mesh interval is smooth, a p method converges at an exponential rate. Finally, an hp method is a combination of an h and a p method in that both the number of mesh intervals and the degree of the approximating polynomial in each mesh interval are allowed to change in order to achieve convergence to the solution of the optimal control problem. An advantage to using an h method over a p method is that an h method NLP is more sparse than a p method NLP. Recent research,[17] however, has demonstrated that both accuracy and computational efficiency can be gained by using an hp method over using either an h method or a p method.

Several different h, p, and hp, pseudospectral methods have been previously developed, with p methods being the most studied of the different approaches.[1–5,7–17,19] Ref. 15 shows the convergence rates for an h LGR method. Refs. 1–5, 7–14, 19 focus on p methods using LG, LGR, and LGL collocation points. Ref. 16 describes a p biased hp LG method with the key result being that a multitude of problems were accurately solved for state and control utilizing fewer collocation points and greater sparsity as compared against a single-interval LG p method. Ref. 17 describes an h biased LGR hp method. The key result of Ref. 17 is that it is advantageous to use higher-degree collocation locally when it is known that the solution within a particular mesh interval is smooth. Moreover, the approach of Ref. 17

leads to a grid refinement algorithm where the computational sparsity of an h method is retained while simultaneously improving the convergence rate in regions where the solution is smooth.

While the research of Refs. 16 and 17 has demonstrated that hp methods provide accurate approximations to the state and control, no methodology has been developed for costate estimation using a general multiple-interval formulation of a pseudospectral method. The motivation of this research is to analyze the effectiveness of hp costate estimation methods using LG and LGR collocation points. To this end, the objective of this paper is to derive a costate estimation approach for continuous-time nonlinear optimal control problems using multiple-interval LG and LGR pseudospectral methods, and to assess the accuracy of the approach. In a manner similar to that of Refs. 16 and 17, the continuous-time optimal control problem is transcribed to an NLP using multiple-interval versions of the Gauss and Radau (that is, LG and LGR) pseudospectral methods.[10–14] The Karush-Kuhn-Tucker (KKT) conditions of the LG and LGR NLPs are then derived. Transformations of the NLP Lagrange multipliers to the costate of the continuous-time optimal control problem are then derived, leading to multiple-interval LG and LGR transformed adjoint systems.[20] Because the multiple-interval methods derived in this paper require continuity in the state at the mesh points, the formulation is significantly different from previously developed p method costate estimation procedures such as those found in Refs. 4, 11, 13, and 14. Specifically, for a continuous optimal costate it is found that the LG and LGR transformed adjoint systems are discrete representations of the continuous-time first-order optimality conditions. When the optimal costate is discontinuous, however, the transformed adjoint systems are *not* discrete representations of the continuous-time first-order optimality conditions if a mesh point is at the location of discontinuity in the costate. In this latter case, it is found by example that the costate estimate is significantly more accurate when mesh points are located at the discontinuities than when mesh points are not placed at the discontinuities. Two examples are studied to analyze the accuracy of the costate estimation methods derived in this paper. In particular, the second example is a common type of problem that arises in aerospace engineering of minimum-energy control subject to a state inequality path constraint.

This paper is organized as follows: In Section 2 we define our notation and conventions. In Section 3 we state both the optimal control problem being approximated by our hp approach and the corresponding first-order optimality conditions. In Section 4 we define the nonlinear

programming problem (NLP) that arises from the *hp* LG and *hp* LGR collocation methods used to discretize the continuous Bolza problem defined in Section 3. In addition, we derive the first-order optimality conditions of these two NLPs and provide a transformation of the NLP Lagrange multipliers to the costate of the continuous-time optimal control problem. In Section 5 we study the accuracy of the costate estimation procedure on two examples. The first example has a smooth solution while the second example has a discontinuous optimal costate. Finally, in Sections 6 and 7 we give a discussion of results and conclusions, respectively.

2   Conventions and Notation

For each method derived in this paper, $(\tau_1, \ldots, \tau_{N_k}) \in [-1, +1]$ denote the LG or LGR quadrature (collocation) points in interval k.[21] It should be noted that the LG quadrature points lie on the interval $(-1, +1)$ while the LGR quadrature points lie on the interval $[-1, +1)$. In order to properly account for the boundary conditions, the LG method includes the noncollocated points $\tau_0 = -1$ and $\tau_{N_k+1} = +1$ while the LGR method includes the noncollocated point $\tau_{N_k+1} = +1$. The manner in which these noncollocated points are used has been described in great detail in Refs. 13, 14, 18. For either method, the state in mesh interval k is approximated using a basis of Lagrange polynomials

$$L_i(\tau) = \prod_{\substack{l=Z \\ l \neq i}}^{N_k+Z} \frac{\tau - \tau_l^{(k)}}{\tau_i - \tau_l^{(k)}}, \quad Z \leq i \leq N_k, \tag{1}$$

where $Z = 0$ for the LG method and $Z = 1$ for the LGR method. Next, all vector functions of time are denoted as *row* vectors; that is,

$$\mathbf{y}(\tau) = [y_1(\tau), \cdots, y_n(\tau)] \in \mathbb{R}^n.$$

We define the approximation of the state, control, and the NLP Lagrange multipliers corresponding to the discrete approximation to the dynamics at $\tau = \tau_i$ as $\mathbf{X}_i$, $\mathbf{U}_i$, and $\mathbf{\Lambda}_i$ respectively. Moreover, in the LG method, $\mathbf{X}^{(k)}$ is an $(N_k + 1) \times n$ matrix corresponding to the state at the initial point plus the LG points in mesh interval k, while in the LGR method, $\mathbf{X}^{(k)}$ is an $(N_k + 1) \times n$ matrix corresponding to the state at the LGR points and the final point in mesh interval k. Similarly, $\mathbf{U}^{(k)}$ and $\mathbf{\Lambda}^{(k)}$ are $N_k \times n$ matrices that correspond to the control and NLP Lagrange multipliers, respectively, at either the LG or LGR points in

mesh interval k. Finally, the notation $\mathbf{A}_{i:j}$ attached to any matrix $\mathbf{A}$ denotes *rows i through j* of the matrix $\mathbf{A}$ and $\mathbf{A}_i$ denotes *row i* of the matrix $\mathbf{A}$.

In addition to the above vector and matrix conventions, the operation $\langle \mathbf{a}, \mathbf{b} \rangle$ is used to denote the standard inner product between the vectors $\mathbf{a} \in \mathbb{R}^n$ and $\mathbf{b} \in \mathbb{R}^n$. Furthermore, if $\mathbf{f} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$, then $\nabla \mathbf{f}$ is the m by n Jacobian matrix whose i^{th} row is $\nabla \mathbf{f}_i$. Using this convention, the gradient of a scalar-valued function is a *row* vector. If $\phi : \mathbb{R}^{m \times n} \longrightarrow \mathbb{R}$ and $\mathbf{X}$ is an m by n matrix, then $\nabla \phi$ denotes the m by n matrix whose (i, j) element is $(\nabla \phi(\mathbf{X}))_{ij} = \partial \phi(\mathbf{X}) / \partial X_{ij}$. Finally, the Kronecker delta function is defined by $\delta_{ii} = 1$ and $\delta_{ij} = 0$ if $i \neq j$.

3   Bolza Optimal Control Problem

Without loss of generality, consider the following fixed-time nonlinear optimal control problem in Bolza form. Minimize the cost functional

$$J = \Phi(\mathbf{x}(t_0), \mathbf{x}(t_f)) + \int_{t_0}^{t_f} g(\mathbf{x}(t), \mathbf{u}(t)) \, dt, \tag{2}$$

subject to the dynamic constraints

$$\frac{d\mathbf{x}}{dt} = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \tag{3}$$

the inequality path constraints

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

and the boundary conditions (i.e., the event constraints)

$$\boldsymbol{\phi}(\mathbf{x}(t_0), \mathbf{x}(t_f)) = \mathbf{0}, \tag{5}$$

where $\mathbf{x}(t) \in \mathbb{R}^n$, $\mathbf{u}(t) \in \mathbb{R}^m$, $\mathbf{C} : \mathbb{R}^n \times \mathbb{R}^m \longrightarrow \mathbb{R}^s$, $\boldsymbol{\phi} : \mathbb{R}^n \times \mathbb{R}^n \longrightarrow \mathbb{R}^q$, and t_0 and t_f are fixed. Furthermore, we consider only problems where the functions g, $\mathbf{f}$ and $\mathbf{C}$ are not explicit functions of time, and the functions Φ, and $\boldsymbol{\phi}$ are not explicit functions of t_0 and t_f. It is noted, however, that the results shown in this paper are easily extendable to problems with a varying initial and/or final time and/or problems where time appears explicitly in Eqs. (2)–(5).

Suppose now that the continuous Bolza problem of Eqs. (2)-(5) is divided into K *mesh intervals* $[t_{k-1}, t_k]$, $k = 1, \ldots, K$, where $t_0 < t_1 < t_2 < \cdots < t_K = t_f$ are the *mesh points*. In

each mesh interval $t \in [t_{k-1}, t_k]$ is transformed to $\tau \in [-1, +1]$ via the affine transformation

$$\tau = \frac{2t - (t_k + t_{k-1})}{t_k - t_{k-1}}, \tag{6}$$

from which we obtain

$$\frac{d\tau}{dt} = \frac{2}{t_k - t_{k-1}}, \quad 1 \le k \le K. \tag{7}$$

Next, let $\mathbf{x}^{(k)}(\tau)$ and $\mathbf{u}^{(k)}(\tau)$ be the state and control, respectively, in mesh interval k. Using Eq. (7), the Bolza optimal control problem in Eqs. (2)–(5) can be written as follows. First, the cost functional of Eq. (2) can be written as

$$J = \Phi(\mathbf{x}^{(1)}(-1), \mathbf{x}^{(K)}(+1)) + \sum_{k=1}^{K} \frac{t_k - t_{k-1}}{2} \int_{-1}^{+1} g(\mathbf{x}^{(k)}(\tau), \mathbf{u}^{(k)}(\tau)) \, d\tau. \tag{8}$$

Next, the dynamics of Eq. (3) and the path constraints of (4) are given in terms of τ in mesh interval $k \in [1, \ldots, K]$, respectively, as

$$\frac{d\mathbf{x}^{(k)}(\tau)}{d\tau} \equiv \dot{\mathbf{x}}^{(k)}(\tau) = \frac{t_k - t_{k-1}}{2} \mathbf{f}(\mathbf{x}^{(k)}(\tau), \mathbf{u}^{(k)}(\tau)), \tag{9}$$

$$\mathbf{C}(\mathbf{x}^{(k)}(\tau), \mathbf{u}^{(k)}(\tau)) \le \mathbf{0}. \tag{10}$$

Furthermore, the boundary conditions of Eq. (5) are given as

$$\phi(\mathbf{x}^{(1)}(-1), \mathbf{x}^{(K)}(+1)) = \mathbf{0}. \tag{11}$$

Finally, the state is assumed to be continuous at an interior mesh point, that is,

$$\mathbf{x}^{(k)}(+1) = \mathbf{x}^{(k+1)}(-1), \quad 1 \le k \le K - 1. \tag{12}$$

3.1   First-Order Optimality Conditions of Bolza Problem

In this section, we develop the first-order optimality conditions from the calculus of variations for the multiple-interval continuous-time Bolza optimal control problem as given in Eqs. (8)–(11). First, the augmented Hamiltonian in interval k is defined as

$$H^{(k)}(\mathbf{x}^{(k)}, \boldsymbol{\lambda}^{(k)}, \mathbf{u}^{(k)}, \boldsymbol{\gamma}^{(k)}) = g^{(k)} + \langle \boldsymbol{\lambda}^{(k)}, \mathbf{f}^{(k)} \rangle - \langle \boldsymbol{\gamma}^{(k)}, \mathbf{C}^{(k)} \rangle, \tag{13}$$

where $\boldsymbol{\lambda}^{(k)}(\tau) \in \mathbb{R}^n$ and $\boldsymbol{\gamma}^{(k)}(\tau) \in \mathbb{R}^s$. Applying the continuous-time first-order optimality conditions from the calculus of variations,[22, 23] we obtain

$$\nabla_{\mathbf{u}} H^{(k)} = \nabla_{\mathbf{u}} \left(g^{(k)} + \langle \boldsymbol{\lambda}^{(k)}, \mathbf{f}^{(k)} \rangle - \langle \boldsymbol{\gamma}^{(k)}, \mathbf{C}^{(k)} \rangle \right) = \mathbf{0}, \tag{14}$$

$$-\frac{t_k - t_{k-1}}{2} \nabla_{\mathbf{x}} H^{(k)} = -\frac{t_k - t_{k-1}}{2} \nabla_{\mathbf{x}} \left(g^{(k)} + \langle \boldsymbol{\lambda}^{(k)}, \mathbf{f}^{(k)} \rangle - \langle \boldsymbol{\gamma}^{(k)}, \mathbf{C}^{(k)} \rangle \right) = \dot{\boldsymbol{\lambda}}^{(k)}(\tau), \tag{15}$$

$$\boldsymbol{\lambda}^{(1)}(-1) = -\nabla_{\mathbf{x}^{(1)}(-1)} \left(\boldsymbol{\Phi} - \langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle \right), \tag{16}$$

$$\boldsymbol{\lambda}^{(K)}(1) = \nabla_{\mathbf{x}^{(K)}(+1)} \left(\boldsymbol{\Phi} - \langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle \right), \tag{17}$$

$$\boldsymbol{\lambda}^{(k)}(1) = \boldsymbol{\lambda}^{(k+1)}(-1) + \boldsymbol{\rho}^{(k)}, \quad 1 \le k \le K - 1, \tag{18}$$

where $\boldsymbol{\psi} \in \mathbb{R}^q$ and $\boldsymbol{\rho}^{(k)} \in \mathbb{R}^n$ is the jump (discontinuity) in the costate at the mesh points due to active path constraints and interior point constraints. Then, from the complementary slackness condition, the Lagrange multiplier $\boldsymbol{\gamma}^{(k)}$ takes the values

$$\gamma_j^{(k)}(\tau) = 0 \quad \text{when} \quad C_j^{(k)}(\tau) < 0, \quad 1 \le j \le s, \tag{19}$$

$$\gamma_j^{(k)}(\tau) < 0 \quad \text{when} \quad C_j^{(k)}(\tau) = 0, \quad 1 \le j \le s. \tag{20}$$

4 LG & LGR Discretizations of Bolza Problem

In this section we consider the discretization of the continuous Bolza optimal control problem of Section 3 using multiple-interval versions of the previously developed *Gauss* (LG) and *Radau* (LGR) pseudospectral methods.[10–14] First, the optimal control problem is transcribed to a finite-dimensional NLP for each discretization method. Next, the Karush-Kuhn-Tucker (KKT) conditions of the LG and LGR NLPs, along with the corresponding transformed adjoint systems,[20] are derived for the *hp* LG and *hp* LGR methods. We show that if the costate is continuous, the transformed adjoint systems for both the LG and LGR *hp* methods are discrete approximations of the continuous-time first-order optimality conditions. If, however, the costate is discontinuous, we show that the LG and LGR transformed adjoint systems are not discrete approximations of the continuous-time first-order optimality conditions.

4.1 Legendre-Gauss Discretization

In the Gauss (LG) pseudospectral method, the state is approximated in each mesh interval k as

$$\mathbf{x}^{(k)}(\tau) \approx \mathbf{X}^{(k)}(\tau) = \sum_{j=0}^{N_k} \mathbf{X}_j^{(k)} L_j^{(k)}(\tau). \tag{21}$$

Differentiating $\mathbf{X}^{(k)}(\tau)$ in Eq. (21) with respect to τ, we obtain

$$\frac{d\mathbf{X}^{(k)}(\tau)}{d\tau} \equiv \dot{\mathbf{X}}^{(k)}(\tau) = \sum_{j=0}^{N_k} \mathbf{X}_j^{(k)} \dot{L}_j^{(k)}(\tau). \tag{22}$$

Collocating the right-hand side of the state dynamics of Eq. (9) with the derivative of the state approximation in Eq. (22) at the N_k LG points, we have

$$\sum_{j=0}^{N_k} \mathbf{X}_j^{(k)} D_{ij}^{(k)} - \frac{t_k - t_{k-1}}{2} \mathbf{f}_i^{(k)} = \mathbf{0}, \quad 1 \leq i \leq N_k \tag{23}$$

where $\mathbf{f}_i^{(k)} = \mathbf{f}(\mathbf{X}_i^{(k)}, \mathbf{U}_i^{(k)})$ and

$$D_{ij}^{(k)} = \dot{L}_j^{(k)}(\tau_i^{(k)}), \quad 1 \leq i \leq N_k, \quad 0 \leq j \leq N_k \tag{24}$$

is the $N_k \times (N_k+1)$ *Gauss pseudospectral differentiation matrix*[7, 10–12] in the kth mesh interval. Next, the cost functional of Eq. (8) is approximated using a multiple-interval LG quadrature as

$$J \approx \Phi(\mathbf{X}_0^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) + \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{t_k - t_{k-1}}{2} w_i^{(k)} g_i^{(k)}, \tag{25}$$

where $g_i^{(k)} = g(\mathbf{X}_i^{(k)}, \mathbf{U}_i^{(k)})$. The path constraints of Eq. (10) are enforced at the N_k LG points in mesh interval k as

$$\mathbf{C}_i \equiv \mathbf{C}(\mathbf{X}_i^{(k)}, \mathbf{U}_i^{(k)}) \leq \mathbf{0}, \quad 1 \leq i \leq N_k. \tag{26}$$

Finally, the boundary conditions are approximated as

$$\phi(\mathbf{X}_0^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) = \mathbf{0}. \tag{27}$$

In addition, an NLP variable corresponding to the state at the terminal point of each mesh interval k, $\mathbf{X}^{(k)}(+1) = \mathbf{X}_{N_k+1}^{(k)}$, is included by adding the following Legendre-Gauss quadrature approximation to the the state at $\tau = +1$:

$$\mathbf{X}_{N_k+1}^{(k)} = \mathbf{X}_0^{(k)} + \frac{t_k - t_{k-1}}{2} \sum_{i=1}^{N_k} w_i^{(k)} \mathbf{f}_i^{(k)}, \quad 1 \leq k \leq K. \tag{28}$$

It is noted that state continuity at the mesh points $k \in [1, \ldots, K-1]$ is enforced via the constraint

$$\mathbf{X}_{N_k+1}^{(k)} = \mathbf{X}_0^{(k+1)}. \tag{29}$$

When implementing the Gauss pseudospectral method, a single variable is used for the value of the state at the end of mesh interval k and the start of mesh interval $k + 1$, that is,

9

$\mathbf{X}_{N_k+1}^{(k)} \equiv \mathbf{X}_0^{(k+1)}$, $1 \leq k \leq K-1$. Hence, redundant variables defining the state at the interior mesh points are eliminated in the discretization and Eq. (28) can be written as

$$\mathbf{X}_0^{(k+1)} = \mathbf{X}_0^{(k)} + \frac{t_k - t_{k-1}}{2} \sum_{i=1}^{N_k} w_i^{(k)} \mathbf{f}_i^{(k)}, \quad 1 \leq k \leq K-1 \tag{30}$$

$$\mathbf{X}_{N_K+1}^{(K)} = \mathbf{X}_0^{(K)} + \frac{t_K - t_{K-1}}{2} \sum_{i=1}^{N_K} w_i^{(K)} \mathbf{f}_i^{(K)}.$$

The NLP that arises from the Gauss pseudospectral approximation is to minimize the cost function of Eq. (25) subject to the algebraic constraints of Eqs. (23), (26), (27) and (30).

4.2   KKT Conditions for Legendre-Gauss Discretization

The Karush-Kuhn-Tucker (KKT) conditions of the NLP of Section 4.1 are now derived. The Lagrangian of the NLP is given as

$$\begin{aligned}
\mathcal{L} = {} & \Phi(\mathbf{X}_0^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) - \langle \mathbf{\Psi}, \boldsymbol{\phi}(\mathbf{X}_0^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) \rangle \\
& + \sum_{k=1}^{K} \sum_{j=1}^{N_k} \left(\frac{t_k - t_{k-1}}{2} w_j^{(k)} g_j^{(k)} - \langle \mathbf{\Gamma}_j^{(k)}, \mathbf{C}_j^{(k)} \rangle \right. \\
& \left. - \langle \mathbf{\Lambda}_j^{(k)}, \mathbf{D}_{j,1:N_k}^{(k)} \mathbf{X}_{1:N_k}^{(k)} + D_{j,0}^{(k)} \mathbf{X}_0^{(k)} - \frac{t_k - t_{k-1}}{2} \mathbf{f}_j^{(k)} \rangle \right) \\
& - \sum_{k=1}^{K-1} \langle \boldsymbol{\pi}^{(k)}, \mathbf{X}_0^{(k+1)} - \mathbf{X}_0^{(k)} - \frac{t_k - t_{k-1}}{2} \sum_{j=1}^{N_k} w_j^{(k)} \mathbf{f}_j^{(k)} \rangle \\
& - \langle \boldsymbol{\pi}^{(K)}, \mathbf{X}_{N_k+1}^{(K)} - \mathbf{X}_0^{(K)} - \frac{t_K - t_{K-1}}{2} \sum_{j=1}^{N_K} w_j^{(K)} \mathbf{f}_j^{(K)} \rangle,
\end{aligned} \tag{31}$$

where $\mathbf{\Lambda}^{(k)}$, $\mathbf{\Gamma}^{(k)}$, $\mathbf{\Psi}$, and $\boldsymbol{\pi}^{(k)}$ are the Lagrange multipliers associated, respectively, with the discretized dynamic constraints of Eq. (23), the discretized path constraints of Eq. (26), the discretized event constraints of Eq. (27), and the quadrature constraints of Eq. (30) in mesh interval k. The KKT conditions of the NLP are obtained by differentiating $\mathcal{L}$ with respect to $\mathbf{X}_j^{(k)}$, $0 \leq j \leq N_k + 1$, and $\mathbf{U}_j^{(k)}$, $1 \leq j \leq N_k$, in each interval k and setting these derivatives

equal to zero. The KKT conditions are given as

$$
\mathbf{0} \;=\; \nabla_{\mathbf{U}} \left(w_j^{(k)} g_j^{(k)} + \langle \mathbf{\Lambda}_j^{(k)} + w_j^{(k)} \boldsymbol{\pi}^{(k)}, \mathbf{f}_j^{(k)} \rangle \right. \tag{32}
$$

$$
\left. - \frac{2}{t_k - t_{k-1}} \langle \mathbf{\Gamma}_j^{(k)}, \mathbf{C}_j^{(k)} \rangle \right), \quad 1 \le k \le K, \quad 1 \le j \le N_k
$$

$$
\mathbf{D}_j^{(k)^\mathsf{T}} \mathbf{\Lambda}^{(k)} \;=\; \frac{t_k - t_{k-1}}{2} \nabla_{\mathbf{X}} \left(w_j^{(k)} g_j^{(k)} + \langle \mathbf{\Lambda}_j^{(k)} + w_j^{(k)} \boldsymbol{\pi}^{(k)}, \mathbf{f}_j^{(k)} \rangle \right. \tag{33}
$$

$$
\left. - \frac{2}{t_k - t_{k-1}} \langle \mathbf{\Gamma}_j^{(k)}, \mathbf{C}_j^{(k)} \rangle \right), \quad 1 \le k \le K, \quad 1 \le j \le N_k
$$

$$
\mathbf{D}_0^{(1)^\mathsf{T}} \mathbf{\Lambda}^{(1)} - \boldsymbol{\pi}^{(1)} \;=\; \nabla_{\mathbf{X}_0^{(1)}} (\Phi - \langle \mathbf{\Psi}, \boldsymbol{\phi} \rangle), \tag{34}
$$

$$
\mathbf{D}_0^{(k)^\mathsf{T}} \mathbf{\Lambda}^{(k)} - \boldsymbol{\pi}^{(k)} \;=\; -\boldsymbol{\pi}^{(k-1)}, \quad 2 \le k \le K \tag{35}
$$

$$
\boldsymbol{\pi}^{(K)} \;=\; \nabla_{\mathbf{X}_{N_K+1}^{(K)}} (\Phi - \langle \mathbf{\Psi}, \boldsymbol{\phi} \rangle), \tag{36}
$$

where $\mathbf{D}_j^\mathsf{T}$ is the jth row of $\mathbf{D}^\mathsf{T}$. Next, consider the following change of variables from continuous to discrete time:

$$
\boldsymbol{\psi} \;=\; \mathbf{\Psi}, \tag{37}
$$

$$
\boldsymbol{\gamma}_j^{(k)} \;=\; \frac{\mathbf{\Gamma}_j^{(k)}}{w_j^{(k)}} \frac{2}{t_k - t_{k-1}}, \quad 1 \le k \le K, \quad 1 \le j \le N_k \tag{38}
$$

$$
\boldsymbol{\lambda}_j^{(k)} \;=\; \frac{\mathbf{\Lambda}^{(k)}}{w_j^{(k)}} + \boldsymbol{\pi}^{(k)}, \quad 1 \le k \le K, \quad 1 \le j \le N_k. \tag{39}
$$

Furthermore, define the matrix $\mathbf{D}^{\dagger(k)}$ as

$$
D_{ij}^{\dagger(k)} \;=\; -\frac{w_j^{(k)}}{w_i^{(k)}} D_{ji}^{(k)}, \quad 1 \le k \le K, \quad 1 \le i, j \le N_k \tag{40}
$$

$$
D_{i,N_k+1}^{\dagger(k)} \;=\; -\sum_{j=1}^{N_k} D_{ij}^{\dagger(k)}, \quad 1 \le i \le N_k. \tag{41}
$$

From Eq. (36), the costate approximation at $t = t_f$ is obtained as

$$
\boldsymbol{\lambda}_{N_K+1}^{(K)} = \boldsymbol{\pi}^{(K)}. \tag{42}
$$

Next, let

$$
\boldsymbol{\lambda}_0^{(1)} \;=\; -\mathbf{D}_0^{(1)^\mathsf{T}} \mathbf{\Lambda}^{(1)} + \boldsymbol{\pi}^{(1)} \tag{43}
$$

11

be the costate approximation at $t = t_0$. Substituting the change of variables given in Eqs. (37) (41), (42), and (43) into (32) (36), the LG transformed adjoint system is given as

$$0 = \nabla_{\mathbf{U}} H(\mathbf{X}_j^{(k)}, \mathbf{U}_j^{(k)}, \boldsymbol{\gamma}_j^{(k)}, \boldsymbol{\lambda}_j^{(k)}), \tag{44}$$

$$1 \le k \le K, \quad 1 \le j \le N_k$$

$$(\mathbf{D}_{j,1:N_k}^{\dagger(k)} \boldsymbol{\lambda}^{(k)} + D_{j,N_k+1}^{\dagger(k)} \boldsymbol{\pi}^{(k)}) = -\frac{t_k - t_{k-1}}{2} \nabla_{\mathbf{X}} H(\mathbf{X}_j^{(k)}, \mathbf{U}_j^{(k)}, \boldsymbol{\gamma}_j^{(k)}, \boldsymbol{\lambda}_j^{(k)}), \tag{45}$$

$$1 \le k \le K, \quad 1 \le j \le N_k$$

$$\boldsymbol{\lambda}_0^{(1)} = -\nabla_{\mathbf{X}_0^{(1)}} (\Phi - \langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle), \tag{46}$$

$$\boldsymbol{\lambda}_{N_K+1}^{(K)} = \nabla_{\mathbf{X}_{N_K+1}^{(K)}} (\Phi - \langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle), \tag{47}$$

$$\boldsymbol{\pi}^{(k-1)} = \boldsymbol{\pi}^{(k)} + \frac{t_k - t_{k-1}}{2} \sum_{j=1}^{N_k} w_j^{(k)} \nabla_{\mathbf{X}} H(\mathbf{X}_j^{(k)}, \mathbf{U}_j^{(k)}, \boldsymbol{\gamma}_j^{(k)}, \boldsymbol{\lambda}_j^{(k)}), \tag{48}$$

$$2 \le k \le K,$$

where it is noted that Eq. (48) is formed by comparing Eqs. (33) and (35) together with the identity[14]

$$D_{i,0}^{(k)} = -\sum_{j=1}^{N_k} D_{ij}^{(k)}, \quad 1 \le i \le N_k. \tag{49}$$

Next, setting $k = K$ in Eq. (48) and noting that $\boldsymbol{\lambda}_{N_K+1}^{(K)} = \boldsymbol{\pi}^{(K)}$, we obtain

$$\boldsymbol{\lambda}_0^{(K)} = \boldsymbol{\pi}^{(K-1)}. \tag{50}$$

Suppose now that the costate is continuous at the mesh points $k = 1, \ldots, K - 1$. Then the right-hand side of Eq. (48) is an approximation for the costate at the beginning of mesh intervals $k = 1, \ldots, K - 1$, that is

$$\boldsymbol{\lambda}_0^{(k+1)} = \boldsymbol{\lambda}_{N_k+1}^{(k)} = \boldsymbol{\pi}^{(k)}, \quad (k = 1, \ldots, K - 1). \tag{51}$$

It is noted that the transformed adjoint system is a discrete representation of the continuous-time first-order necessary conditions if the costate is continuous at every mesh point.

Suppose now that the costate is discontinuous at a particular mesh point $M \in [1, \ldots, K - 1]$. Then

$$\boldsymbol{\pi}^{(M)} = \boldsymbol{\lambda}_0^{(M+1)} = \boldsymbol{\lambda}_{N_M+1}^{(M)} - \boldsymbol{\rho}^{(M)} \tag{52}$$

where $\boldsymbol{\rho}^{(M)}$ is the difference between $\boldsymbol{\lambda}_{N_M+1}^{(M)}$ and $\boldsymbol{\lambda}_0^{(M+1)}$. Therefore, if the costate is discontinuous at mesh point M, $\boldsymbol{\pi}^{(M)} \ne \boldsymbol{\lambda}_{N_M+1}^{(M)}$ in Eq. (45). Hence, the left-hand side of Eq. (45)

is not a discrete approximation to the continuous-time costate dynamics. As a result, the transformed adjoint system given in Eqs. (44)–(48) is not a discrete representation of the first-order optimality conditions given in Eqs. (14)–(18).

4.3 Legendre-Gauss-Radau Discretization

In the Radau (LGR) pseudospectral method, the state is approximated in each mesh interval k as

$$\mathbf{x}^{(k)}(\tau) \approx \mathbf{X}^{(k)}(\tau) = \sum_{j=1}^{N_k+1} \mathbf{X}_j^{(k)} L_j^{(k)}(\tau). \tag{53}$$

Differentiating $\mathbf{X}^{(k)}(\tau)$ in Eq. (53) with respect to τ, we obtain

$$\frac{d\mathbf{X}^{(k)}(\tau)}{d\tau} \equiv \dot{\mathbf{X}}^{(k)}(\tau) = \sum_{j=1}^{N_k+1} \mathbf{X}_j^{(k)} \dot{L}_j^{(k)}(\tau). \tag{54}$$

Collocating the state dynamics of Eq. (9) with the derivative of the state approximation in Eq. (54) at the N_k LGR points, we have

$$\sum_{j=1}^{N_k+1} \mathbf{X}_j^{(k)} D_{ij}^{(k)} - \frac{t_k - t_{k-1}}{2} \mathbf{f}_i^{(k)} = \mathbf{0}, \quad 1 \le i \le N_k, \tag{55}$$

where

$$D_{ij}^{(k)} = \dot{L}_j^{(k)}(\tau_i^{(k)}), \qquad 1 \le i \le N_k, \quad 1 \le j \le N_k + 1 \tag{56}$$

is the $N_k \times (N_k + 1)$ *Radau pseudospectral differentiation matrix*[13] in mesh interval k. The cost functional of Eq. (8) is then approximated using a multiple-interval LGR quadrature as

$$J \approx \Phi(\mathbf{X}_1^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) + \sum_{k=1}^{K} \sum_{i=1}^{N_k} \frac{t_k - t_{k-1}}{2} w_i^{(k)} g_i^{(k)}. \tag{57}$$

Furthermore, the path constraints of Eq. (4) are enforced at the N_k LGR points in mesh interval k as

$$\mathbf{C}_i^{(k)} \equiv \mathbf{C}(\mathbf{X}_i^{(k)}, \mathbf{U}_i^{(k)}) \le \mathbf{0}, \quad 1 \le i \le N_k. \tag{58}$$

Finally, the boundary conditions are approximated as

$$\phi(\mathbf{X}_1^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) = \mathbf{0}. \tag{59}$$

Continuity in the state at the mesh points $k \in [1, \ldots, K - 1]$ is imposed via the equality constraint

$$\mathbf{X}_{N_k+1}^{(k)} = \mathbf{X}_1^{(k+1)}. \tag{60}$$

13

The NLP that arises from the Radau pseudospectral approximation is to minimize the cost function of Eq. (57) subject to the algebraic constraints of Eqs. (55), (58), (59) and (60). As with the multiple-interval form of the Gauss pseudospectral method, the same NLP variable is used for both $\mathbf{X}_{N_k+1}^{(k)}$ and $\mathbf{X}_1^{(k+1)}$, thus eliminating the need for Eq. (60) and removing any redundant variables in the discretization. The constraints of Eq. (55) are then written as

$$
\begin{aligned}
\mathbf{D}_{j,1:N_k}^{(k)}\mathbf{X}_{1:N_k}^{(k)} + D_{j,N_k+1}^{(k)}\mathbf{X}_1^{(k+1)} - \frac{t_k - t_{k-1}}{2}\mathbf{f}_j^{(k)} &= \mathbf{0} \quad 1 \le k \le K-1, \\
\mathbf{D}_{j,1:N_K}^{(K)}\mathbf{X}_{1:N_K}^{(K)} + D_{j,N_K+1}^{(K)}\mathbf{X}_{N_K+1}^{(K)} - \frac{t_K - t_{K-1}}{2}\mathbf{f}_j^{(K)} &= \mathbf{0}
\end{aligned}
\tag{61}
$$

4.4 KKT Conditions for Legendre-Gauss-Radau Discretization

The KKT conditions of the NLP of Section 4.3 are now derived. The Lagrangian of the LGR NLP is given as

$$
\begin{aligned}
\mathcal{L} = {} & \Phi(\mathbf{X}_1^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) - \langle \mathbf{\Psi}, \boldsymbol{\phi}(\mathbf{X}_1^{(1)}, \mathbf{X}_{N_K+1}^{(K)}) \rangle \\
& + \sum_{k=1}^{K}\sum_{j=1}^{N_k} \frac{t_k - t_{k-1}}{2}w_j^{(k)}g_j^{(k)} - \langle \mathbf{\Gamma}_j^{(k)}, \mathbf{C}_j^{(k)} \rangle \\
& - \sum_{k=1}^{K-1}\sum_{j=1}^{N_k} \langle \mathbf{\Lambda}_j^{(k)}, \mathbf{D}_{j,1:N_k}^{(k)}\mathbf{X}_{1:N_k}^{(k)} + D_{j,N_k+1}^{(k)}\mathbf{X}_1^{(k+1)} - \frac{t_k - t_{k-1}}{2}\mathbf{f}_j^{(k)} \rangle \\
& - \sum_{j=1}^{N_K} \langle \mathbf{\Lambda}_j^{(K)}, \mathbf{D}_{j,1:N_K}^{(K)}\mathbf{X}_{1:N_K}^{(K)} + D_{j,N_K+1}^{(K)}\mathbf{X}_{N_K+1}^{(K)} - \frac{t_K - t_{K-1}}{2}\mathbf{f}_j^{(K)} \rangle
\end{aligned}
\tag{62}
$$

where $\mathbf{\Lambda}^{(k)}$, $\mathbf{\Gamma}^{(K)}$, $\mathbf{\Psi}$, and are the Lagrange multipliers associated, respectively, with the discretized dynamic constraints of Eq. (55), the discretized path constraints of Eq. (58), and the discretized event constraints of Eq. (59) in mesh interval k, respectively. The KKT

14

conditions of the multiple-interval Legendre-Gauss-Radau discretization are then given as

$$
\mathbf{0} = \nabla_{\mathbf{U}} \left(w_j^{(k)} g_j^{(k)} + \langle \mathbf{\Lambda}_j^{(k)}, \mathbf{f}_j^{(k)} \rangle \right. \tag{63}
$$

$$
\left. - \frac{2}{t_k - t_{k-1}} \langle \mathbf{\Gamma}_j^{(k)}, \mathbf{C}_j^{(k)} \rangle \right), \quad 1 \le k \le K, \quad 1 \le j \le N_k
$$

$$
\mathbf{D}_j^{(1)^{\mathsf{T}}} \mathbf{\Lambda}^{(1)} = \frac{t_1 - t_0}{2} \nabla_{\mathbf{X}} \left(w_j^{(1)} g_j^{(1)} + \langle \mathbf{\Lambda}_j^{(1)}, \mathbf{f}_j^{(1)} \rangle \right. \tag{64}
$$

$$
\left. - \frac{2}{t_1 - t_0} \langle \mathbf{\Gamma}_j^{(1)}, \mathbf{C}_j^{(1)} \rangle \right) - \delta_{1j} \left(-\nabla_{\mathbf{X}_1^{(1)}} \Phi + \nabla_{\mathbf{X}_1^{(1)}} \langle \mathbf{\Psi}, \boldsymbol{\phi} \rangle \right), \quad 1 \le j \le N_1
$$

$$
\mathbf{D}_j^{(k)^{\mathsf{T}}} \mathbf{\Lambda}^{(k)} = \frac{t_k - t_{k-1}}{2} \nabla_{\mathbf{X}} \left(w_j^{(k)} g_j^{(k)} + \langle \mathbf{\Lambda}_j^{(k)}, \mathbf{f}_j^{(k)} \rangle \right. \tag{65}
$$

$$
\left. - \frac{2}{t_k - t_{k-1}} \langle \mathbf{\Gamma}_j^{(k)}, \mathbf{C}_j^{(k)} \rangle \right) - \delta_{1j} \mathbf{D}_{N_{k-1}+1}^{(k-1)^{\mathsf{T}}} \mathbf{\Lambda}^{(k-1)}, \quad 2 \le k \le K, \quad 1 \le j \le N_k
$$

$$
\mathbf{D}_{N_K+1}^{(K)^{\mathsf{T}}} \mathbf{\Lambda}^{(K)} = \nabla_{\mathbf{X}_{N_K+1}^{(K)}} (\Phi - \langle \mathbf{\Psi}, \boldsymbol{\phi} \rangle). \tag{66}
$$

Next, consider the following change of variables from continuous to discrete time:

$$
\boldsymbol{\psi} = \mathbf{\Psi}, \tag{67}
$$

$$
\boldsymbol{\gamma}_j^{(k)} = \frac{2}{t_k - t_{k-1}} \frac{\mathbf{\Gamma}_j^{(k)}}{w_j^{(k)}}, \quad 1 \le k \le K, \quad 1 \le j \le N_k, \tag{68}
$$

$$
\boldsymbol{\lambda}_j^{(k)} = \frac{\mathbf{\Lambda}_j^{(k)}}{w_k^{(k)}}, \quad 1 \le k \le K, \quad 1 \le j \le N_k, \tag{69}
$$

$$
\boldsymbol{\pi}^{(k)} = \mathbf{D}_{N_k+1}^{(k)^{\mathsf{T}}} \mathbf{\Lambda}^{(k)}, \quad 1 \le k \le K. \tag{70}
$$

In addition, let the matrix $\mathbf{D}^{\dagger(k)}$ be defined as

$$
D_{11}^{\dagger(k)} = -D_{11}^{(k)} - \frac{1}{w_1^{(k)}}, \tag{71}
$$

$$
D_{ij}^{\dagger(k)} = -\frac{w_j^{(k)}}{w_i^{(k)}} D_{ji}^{(k)} \quad \text{otherwise.} \tag{72}
$$

Substituting Eqs. (67)–(72) into Eqs. (63)–(66), the transformed adjoint system is given as

$$
\mathbf{0} = \nabla_{\mathbf{U}} H(\mathbf{X}_j^{(k)}, \mathbf{U}_j^{(k)}, \boldsymbol{\lambda}_j^{(k)}, \boldsymbol{\gamma}_j^{(k)}), \quad 1 \le k \le K, \quad 1 \le j \le N_k \tag{73}
$$

$$
\mathbf{D}_j^{\dagger(1)} \boldsymbol{\lambda}^{(1)} = -\frac{t_1 - t_0}{2} \nabla_{\mathbf{X}} H(\mathbf{X}_j^{(1)}, \mathbf{U}_j^{(1)}, \boldsymbol{\lambda}_j^{(1)}, \boldsymbol{\gamma}_j^{(1)}) \tag{74}
$$

$$
+ \frac{\delta_{1j}}{w_1^{(1)}} (-\nabla_{\mathbf{X}_1^{(1)}} \Phi + \nabla_{\mathbf{X}_1^{(1)}} \langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle - \boldsymbol{\lambda}_1^{(1)}),
$$

$$
\mathbf{D}_j^{\dagger(k)} \boldsymbol{\lambda}^{(k)} = -\frac{t_k - t_{k-1}}{2} \nabla_{\mathbf{X}} H(\mathbf{X}_j^{(k)}, \mathbf{U}_j^{(k)}, \boldsymbol{\lambda}_j^{(k)}, \boldsymbol{\gamma}_j^{(k)}) \tag{75}
$$

$$
+ \frac{\delta_{1j}}{w_1^{(k)}} (\boldsymbol{\pi}^{(k-1)} - \boldsymbol{\lambda}_1^{(k)}), \quad 2 \le k \le K, \quad 1 \le j \le N_k
$$

$$
\boldsymbol{\pi}^{(K)} = \nabla_{\mathbf{X}_{N_K+1}^{(K)}} (\mathbf{\Phi} - \langle \boldsymbol{\psi}, \boldsymbol{\phi} \rangle). \tag{76}
$$

15

Analogous to Ref. 13, we expand Eq. (70) using Eqs. (64) and (65) to obtain the following relationships:

$$-\nabla_{\mathbf{X}_1^{(1)}}\Phi + \nabla_{\mathbf{X}_1^{(1)}}\langle\boldsymbol{\psi},\boldsymbol{\phi}\rangle = \boldsymbol{\pi}^{(1)} + \frac{t_1 - t_0}{2}\sum_{j=1}^{N_1} w_j^{(1)}\nabla_{\mathbf{x}}H(\mathbf{X}_j^{(1)},\mathbf{U}_j^{(1)},\boldsymbol{\lambda}_j^{(1)},\boldsymbol{\gamma}_j^{(1)}). \qquad (77)$$

$$\boldsymbol{\pi}^{(k-1)} = \boldsymbol{\pi}^{(k)} + \frac{t_k - t_{k-1}}{2}\sum_{j=1}^{N_k} w_j^{(k)}\nabla_{\mathbf{X}}H(\mathbf{X}_j^{(k)},\mathbf{U}_j^{(k)},\boldsymbol{\lambda}_j^{(k)},\boldsymbol{\gamma}_j^{(k)}). \quad 2 \le k \le K \qquad (78)$$

Next, the costate approximation at $t = t_f$ is given as

$$\boldsymbol{\lambda}_{N_K+1}^{(K)} = \nabla_{\mathbf{X}_{N_K+1}^{(K)}}(\Phi - \langle\boldsymbol{\psi},\boldsymbol{\phi}\rangle) \qquad (79)$$

from which we obtain

$$\boldsymbol{\lambda}_{N_K+1}^{(K)} = \boldsymbol{\pi}^{(K)}. \qquad (80)$$

The right-hand side of Eq. (78) for $k = K$ is then an approximation to $\boldsymbol{\lambda}_1^{(K)}$, and therefore, $\boldsymbol{\lambda}_1^{(K)} = \boldsymbol{\pi}^{(K-1)}$. Consequently, the final term in Eq. (75) disappears for $k = K$. Furthermore, if the costate is continuous at the mesh point $k = K - 1$, we obtain

$$\boldsymbol{\lambda}_{N_{K-1}+1}^{(K-1)} = \boldsymbol{\lambda}_1^{(K)} = \boldsymbol{\pi}^{(K-1)}. \qquad (81)$$

Furthermore, for $k = K - 1$, the right hand side of Eq. (78) becomes an approximation to $\boldsymbol{\lambda}_1^{(K-1)}$. Thus, we have

$$\boldsymbol{\lambda}_1^{(K-1)} = \boldsymbol{\pi}^{(K-2)}. \qquad (82)$$

If the costate is continuous at every mesh point, it then follows that

$$\boldsymbol{\lambda}_{N_k+1}^{(k)} = \boldsymbol{\lambda}_1^{(k+1)} = \boldsymbol{\pi}^{(k)}, \quad 1 \le k \le K - 1. \qquad (83)$$

As a result, the left hand side of Eq. (77) then is an approximation to the costate at t_0. Therefore, for the case when the costate is continuous at every mesh point, the final term in Eqs. (74) and (75) disappears and the transformed adjoint system is a discrete representation of continuous-time first-order optimality conditions.

Suppose now that the costate is discontinuous at a particular mesh point $M \in [1, \ldots, K - 1]$. Then

$$\boldsymbol{\pi}^{(M)} = \boldsymbol{\lambda}_1^{(M+1)} = \boldsymbol{\lambda}_{N_M+1}^{(M)} - \boldsymbol{\rho}^{(M)} \qquad (84)$$

where $\boldsymbol{\rho}^{(M)}$ is the difference between $\boldsymbol{\lambda}_{N_M+1}^{(M)}$ and $\boldsymbol{\lambda}_1^{(M+1)}$. Therefore, if the costate is discontinuous at mesh point M, $\boldsymbol{\pi}^{(M-1)} \ne \boldsymbol{\lambda}_1^{(M)}$ in Eq. (75). As a result, the transformed adjoint system given in Eqs. (73)–(76) is not a discrete representation of the first-order optimality conditions given in Eqs. (14)–(18).

5 Examples

The costate estimation methods derived in Section 4 are now applied to two optimal control problems that have been studied extensively. In the first example the optimal costate is smooth, while in the second example the optimal costate is discontinuous. The purpose of the examples is to assess the accuracy of the multiple-interval LG and LGR costate estimation procedures as functions of the number of collocation points and the degree of the polynomial approximation and compare them against single interval p methods. A performance analysis on the methods is not given with respect to CPU time, number of collocation points, or NLP density. See Ref. 16 and Ref. 17 for performance analyses of multiple-interval LG and LGR methods.

The examples were solved using the open-source program *General Pseudospectral Optimal Control Software*[7] (GPOPS) with the NLP solver SNOPT.[24] All computations were performed using a 2.5 GHz Core 2 Duo Macbook Pro running Mac OS-X 10.5.8 with MAT-LAB R2009b. The p method discretizations were solved utilizing a fixed number of intervals with a variable number of collocation points in each interval. For the first example and the first part of the second example, a single interval was used. In the second part of the second example three intervals were used and these three intervals were connected at the locations of the costate discontinuity. The h method discretizations were solved utilizing a fixed low-degree approximation in each interval using a variable number of intervals. The hp method discretizations were solved utilizing a variable number of collocation points per interval and a variable number of intervals. Throughout this section, the terminology "$h - x$" denotes an h method of degree x in every mesh interval, while "$hp - x$" denotes an hp method with a *minimum* polynomial degree of x in each interval.

5.1 Example 1

Consider the following optimal control problem taken from Ref. 13. Minimize the cost functional

$$J = \frac{1}{2} \int_0^{t_f} (y + u^2) dt \tag{85}$$

subject to the dynamic constraint

$$\dot{y} = 2y + 2u\sqrt{y} \tag{86}$$

and the boundary conditions

$$y(0) \;=\; 2 \;\;,\;\; y(t_f) \;=\; 1, \tag{87}$$

where $t_f = 5$. The optimal state and costate for this example are $y^*(t) = (x^*(t))^2$ and $\lambda_y^*(t) = \lambda_x^*(t)/(2x^*(t))$, respectively, where $x^*(t)$ and $\lambda_x^*(t)$ are given as

$$\begin{bmatrix} x^*(t) \\ \lambda_x^*(t) \end{bmatrix} = \exp(\mathbf{A}t) \begin{bmatrix} x_0 \\ \lambda_{x0} \end{bmatrix}, \quad \mathbf{A} = \begin{bmatrix} 1 & -1 \\ -1 & -1 \end{bmatrix}, \tag{88}$$

and

$$x_0 \;=\; \sqrt{2}, \quad x_f = 1, \tag{89}$$

$$\mathbf{B} \;=\; \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix} = \exp(\mathbf{A}t_f), \tag{90}$$

$$\lambda_{x0} \;=\; \frac{x_f - B_{11}x_0}{B_{12}}. \tag{91}$$

In this example we analyze the costate errors using the LG and LGR p, h–2, h–3, and h–4 methods. It is noted that the h–x methods are constructed with a uniformly distributed grid of mesh points. Because this example has a smooth solution, it is expected that a p method will converge the fastest. Furthermore, because the optimal costate is continuous, it is expected that the multiple-interval LG and LGR methods (i.e., h) will provide accurate approximations to the optimal costate because the transformed adjoint systems defined in Section 4 are discrete representations of the continuous-time first-order optimality conditions.

Figs. 1a and 1b show the base ten logarithm of the maximum absolute costate error as a function of the number of collocation points for LG and LGR p and h–x methods, respectively. Because the solution to this problem is smooth (that is, $\mathbf{x}^*(t)$, $\mathbf{u}^*(t)$, and $\boldsymbol{\lambda}^*(t)$ are all smooth), the p method converges the fastest using either discretization scheme. Next, it is seen in Figs. 1a and 1b that the convergence rate using the h–x methods is much slower than the convergence rate using the p methods. For the h–x methods, the rate of convergence increases with the degree of approximating polynomial. A closer examination of Figs. 1a and 1b also reveals that the costate estimate using an LG h–x method converges faster than the corresponding costate estimate using an LGR h–x method.

5.2 Example 2

Consider the following optimal control minimum-energy optimal control problem with an inequality state constraint taken from Ref. 23. Minimize the cost functional

$$J = \frac{1}{2} \int_0^1 a^2 dt \tag{92}$$

subject to the dynamic constraints

$$\dot{x} = v, \tag{93}$$

$$\dot{v} = a, \tag{94}$$

the boundary conditions

$$x(0) = x(1) = 0, \tag{95}$$

$$v(0) = -v(1) = 1, \tag{96}$$

and the constraint $x(t) \leq \ell$. The optimal costate for this example is

$$
\lambda_x^* = \begin{cases} \frac{2}{9\ell^2} & , & 0 \leq t \leq 3\ell, \\ 0 & , & 3\ell \leq t \leq 1 - 3\ell, \\ -\frac{2}{9\ell^2} & , & 1 - 3\ell \leq t \leq 1, \end{cases} \tag{97}
$$

$$
\lambda_v^* = \begin{cases} \frac{2}{3\ell}\left(1 - \frac{t}{3\ell}\right) & , & 0 \leq t \leq 3\ell, \\ 0 & , & 3\ell \leq t \leq 1 - 3\ell, \\ \frac{2}{3\ell}\left(1 - \frac{1-t}{3\ell}\right) & , & 1 - 3\ell \leq t \leq 1. \end{cases} \tag{98}
$$

Unlike the optimal solution to Example 1, where the optimal solution was smooth, it is seen in Eqs. (97) and (98) that $\lambda_x^*(t)$ is discontinuous while $\lambda_v^*(t)$ is continuous but not smooth. In this example, we study the numerical approximations obtained for $\ell = 1/12$. Furthermore, we consider discretizations where no mesh points are at the locations of the discontinuity in $\lambda_x^*(t)$ and discretizations where mesh points *are* located at the discontinuity in $\lambda_x^*(t)$.

5.2.1 Case 1: No Mesh Points at Costate Discontinuity

In this case we analyze LG and LGR p, uniformly spaced h–2, and hp–3 methods. In general, mesh points will not be located exactly at a costate discontinuity. Figs. 2a–3b show the base ten logarithm of the maximum absolute error at the collocation points for the costate as a function of the number of collocation points using LG and LGR p methods and uniformly

distributed h–2 methods, respectively. First, it is seen that essentially no difference is seen in the convergence rates of the LG and LGR p and h–2 methods. Next, while λ_v converges (albeit slowly), the first component of the costate does not converge when a mesh point is not located at the discontinuity.

Suppose now that we attempt to improve the convergence of the costate by applying the hp–3 method of Ref. 17 to this example, where it is noted that the hp–adaptive method of Ref. 17 is designed to capture discontinuities in the control and potential nonsmoothness in the state. Table 2 shows the maximum costate errors for various accuracy tolerances ϵ, while Figs. 4a and 4b show the costate approximation and the location of the mesh points on the final grid for $\epsilon = 10^{-4}$. While the hp–adaptive method places mesh points near the costate discontinuity, it is seen that the approximation still leads to inaccuracies at the point of the costate discontinuity (in particular, see the value of $\lambda_x(t)$ at $t \approx 0.25$ in Fig. 4a). Thus, even though the hp–adaptive method refines the mesh near the discontinuity in λ_x, the errors in λ_x are still large near the discontinuities because no mesh points are located at the discontinuity itself. Finally, it is noted for $\epsilon = 10^{-6}$ that the errors in the state and control are $\mathcal{O}(10^{-6})$ and $\mathcal{O}(10^{-3})$, respectively. This last fact demonstrates that placing the mesh points near but not exactly at the point of the costate discontinuity significantly decreases the error in the state and control, but has a significantly lesser effect on the accuracy of the costate. In the next section we show the results when a mesh point is located at the location of the costate discontinuity.

5.2.2 Case 2: Location of Costate Discontinuity Known

Suppose now that we use the fact that we know the locations of the discontinuities in $\lambda_x^*(t)$ to aid in improving the accuracy of the costate. Specifically, let us apply the h–3 and p methods to this problem by dividing the time interval into three segments $T_1 = [0, 3\ell]$, $T_2 = [3\ell, 1-3\ell]$, and $T_3 = [1-3\ell, 1]$. In the h–3 method, T_1 and T_3 are divided into uniformly spaced intervals while three collocation points are used in T_2. In the p method, a single mesh interval is used in each segment T_1, T_2, and T_3.

Figs. 5 and 6 show the maximum costate errors using the LG and LGR h–3 and p methods as a function of the number of mesh intervals (using the h–3 method) and as a function of the polynomial degree (for the p method) when a mesh point is located at each of the costate discontinuities. It is seen that the accuracy in the costate is improved dramatically when

mesh points coincide with the discontinuities. Interestingly, the LG method produces errors that are near machine precision using a small number of mesh intervals (in the case of the h–3 method) or using a small polynomial degree (in the case of the p method). Moreover, as the number of approximating intervals or degree of approximation is increased, it is interesting to see the accuracy of the LG method *decreases*. For the LGR method, as the number of approximating intervals or degree of approximation is increased, the LGR method does not significantly increase or decrease in accuracy. Rather, the LGR accuracy lies between $\mathcal{O}(10^{-4})$ and $\mathcal{O}(10^{-12})$. Next, comparing Figs. 5a and 5b, it is seen that the errors in the two components of the costate are nearly identical using an $h - 3$ method. Thus, even though only $\lambda_x^*(t)$ is discontinuous while $\lambda_v^*(t)$ is continuous, the size of the errors in the two components of the costate are the same because the transformed adjoint system is not a discrete representation of the costate dynamics. Furthermore, Figs. 6a and 6b show that using LG and LGR p methods the approximation of $\lambda_v^*(t)$ is slightly more accurate than the approximation of $\lambda_x^*(t)$.

6 Discussion

The two examples studied in Section 5 demonstrate some of the key characteristics of the LG and LGR hp schemes. In Example 1, where the solution is smooth, the LG and LGR costate approximations converge using h, p, and hp methods. These results are consistent with the fact that the costate is continuous, thereby making the LG and LGR transformed adjoint system discrete representations of the continuous-time first-order optimality conditions. Furthermore, and consistent with the fact that the solution to Example 1 is smooth, a p method converges the most rapidly of all of the different discretization methods.

Next, the results obtained for Example 2, where the optimal costate is discontinuous, are quite different from those obtained in Example 1. When mesh points are not located at the discontinuities, the discontinuous components of the costate do not converge for any of the methods. Furthermore, even for the components of the costate that are continuous, the convergence is extremely slow. In order to improve the costate approximation in Example 2, mesh points are located at the costate discontinuities. While improving the placement of the mesh points still does not result in the transformed adjoint systems being discrete representations of the continuous-time first-order optimality conditions, the accuracy of the

LG and LGR costate approximations is much higher with proper mesh point placement than it was when no mesh points were located at the costate discontinuities. Finally, even though in Example 2 the second component of the optimal costate was continuous, the errors in this second costate component were similar to those of the first (discontinuous) costate component.

In general, it is expected that the costate approximation methods derived in this paper will lead to accurate approximations for problems whose optimal costate are continuous. In the case where the optimal costate is discontinuous, however, the accuracy is expected to depend greatly on the placement of the mesh points. In particular, if mesh points are not placed at the discontinuities in the costate, it is expected that the costate errors will be large even if the corresponding errors in the state and control are significantly smaller.

7   Conclusions

Methods have been presented for costate estimation using multiple-interval collocation at Legendre-Gauss and Legendre-Gauss-Radau points. A mapping from the Lagrange multipliers of the multiple-interval discretized optimal control problem to the costate of the continuous-time optimal control problem has been derived for each method. The requirement to enforce continuity in the state at the mesh points leads to a significantly different transformation of the Lagrange multipliers of the pseudospectral nonlinear programming problem than is obtained using a global pseudospectral method. If the costate is continuous, the transformed adjoint system is a discrete representation of the continuous-time first-order optimality conditions. If, however, the costate is discontinuous, the transformed adjoint system is not an exact discrete representation of the continuous-time first-order optimality conditions. It was found that, even though in the case of a discontinuous costate the transformed adjoint system was not a discrete representation of the continuous-time first-order optimality conditions, the accuracy of the approximation could be improved if mesh points were placed at the locations of the discontinuities.

Acknowledgments

References

[1] 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, 1995, pp. 1793–1796.

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

[3] Fahroo, F. and Ross, I. M., "A Spectral Patching Method for Direct Trajectory Optimization," *Journal of Astronautical Sciences*, Vol. 48, No. 2-3, Apr.-Sept. 2000, pp. 269–286.

[4] Fahroo, F. and Ross, I. M., "Costate Estimation by a Legendre Pseudospectral Method," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 2, 2001, pp. 270–277.

[5] Ross, I. M. and Fahroo, F., "Pseudospectral Knotting Methods for Solving Optimal Control Problems," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 3, 2004, pp. 397–405.

[6] Gong, Q., Ross, I. M., Kang, W., and Fahroo, F., "Connections Between the Covector Mapping Theorem and Convergence of Pseudospectral Methods," *Computational Optimization and Applications*, Vol. 41, No. 3, December 2008, pp. 307–335.

[7] Rao, A. V., Benson, D. A., Darby, C. L., Francolin, C., Patterson, M. A., Sanders, I., and Huntington, G. T., "Algorithm 902: GPOPS, A Matlab Software for Solving Multiple-Phase Optimal Control Problems Using the Gauss Pseudospectral Method," *ACM Transactions on Mathematical Software*, Vol. 37, No. 2, April–June 2010, pp. 22:1–22:39.

[8] Williams, P., "Jacobi Pseudospectral Method for Solving Optimal Control Problems," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 2, 2004, pp. 293–297.

[9] Williams, P., "Application of Pseudospectral methods for Receding Horizon Control," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 2, 2004, pp. 310–314.

[10] Benson, D. A., *A Gauss Pseudospectral Transcription for Optimal Control*, Ph.D. thesis, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2004.

[11] 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, November-December 2006, pp. 1435–1440.

[12]Huntington, G. T., *Advancement and Analysis of a Gauss Pseudospectral Transcription for Optimal Control*, Ph.D. thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, 2007.

[13]Garg, D., Patterson, M. A., Darby, C. L., Francolin, C., Huntington, G. T., Hager, W. W., and Rao, A. V., "Direct Trajectory Optimization and Costate Estimation of Finite-Horizon and Infinite-Horizon Optimal Control Problems via a Radau Pseudospectral Method," *Computational Optimization and Applications*, Published Online: October 6, 2009. DOI 10.1007/s10589-009-9291-0. http://www.springerlink.com/content/n851q6n343p9k60k/.

[14]Garg, D., Patterson, M. A., Hager, W. W., Rao, A. V., Benson, D. A., and Huntington, G. T., "A Unified Framework for the Numerical Solution of Optimal Control Problems Using Pseudospectral Methods," *Automatica*, Vol. 46, No. 11, December 2010, pp. 1843–1851.

[15]Kameswaran, S. and Biegler, L. T., "Convergence Rates for Direct Transcription of Optimal Control Problems Using Collocation at Radau Points," *Computational Optimization and Applications*, Vol. 41, No. 1, 2008, pp. 81–126.

[16]Darby, C. L., Hager, W. W., and Rao, A. V., "An hp-Adaptive Pseudospectral Method for Solving Optimal Control Problems," *Optimal Control Applications and Methods*, Accepted for Publication June 2010 (in press).

[17]Darby, C. L., Hager, W. W., and Rao, A. V., "A Variable Order hp-Adaptive Pseudospectral Method for Solving Optimal Control Problems," *Journal of Spacecraft and Rockets*, Accepted for Publication November 2010 (in press).

[18]Garg, D., Hager, W. W., and Rao, A. V., "Pseudospectral Methods for Solving Infinite-Horizon Optimal Control Problems," *Automatica*, Accepted For Publication, November 2010.

[19]Gong, Q., Kang, W., and Ross, I. M., "A Pseudospectral Method for the Optimal Control of Feedback Linearizable Systems," *IEEE Transactions on Automatic Control*, Vol. 51, No. 7, July 2006, pp. 1115–1129.

[20]Hager, W. W., "Runge-Kutta Methods in Optimal Control and the Transformed Adjoint System," *Numerische Mathematik*, Vol. 87, 2000, pp. 247–282.

[21]Abramowitz, M. and Stegun, I., *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover Publications, New York, 1965.

[22]Kirk, D. E., *Optimal Control Theory: An Introduction*, Dover Publications, Mineola, New York, 2004.

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

[24]Gill, P. E., Murray, W., and Saunders, M. A., *User's Guide for SNOPT Version 7: Software for Large Scale Nonlinear Programming*, February 2006.
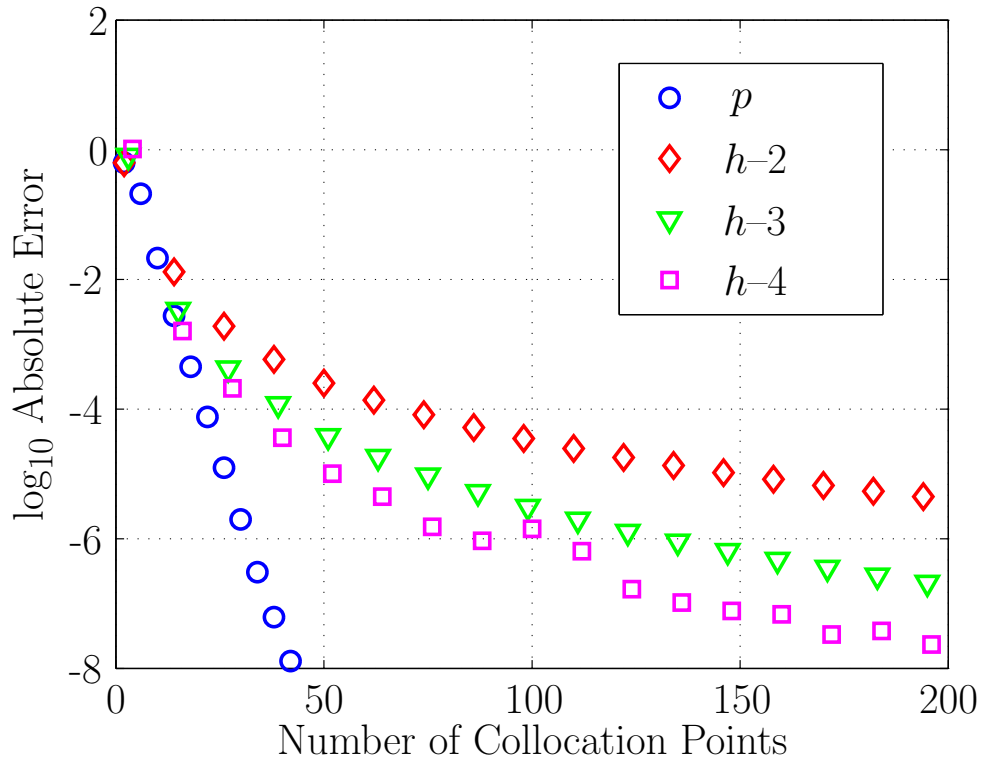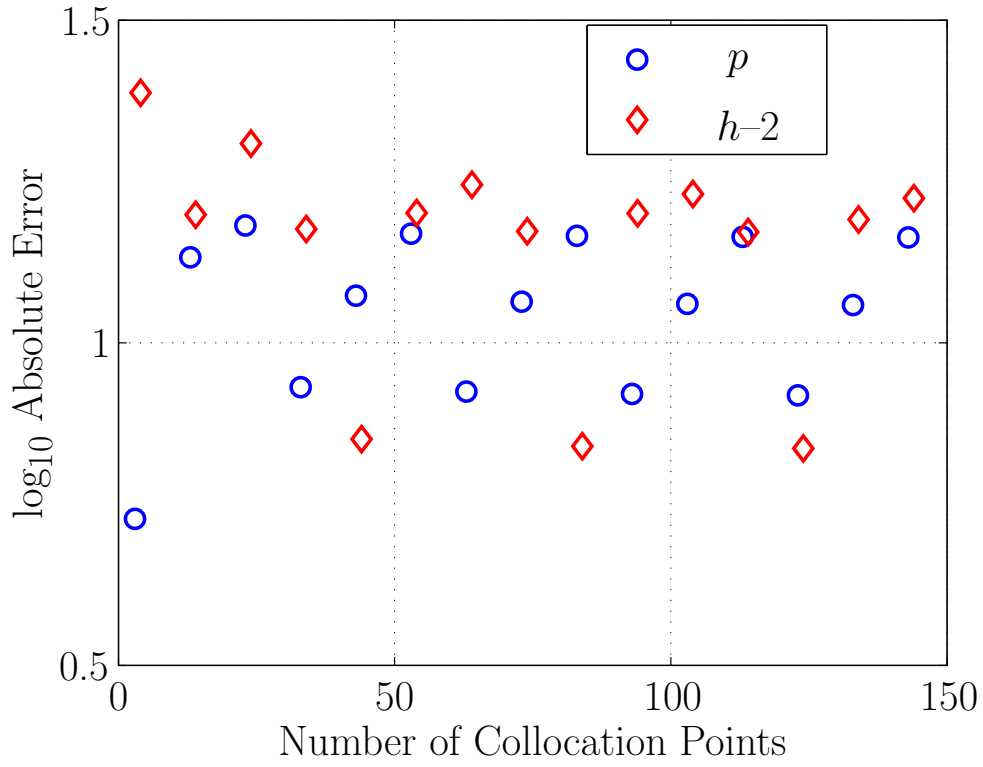
(a) Base Ten Logarithm of Maximum Absolute Error vs. Number of Collocation Points for the Costate Using the LG Method.
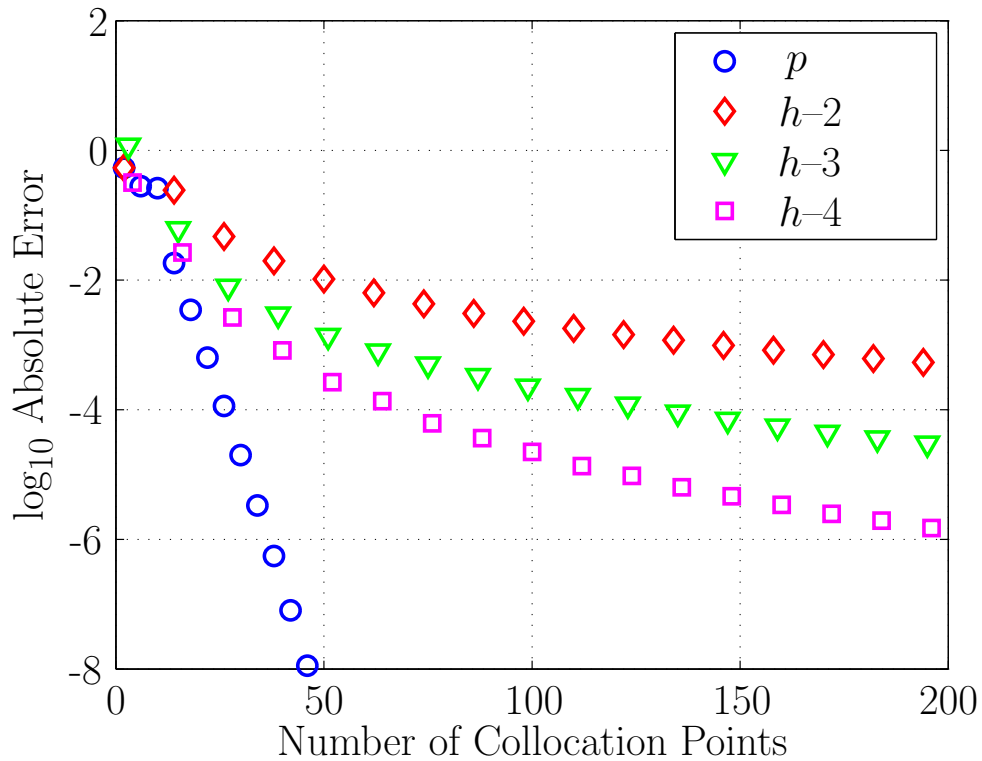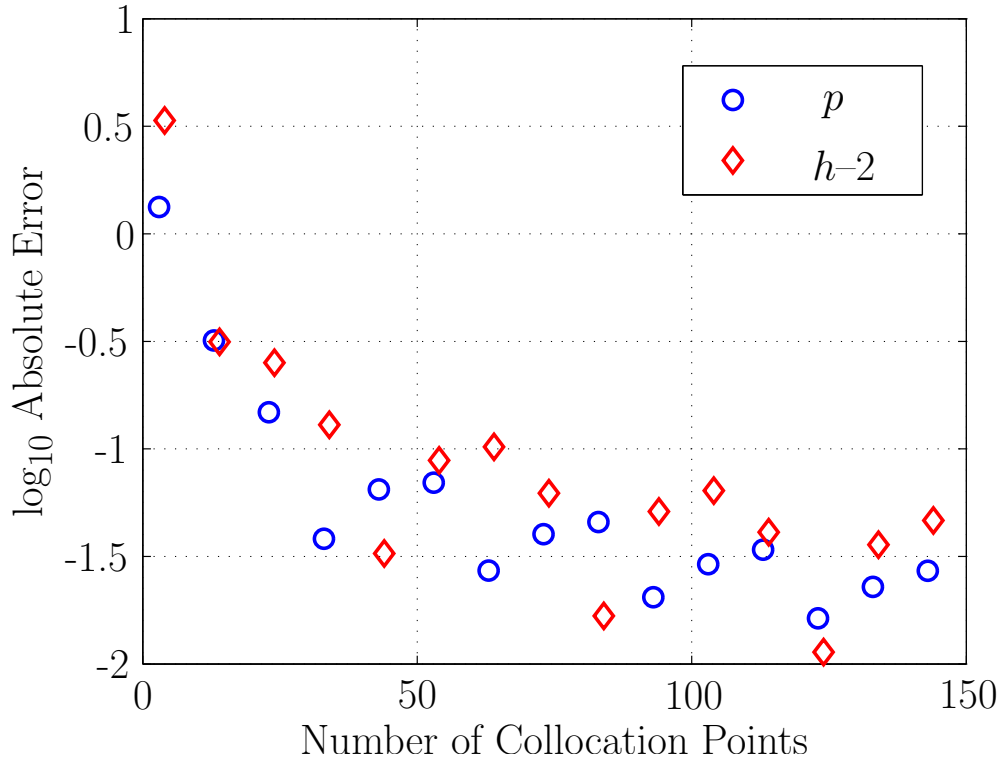

(b) Base Ten Logarithm of Maximum Absolute Error vs. Number of Collocation Points for the Costate Using the LGR Method.

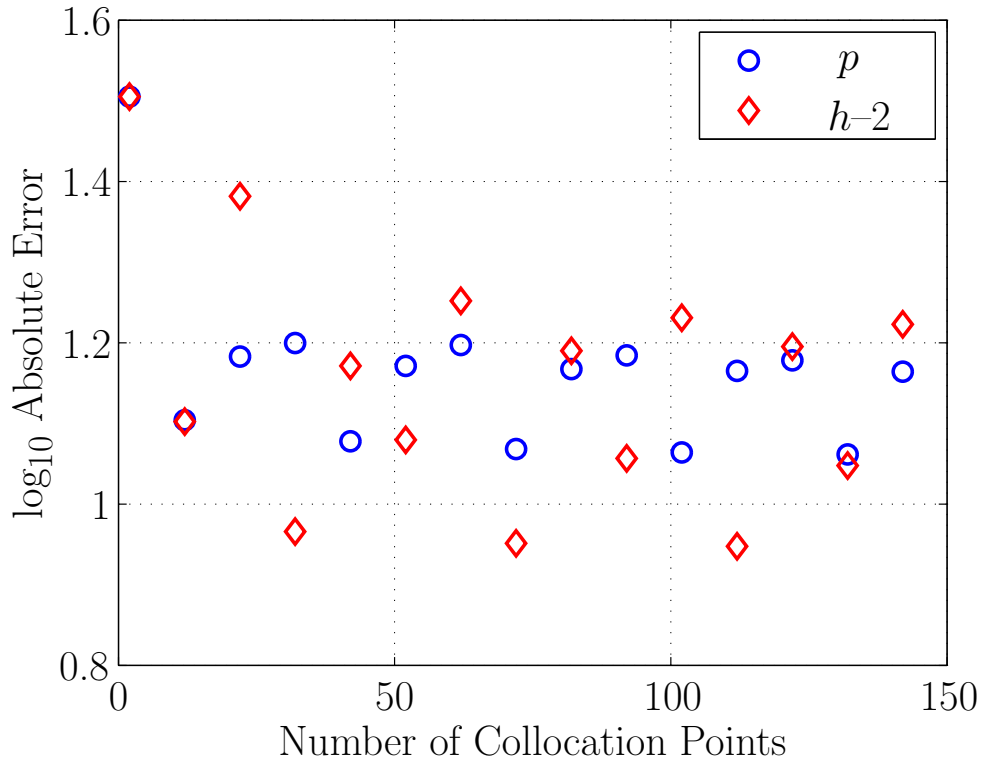Figure 1: Base Ten Logarithm of LG and LGR p-, h–2, h–3, and h–4 Absolute Costate Errors for Example 1.

(a) Base Ten Logarithm of Maximum Absolute Error in $\lambda_x^*(t)$ vs. Number of Collocation Points Using the LG Method.
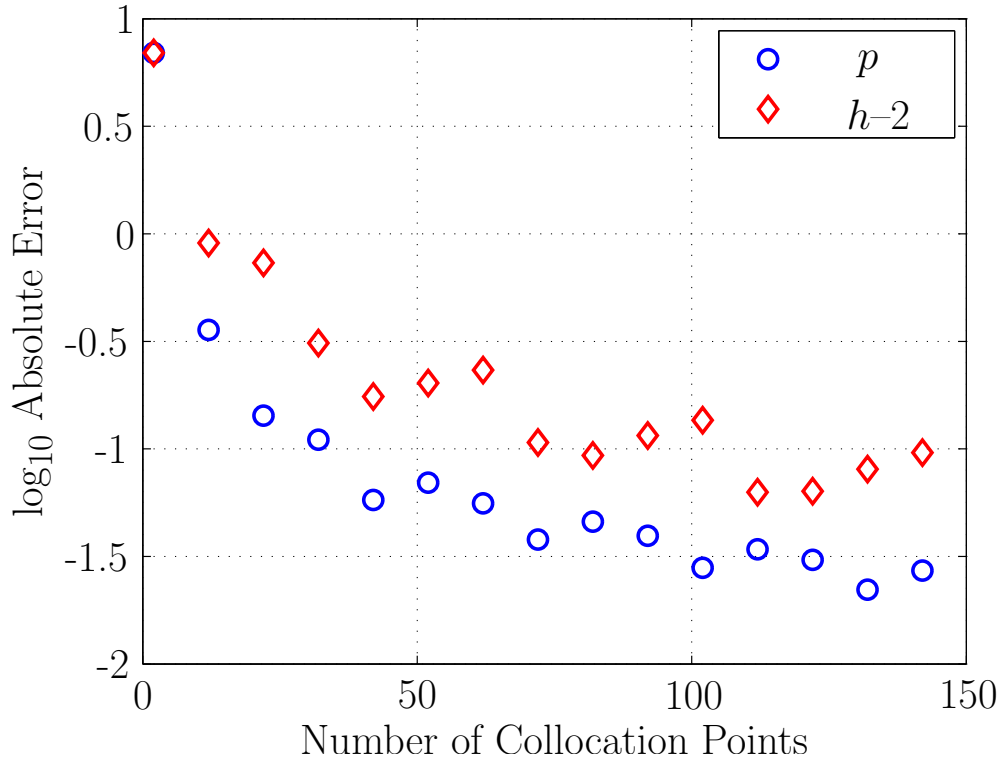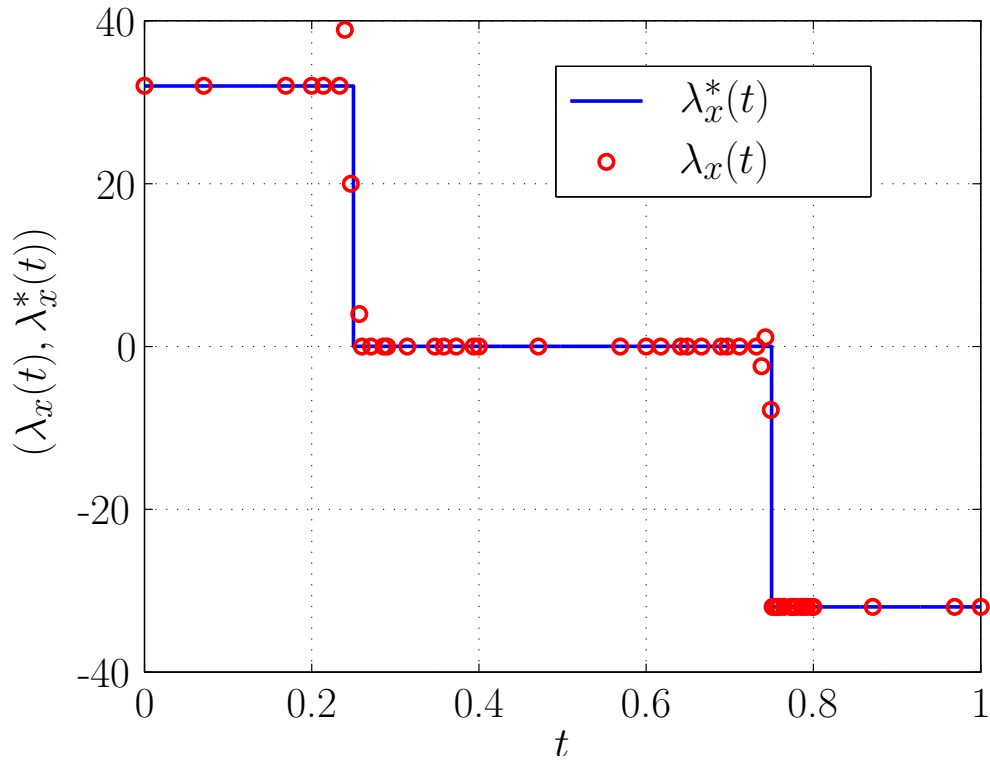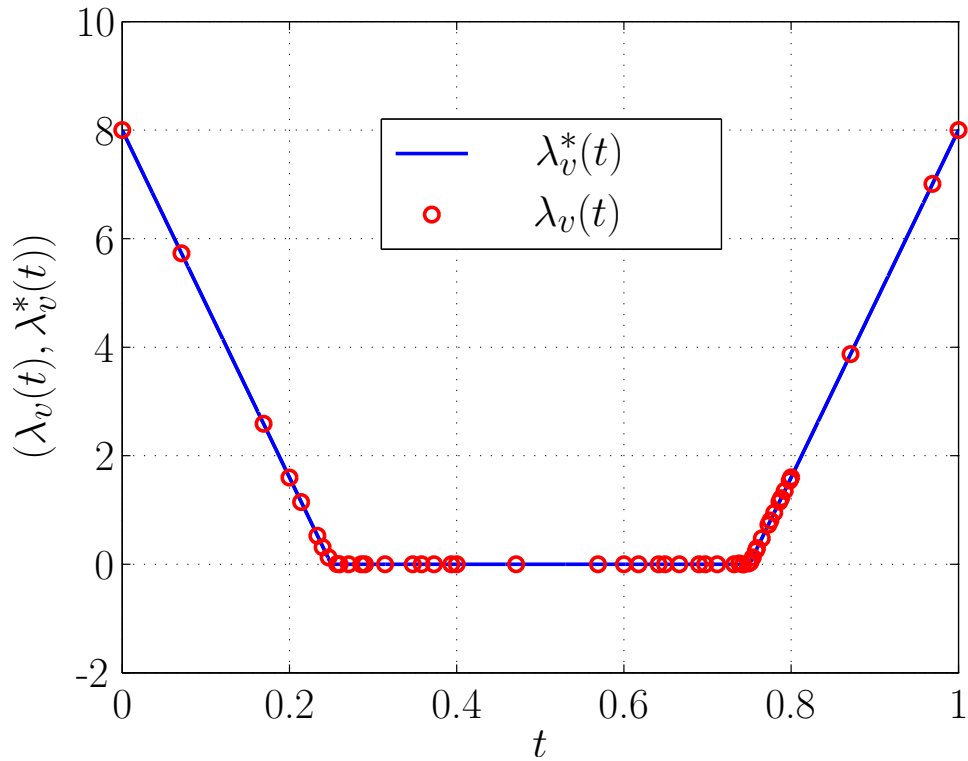

(b) Base Ten Logarithm of Maximum Absolute Error in $\lambda_v^*(t)$ vs. Number of Collocation Points Using the LG Method.

Figure 2: Base Ten Logarithm of LG Uniformly Spaced h–2 and p method Absolute Costate Errors for Example 2 When Mesh Points are Not Located at the Costate Discontinuity.
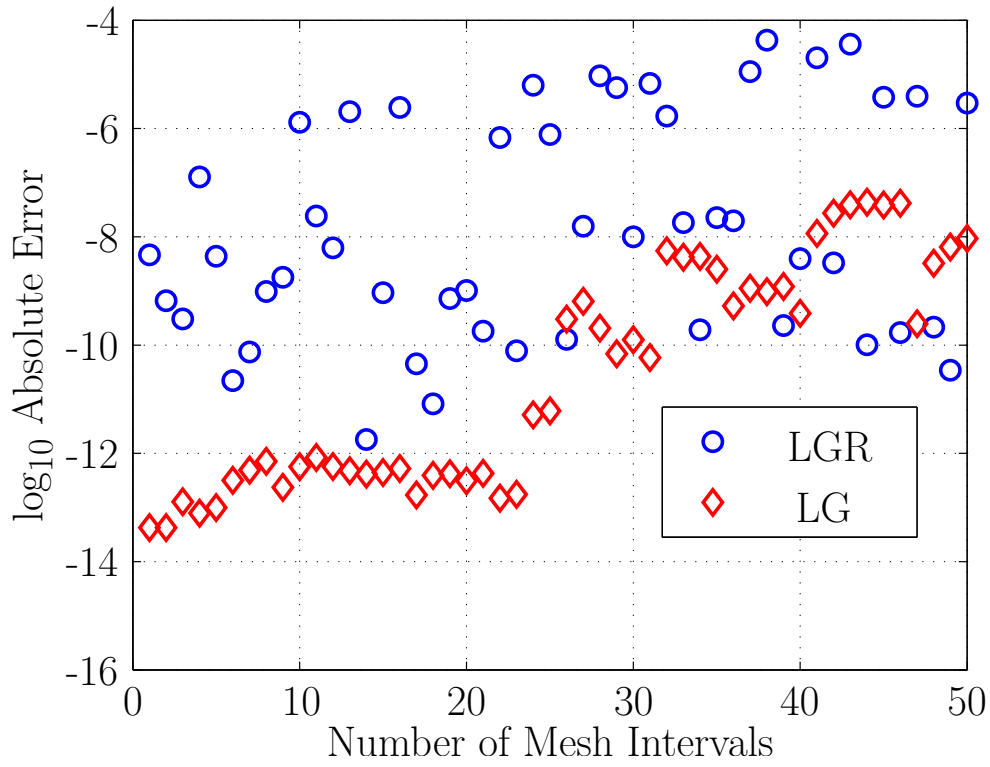
(a) Base Ten Logarithm of Maximum Absolute Error in $\lambda_x^*(t)$ vs. Number of Collocation Points Using the LGR Method.

(b) Base Ten Logarithm of Maximum Absolute Error in $\lambda_v^*(t)$ vs. Number of Collocation Points Using the LGR Method.

Figure 3: Base Ten Logarithm of LGR Uniformly Spaced h–2 and p method Absolute Costate Errors for Example 2 When Mesh Points are Not Located at the Costate Discontinuity.
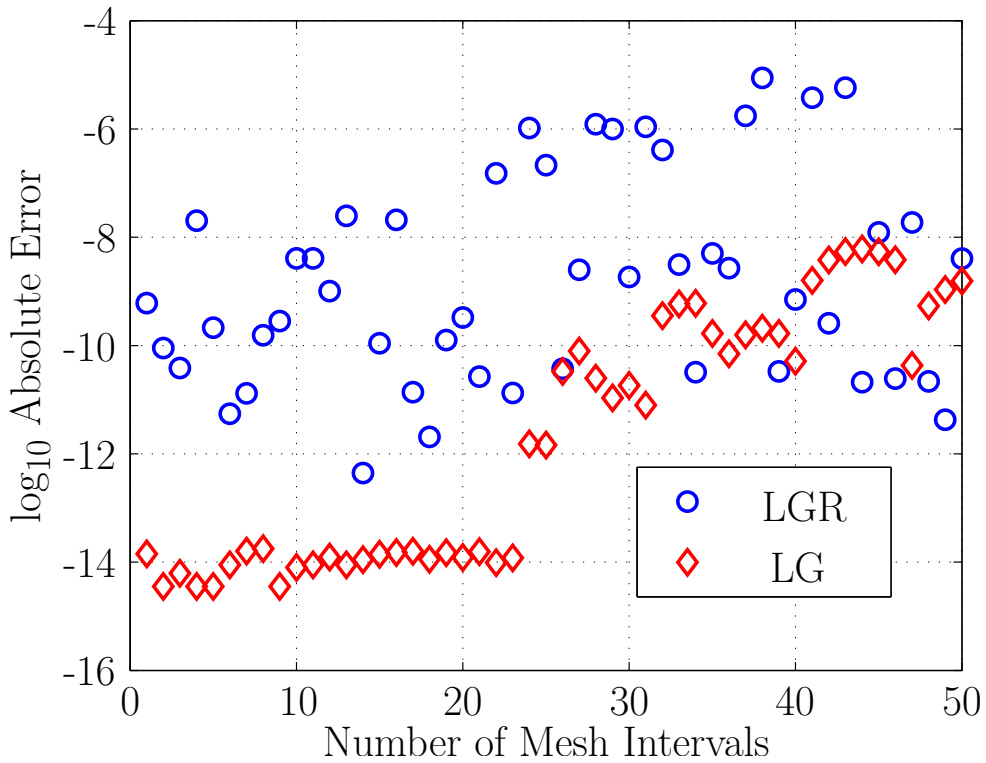
(a) $\lambda_x^*(t)$ and $\lambda_x(t)$.

(b) $\lambda_v^*(t)$ and $\lambda_v(t)$.

Figure 4: Costate Approximation Using hp–3 Method with $\epsilon = 10^{-4}$ for Example 2 Alongside Optimal Solution.
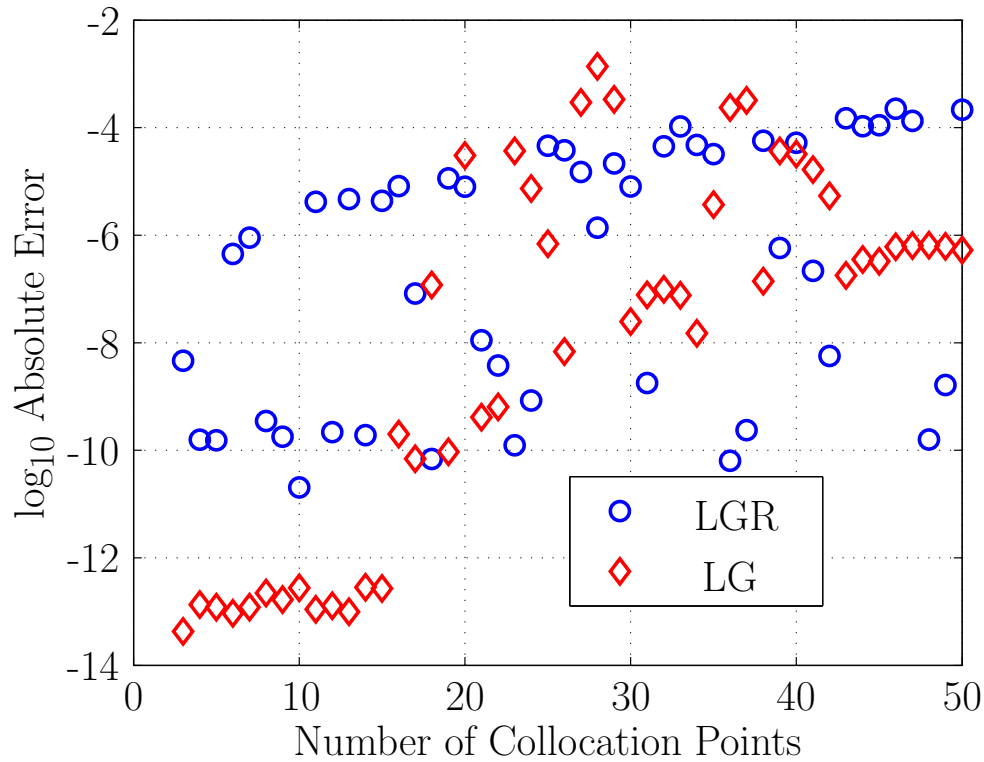
(a) Base Ten Logarithm of Maximum Absolute Error in $\lambda_x^*(t)$ vs. Number of Mesh Intervals in Segments T_1 and T_3 Using the LG and LGR h–3 Methods.
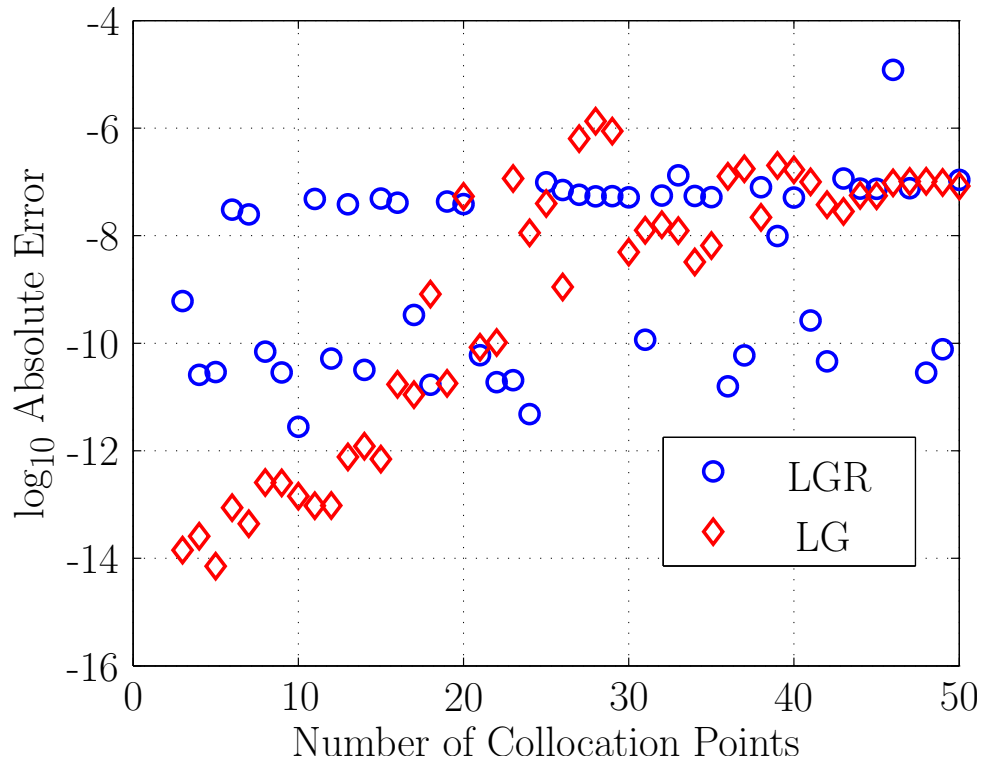

(b) Base Ten Logarithm of Maximum Absolute Error in $\lambda_v^*(t)$ vs. Number of Mesh Intervals in Segments T_1 and T_3 Using the LG and LGR h–3 Methods.

Figure 5: Base Ten Logarithm of LG and LGR h–3 Method Absolute Costate Errors for Example 2 When Mesh Points are Located at the Costate Discontinuity.

(a) Base Ten Logarithm of Maximum Absolute Error in $\lambda_x^*(t)$ vs. Degree of Approximating Polynomial in Segments T_1, T_2, and T_3 Using the LG and LGR p Methods.

(b) Base Ten Logarithm of Maximum Absolute Error in $\lambda_v^*(t)$ vs. Degree of Approximating Polynomial in Segments T_1, T_2, and T_3 Using the LG and LGR p Methods.

Figure 6: Base Ten Logarithm of LG and LGR p Method Absolute Costate Errors for Example 2 When Mesh Points are Located at the Costate Discontinuity.

ϵ	λ_x	λ_v	Number of Collocation Points
10^{-2}	1.39×10^1	2.02×10^{-1}	30
10^{-4}	2.32×10^1	2.99×10^{-2}	58
10^{-6}	1.30×10^1	2.32×10^{-3}	87

Table 2: Summary of Maximum Absolute Costate Errors for Example 2 Using an hp–3 Method for Various Accuracy Tolerances, ϵ.