Modified Legendre-Gauss-Radau Collocation Method for Solving Optimal Control Problems with Nonsmooth Solutions

Joseph D. Eide∗
William W. Hager†
Anil V. Rao‡

University of Florida
Gainesville, FL 32611

Abstract
A new method is developed for solving optimal control problems whose solutions are nonsmooth. The method developed in this paper employs a modified form of the Legendre-Gauss-Radau orthogonal direct collocation method. This modified Legendre-Gauss-Radau method adds two variables and two constraints at the end of a mesh interval when compared with a previously developed standard Legendre-Gauss-Radau collocation method. These new variables are the time and the control at the end of each mesh interval. The two additional constraints are a collocation condition on each differential equation that is a function of control and an inequality constraint on the control at the end of each mesh interval. These additional constraints modify the search space of the nonlinear programming problem such that an accurate approximation to the location of the nonsmoothness is obtained. The transformed adjoint system of the modified Legendre-Gauss-Radau method is then developed. Using this transformed adjoint system, a method is developed to transform the Lagrange multipliers of the nonlinear programming problem to the costate of the optimal control problem. Furthermore, it is shown that the costate estimate satisfies the Weierstrass-Erdmann optimality conditions. Finally, the method developed in this paper is demonstrated on an example whose solution is nonsmooth.

1 Introduction

Over the past two decades, direct collocation methods have become increasingly popular for computing the numerical solution of constrained optimal control problems. A direct collocation method is an implicit simulation method where the state and control are both parameterized and the constraints in the continuous optimal control problem are enforced at a specially chosen set of collocation points. This approximation of the continuous optimal control problem leads to a finite-dimensional nonlinear programming problem (NLP)[1], and the NLP is solved using well known software [2, 3]. Originally, direct collocation methods were developed as h methods (for example, Euler or Runge-Kutta methods) where the time interval is divided into a mesh and the state is approximated using the same fixed-degree polynomial in each mesh

∗Ph.D. Student, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250. Email: gatoreide@ufl.edu
†Distinguished Professor, Department of Mathematics, University of Florida, Gainesville, FL 32611-6250. Fellow, Society for Industrial and Applied Mathematics. E-mail: hager@ufl.edu.
‡Professor, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250. Erich Farber Faculty Fellow and University Term Professor. Associate Fellow, AIAA. E-mail: anilvrao@ufl.edu. Corresponding Author.
interval. Convergence in an $h$ method is then achieved by increasing the number and placement of the mesh points \[1, 4, 5\]. More recently, a great deal of research has been done in the class of direct Gaussian quadrature orthogonal collocation methods \[6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17\]. In a Gaussian quadrature collocation method, the state is typically approximated using a Lagrange polynomial where the support points of the Lagrange polynomial are chosen to be points associated with a Gaussian quadrature. Originally, Gaussian quadrature collocation methods were implemented as $p$ methods using a single interval. Convergence of the $p$ method was then achieved by increasing the degree of the polynomial approximation. For problems whose solutions are smooth and well-behaved, a Gaussian quadrature collocation method has a simple structure and converges at an exponential rate \[18, 19, 20\]. The most well developed Gaussian quadrature methods are those that employ either Legendre-Gauss (LG) points \[10, 21\], Legendre-Gauss-Radau (LGR) points \[11, 12, 13, 22\], or Legendre-Gauss-Lobatto (LGL) points \[6\]. In addition, a convergence theory has recently been developed using Gaussian quadrature collocation. Research on this theory has demonstrated that, under certain assumptions of the smoothness and coercivity, an $hp$ Gaussian quadrature method that employs either LG or LGR collocation points converges to a local minimizer of the optimal control problem \[23, 24, 25, 26, 27\].

While Gaussian quadrature orthogonal collocation methods are well suited to solving optimal control problems whose solutions are smooth, it is often the case that the solution of an optimal control problem has a nonsmooth optimal control \[28\]. The difficulty in solving problems with nonsmooth control lies in determining when the nonsmoothness occurs. For example, dynamical systems where the control appears linearly or problems that have state inequality path constraints often have solutions where the control and state may be nonsmooth. One approach to handling nonsmoothness is to employ a mesh refinement method where the optimal control problem is partitioned into a mesh and a mesh that meets a specified solution accuracy tolerance is obtained iteratively. In the context of Gaussian quadrature collocation, $hp$-adaptive mesh refinement methods \[14, 17, 29, 30, 31\] have been developed in order to improve accuracy in a wide variety of optimal control problems including those whose solutions are nonsmooth. It is noted, however, that mesh refinement methods often place an unnecessarily large number of collocation points and mesh intervals near points of nonsmoothness in the solution. Thus, it is beneficial to develop techniques that take advantage of the rapid convergence of a Gaussian quadrature collocation methods in segments where the solution is smooth and only increase the size of the mesh when necessary (thus, maintaining a smaller mesh than might be possible with a standard mesh refinement approach).

For optimal control problems where the solution is nonsmooth the convergence theory developed in Refs. \[23, 24, 25, 26, 27\] is not applicable. Consequently, when the solution of an optimal control problem is nonsmooth, an $hp$ method may not converge to a local minimizer of the optimal control problem. A well studied class of problems where the smoothness and coercivity conditions found in Ref. \[24\] are not met are those where the control appears linearly in the problem formulation \[28, 32, 33, 34\]. One approach for estimating the location of nonsmoothness is to introduce a variable called a break point \[35\] that defines the location of nonsmoothness and to include this variable in the NLP. The key problem that arises by
introducing a break point is that the NLP has an extra degree of freedom. As a result, the NLP may converge to a solution where this additional variable does not correspond to the location of the nonsmoothness. Ref. [36] also developed the concept of a knot using Legendre-Gauss-Lobatto collocation by introducing a variable that defines the switch time and collocating the dynamics at both the end of a mesh interval and the start of the subsequent mesh interval. However, the LGL method used in in Ref. [36] employs a square and singular differentiation matrix. Therefore, unlike the approach of Ref. [35], which used Legendre-Gauss collocation, the scheme used in Ref. [36] is not a Gauss quadrature integrator.

The objective of this research is to develop a new method that employs Gaussian quadrature collocation and accurately approximates the solution of an optimal control problem whose solution is nonsmooth by letting the location of the nonsmoothness be a free variable in the problem. In this paper, an approach is developed to improve upon the approach originally developed in Ref. [35] by gaining a better understanding why an incorrect location of the nonsmoothness in the optimal control is obtained when solving an optimal control problem using Legendre-Gauss-Radau collocation. Specifically, it is shown in this paper that the incorrect nonsmoothness location is obtained due to Lavrentiev phenomenon [37]. Lavrentiev phenomenon occurs in a practical situation when it is desired to minimize a numerical approximation of a continuous (functional) optimization problem. Whenever a numerical approximation of a functional leads to a minimizer that is strictly greater than or less than the true minimizer of the functional, the continuous optimization problem may be subject to Lavrentiev phenomenon [38, 39, 40, 41]. Simple examples of optimization problems that possess Lavrentiev phenomenon are given in Ref. [42], and the concept of Lavrentiev phenomenon has been extended to optimal control through the Lavrentiev gap [32]. The reason that the approximation of the continuous optimization problem has a higher or lower optimal objective arises from the possibility that the space over which the numerical optimization is performed may be different from the space over which the optimization needs to be performed in order to converge to the optimal solution. Therefore, the existence and the behavior of Lavrentiev phenomenon depends upon the choice of the approximation method. Moreover, any numerical scheme that gives rise to Lavrentiev phenomenon must somehow be augmented to compensate for any errors caused by the Lavrentiev phenomenon itself. Initial explorations of Lavrentiev phenomenon using Gaussian quadrature collocation methods have been provided in Refs. [39, 40, 41]. In order to properly account for Lavrentiev phenomenon it is first necessary to understand the circumstances in which it occurs for any given numerical scheme.

It is important to note that the approach developed in this paper is fundamentally different from the approaches developed in Refs. [35] and [36]. The key difference between the approach of this paper and that of Ref. [35] is that the search space is modified to include collocation constraints on the differential equations that are a function of control whereas the approach of Ref. [35] introduces no such additional collocation constraints. Moreover, the key difference between the approach of this paper and the work of Ref. [36] is that the work of Ref. [36] collocates all of the differential equations at the end of a mesh interval where a solution may be nonsmooth whereas in this work collocation constraints are included at the end of a mesh interval on only those differential equations that are a function of control. Second, the method
of Ref. [36] uses Legendre-Gauss-Lobatto which employs a square and singular differentiation matrix. The approach developed in this paper employs Legendre-Gauss-Radau collocation where the differentiation matrix is rectangular and has been shown previously to be a Gaussian quadrature integrator [22].

This paper presents a new method for Gaussian quadrature collocation. In this new method, the standard LGR method is modified to include additional variables and additional constraints at the end of a mesh interval when compared with a previously developed standard Legendre-Gauss-Radau collocation method. The additional variables are the time associated with mesh interval intersections and the value of the control at the end of every mesh interval. The additional constraints are a collocation conditions on those differential equations that are a function of the control and inequality constraints on the control at the endpoint of each mesh interval. It is important to note that the additional constraints are added to only those collocation constraints associated with the differential equations that are functions of the control and are not added to all differential equations. The modified method results in a different control variable at the end of each mesh interval from the control variable at the start of the next mesh interval. A costate estimation method is then developed that transforms the Lagrange multipliers of the NLP to the costate of the optimal control problem [12, 13, 22]. Using this costate estimation method, the transformed adjoint system [12, 13, 22, 43] of the modified LGR collocation method is developed. It is also shown that the state and control obtained from the modified LGR method along with the new costate estimation scheme satisfies one of the necessary Weierstrass-Erdmann conditions when the solution of the optimal control problem is nonsmooth.

This paper is organized as follows: Section 2 provides a brief introduction to solving optimal control problems using an LGR collocation method. Section 3 examines a motivating example to demonstrate the difficulties of solving optimal control problems with nonsmooth solutions using LGR methods. Section 3.2 provides a brief introduction to Lavrentiev phenomenon and examines the polynomial search space of the LGR discretization scheme for an example problem. Section 5 introduces the modified LGR method. The motivating example studied in Section 3 is revisited using the newly developed method to show the improvement in locating the nonsmoothness in the numerical approximation using the modified LGR method. Section 7 introduces a method to transform the Lagrange multipliers of the NLP to the costates of the continuous optimal control problem. Section 8 provides an analysis of the Weierstrass-Erdmann conditions using the method of this paper. Section 9 provides an example that demonstrates the accuracy of the method. Section 10 provides a comparison with other work that focuses on solving optimal control problems with nonsmooth solutions using collocation at Gaussian quadrature points. Finally, Section 11 provides conclusions on this work.

## 2 Legendre-Gauss-Radau Collocation

This paper focuses on second-order controlled dynamical systems of the form \( \ddot{x}(\tau) = f(x(\tau), \dot{x}(\tau), u(\tau)) \). Such a form is quite broad in applicability in that it arises frequently in mechanical systems (Newton-Euler
or Lagrangian mechanics). With such a class of dynamical systems as the focus, consider the following optimal control problem defined on \( \tau \in [-1, +1] \). Minimize the objective functional

\[
J = M(x(-1), v(-1), x(+1), v(+1), t_0, t_f) + \frac{t_f - t_0}{2} \int_{-1}^{+1} L(x(\tau), v(\tau), u(\tau)) d\tau,
\]

subject to the dynamic constraints

\[
\begin{align*}
\dot{x}(\tau) &= \frac{t_f - t_0}{2} v(\tau), \\
\dot{v}(\tau) &= \frac{t_f - t_0}{2} f(x(\tau), v(\tau), u(\tau)),
\end{align*}
\]

inequality path constraints

\[ c(x(\tau), v(\tau), u(\tau)) \leq 0, \]

and boundary conditions

\[ b(x(-1), v(-1), x(+1), v(+1), t_0, t_f) = 0, \]

where \((x(\tau), v(\tau)) \in \mathbb{R}^{2n}\) is the state (such that \(x(\tau) \in \mathbb{R}^n\) and \(v(\tau) \in \mathbb{R}^n\)), \(u(\tau) \in \mathbb{R}^m\) is the control, \(f: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n\), \(c: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\), \(b: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\), \(M: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}\), and \(L: \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}\). For convenience with the mathematical development that follows, all vector quantities are treated as row vectors. For example, \(x(\tau)\) and \(u(\tau)\) are defined as row vectors, respectively, as

\[
\begin{align*}
x(\tau) &= \begin{bmatrix} x_1(\tau) & \cdots & x_n(\tau) \end{bmatrix} \in \mathbb{R}^n, \\
u(\tau) &= \begin{bmatrix} u_1(\tau) & \cdots & u_m(\tau) \end{bmatrix} \in \mathbb{R}^m.
\end{align*}
\]

All other vector quantities are defined in a similar manner to that shown for \(x(\tau)\) and \(u(\tau)\) given in Eq. (5).

Suppose now that the state \((x(\tau), v(\tau))\) is approximated by a polynomial of degree at most \(N\). Let \(\ell_i \ (i = 1, \ldots, N + 1)\) be a basis of Lagrange polynomials given by

\[
\ell_i(\tau) = \prod_{j=1 \atop j \neq i}^{N+1} \frac{\tau - \tau_j}{\tau_i - \tau_j}, \quad (i = 1, \ldots, N + 1).
\]

The \(j^{th}\) component of \(x(\tau)\) and \(v(\tau)\) are then approximated in terms of the Lagrange polynomial basis as

\[
\begin{align*}
x_j(\tau) &\approx X_j(\tau) = \sum_{i=1}^{N+1} X_{ij} \ell_i(\tau), \\
v_j(\tau) &\approx V_j(\tau) = \sum_{i=1}^{N+1} V_{ij} \ell_i(\tau),
\end{align*}
\]

Differentiating \(x_j(\tau)\) and \(v_j(\tau)\) in Eq. (5) and evaluating the result at \(\tau = \tau_k\) gives

\[
\begin{align*}
\dot{x}_j(\tau_k) &\approx \dot{X}_j(\tau_k) = \sum_{i=1}^{N+1} X_{ij} \dot{\ell}_i(\tau_k) = \sum_{i=1}^{N+1} D_{ik} X_{ij}, \\
\dot{v}_j(\tau_k) &\approx \dot{V}_j(\tau_k) = \sum_{i=1}^{N+1} V_{ij} \dot{\ell}_i(\tau_k) = \sum_{i=1}^{N+1} D_{ik} V_{ij}.
\end{align*}
\]

The coefficients \(D_{ik} \ (i = 1, \ldots, N; \ k = 1, \ldots, N + 1\) form the \(N \times (N + 1)\) matrix \(D\) called the LGR differentiation matrix. For convenience \(D\) is partitioned as

\[
D = \begin{bmatrix} D_1 & D_2 & \cdots & D_{N+1} \end{bmatrix} = \begin{bmatrix} D_{1:N} & D_{N+1} \end{bmatrix},
\]
where $D_i$ denotes the $i^{th}$ column of $D$, $D_{1:N} \in \mathbb{R}^{N \times N}$ is an $N \times N$ matrix formed from the first $N$ columns of $D$, and $D_{N+1}$ is the last column of $D$. Thus, unlike the state and control, which are treated as row vectors at an instant of time, in this exposition the differentiation matrix is dealt with column-wise. Using the row vector convention for the state and control, the matrices $X \in \mathbb{R}^{(N+1) \times n}$ and $V \in \mathbb{R}^{(N+1) \times n}$ correspond row-wise to the state approximations at times $(\tau_1, \ldots, \tau_{N+1})$, while the matrix $U \in \mathbb{R}^{N \times m}$ corresponds row-wise to the approximations of the control at times $(\tau_1, \ldots, \tau_N)$. The matrices $X$, $V$, and $U$ are then given, respectively, as

$$X = \begin{bmatrix} X_1 \\ \vdots \\ X_{N+1} \end{bmatrix} \equiv X_{1:N+1},$$

$$V = \begin{bmatrix} V_1 \\ \vdots \\ V_{N+1} \end{bmatrix} \equiv V_{1:N+1},$$

$$U = \begin{bmatrix} U_1 \\ \vdots \\ U_N \end{bmatrix} \equiv U_{1:N},$$

where the notation $Y_{i:j}$ denotes generically rows $i$ through $j$ of the matrix $Y$. Also, the derivative approximations $\dot{X}(\tau)$ and $\dot{V}(\tau)$ at the $k^{th}$ LGR point $\tau_k$ are then given as row vectors, respectively, as

$$\dot{X}(\tau_k) = [DX]_k, \quad \dot{V}(\tau_k) = [DV]_k.$$  

It is noted that the state approximation is exact if the state is a polynomial of degree at most $N$. The LGR approximation of the state leads to the following nonlinear programming problem (NLP) that approximates the optimal control problem given in Eqs. (1)–(4):

$$\text{minimize } J = M(X_1, V_1, X_{N+1}, V_{N+1}, t_0, t_f) + \frac{t_f - t_0}{2} \sum_{i=1}^{N} w_i L(X_i, V_i, U_i),$$  

subject to

$$DX - \frac{t_f - t_0}{2} V_{1:N} = 0,$$

$$DV - \frac{t_f - t_0}{2} f(X_{1:N}, V_{1:N}, U_{1:N}) = 0,$$

$$c(X_{1:N}, V_{1:N}, U_{1:N}) \leq 0,$$

$$b(X_1, V_1, X_{N+1}, V_{N+1}, t_0, t_f) \leq 0,$$

where $w_i, \quad (i = 1, \ldots, N)$ are the LGR quadrature weights (and produce an exact integral if the integrand is a polynomial of degree at most $2N - 2$). Equations (11) and (12) will be referred to as the Legendre-Gauss-Radau collocation method.
3 Standard LGR Collocation When Solution is Nonsmooth

The convergence theory developed for the LGR collocation method described in Section 2, developed in Ref. [44], is applicable only to optimal control problems whose solutions are smooth. It is noted, however, that the convergence theory of Ref. [44] is not applicable to optimal control problems whose solutions are nonsmooth. In this section, LGR collocation will be applied to an example whose solution has a single point of nonsmoothness. The nonsmoothness will be accounted for by dividing the problem into two mesh intervals and including an additional decision variable that defines the mesh interval intersection and additional control at the end of each mesh interval. This example will demonstrate the key issue of Lavrentiev phenomenon [38, 42] that arises from using the LGR collocation method on a problem whose solution is nonsmooth, thus motivating the need to develop a modified LGR collocation method for optimal control problems whose solutions are nonsmooth. In particular, it will be shown how Lavrentiev phenomenon manifests itself when solving an optimal control problem with a nonsmooth solution using LGR collocation.

3.1 Lavrentiev Phenomenon

Lavrentiev phenomenon has been described using the following example [38, 42]: Let

$$ J(u) = \int_0^1 (u^3 - x)^2 \left( \frac{du}{dt} \right) dx $$

be an objective functional to be minimized. Let $u \in A$ where $A$ is the space of all absolutely continuous functions on the interval $[0, 1]$ and $(u(0), u(1)) = (0, 1)$. It has been shown by Maniá [42] that

$$ \inf_{u \in A \cap W} J(u) > \inf_{u \in A} J(u) = 0, $$

where $W$ is the space of all Lipschitz continuous functions on the interval $[0, 1]$. Equation (14) states that a function $u$ from the space of absolutely continuous results in an objective that is strictly less than than any minimizer $u$ that is from the space of Lipschitz continuous functions. Such an occurrence is known as Lavrentiev phenomenon [37]. Lavrentiev phenomenon is a concern when a method to numerically solve an optimization problem does not search over the correct space. Over the years, the definition of Lavrentiev phenomenon has expanded to include problems that do not only involve the space of absolutely continuous and Lipschitz continuous functions [45]. For instance, Guerra [32] examined the space of a singular arc optimal control problem against the space of the optimal control problem created when the singular problem is regularized.

In this paper, the search space of the standard LGR method is compared against the search space of the continuous time problem. In the next section, the search space of the standard LGR method will be defined. Then the search space will be constructed for the example problem presented in Section [5].
3.2 Lavrentiev Gap

In order to examine how Lavrentiev phenomenon manifests in the standard LGR collocation method the search space of the NLP and the search space of the continuous time optimal control problem must be analyzed. Let $\mathcal{P}_N$ be the space of all polynomials of degree $N$ on the interval $\tau \in [-1, +1]$. Furthermore, let $\mathcal{A} \subset \mathcal{P}_N$ be the set of all polynomials of degree $N$ that satisfies the collocation constraints of (12) at each LGR point $(\tau_1, \ldots, \tau_N)$. Next, let $\mathcal{U}^p$ be the set of all control functions such that produce a state approximation that lies in $\mathcal{A}$. Note that any state approximation that arises from a control in $\mathcal{U}^p$ satisfies the collocation constraints of Eq. (12) at only the LGR points. Let $\mathcal{Y} \subset \mathcal{P}_N$ be the set of polynomials such that any polynomial $\mathcal{Y}$ lies in a neighborhood of a solution to the continuous optimal control problem given in Eqs. (1)–(4). Finally, let $\mathcal{U}$ be the set of controls such that any element in $\mathcal{U}$ produces a state that lies in $\mathcal{Y}$. Because $\mathcal{Y}$ is a set of polynomials that lie in a neighborhood of an optimal solution, any state in $\mathcal{Y}$ must also reside in $\mathcal{A}$ (that is, $\mathcal{Y} \subset \mathcal{A}$), while any control that lies in $\mathcal{U}$ must also lie in $\mathcal{U}^p$ (that is, $\mathcal{U} \subset \mathcal{U}^p$). Suppose now that $u^* \in \mathcal{U}$ and $U^* \in \mathcal{U}^p$ are the optimal controls obtained when solving the LGR NLP with allowable search spaces $\mathcal{U}$ and $\mathcal{U}^p$, respectively. Furthermore, let $\mathcal{J}_{u^*}$ and $\mathcal{J}_{U^*}$ be the values of the objective obtained with $u^*$ and $U^*$, respectively. If $\mathcal{J}_{U^*} < \mathcal{J}_{u^*}$, then, because $\mathcal{U} \subset \mathcal{U}^p$, the solution obtained solving the LGR NLP with the allowable search space $\mathcal{U}^p$ exhibits Lavrentiev phenomenon [37, 38] and the optimal control problem possess a Lavrentiev gap [32] defined by $\mathcal{U}^p - \mathcal{U}$. Figure 1 illustrates the Lavrentiev gap.

It has been shown in Refs. [23, 24, 25, 26, 27] that under conditions of smoothness and coercivity, a Gaussian quadrature direct LGR collocation will converge to a local minimizer of the continuous optimal control problem. A locally minimizing solution may not, however, be obtained when the problem does not satisfy such coercivity conditions (for example, a problem with a nonsmooth optimal control). In such a situation, the solution of the LGR NLP may have a lower objective from the optimal objective of the continuous optimal control problem. Now an example is introduced to demonstrate how Lavrentiev gap manifests when solving an optimal control problem using the standard LGR collocation method and allowing the NLP solver to determine the location where the nonsmoothness occurs.
3.3 Motivating Example

Consider the following optimal control problem [14]. Minimize the objective functional

\[ J = t_f \]

subject to the dynamic constraints

\[ \dot{x}(\tau) = \frac{t_f}{2} v(\tau) \quad , \quad \dot{v}(\tau) = \frac{t_f}{2} u(\tau), \]

the inequality path constraints

\[ (0, -10, -1) \leq (x(\tau), v(\tau), u(\tau)) \leq (\infty, 10, +1), \]

and the boundary conditions

\[ (x(-1), x(+1), v(-1), v(+1)) = (10, 0, 0, 0). \]

The optimal solution to the optimal control problem given in Eqs. (15)–(18) is

\[ x^*(\tau) = \frac{t_f}{2} \begin{cases} \frac{x_0 - x_0}{2} (\tau + 1)^2, & -1 \leq \tau \leq \tau^*_s, \\ \frac{x_0}{2} (\tau - 1)^2, & \tau^*_s \leq \tau \leq +1, \end{cases} \]

\[ v^*(\tau) = \frac{t_f}{2} \begin{cases} -\sqrt{x_0} (\tau + 1), & -1 \leq \tau \leq \tau^*_s, \\ +\sqrt{x_0} (\tau - 1), & \tau^*_s \leq \tau \leq +1, \end{cases} \]

\[ u^*(\tau) = \begin{cases} -1, & -1 \leq \tau \leq \tau^*_s, \\ +1, & \tau^*_s \leq \tau \leq +1, \end{cases} \]

where \( \tau^*_s = 0 \) and \( t^*_f = 2\sqrt{x_0} \approx 6.32456 \). It is seen that the \( x^*(\tau) \) trajectory given in Eq. (19) is piecewise quadratic with a single switch in the optimal control. Thus, it should be possible to obtain the exact solution to the problem given in Eqs. (15)–(18) by dividing the time interval into two subintervals as follows: Minimize the objective functional

\[ J = t_f, \]

subject to the dynamic constraints in each interval \( k \in [1, 2], \)

\[ \dot{x}^{(k)}(\tau) = a^{(k)} \frac{t_f}{2} v^{(k)}(\tau) \quad , \quad \dot{v}^{(k)}(\tau) = a^{(k)} \frac{t_f}{2} u^{(k)}(\tau), \]

the inequality path constraints in each interval \( k \in [1, 2], \)

\[ (0, -10, -1) \leq (x^{(k)}(\tau), v^{(k)}(\tau), u^{(k)}(\tau)) \leq (\infty, 10, +1), \]

and the boundary conditions

\[ (x^{(1)}(-1), x^{(2)}(+1), v^{(1)}(-1), v^{(2)}(+1)) = (10, 0, 0, 0), \]
where \( \alpha^{(1)} = (t_s - t_0)/(t_f - t_0) \) and \( \alpha^{(2)} = (t_f - t_s)/(t_f - t_0) \). Suppose now that the LGR collocation method is used to approximate the two-interval optimal control problem of Eqs. (20)–(23). Because the optimal trajectory is piecewise quadratic and the LGR quadrature is exact for polynomials of degree at most \( 2N - 2 \), it should be possible to obtain an exact solution using two collocation points in each subinterval (that is, \( N^{(1)} = N^{(2)} = 2 \)) with \( t_s \) included as a variable in the optimization. Furthermore, the control function, known as the \emph{approximate control}, can be obtained using \( \dot{v}(\tau) \) as
\[
u(\tau) = \dot{v}(\tau).
\)

Figure 2 shows the NLP control and the approximate control obtained by solving the two-interval standard LGR NLP. First, it is seen that the NLP solver returns a switch time in the control that differs significantly from the optimal switch time. In addition, the optimal objective returned by the NLP solver is approximately 6.0 which is \emph{less} than the known optimal objective \( 2\sqrt{x_0} \approx 6.32456 \). Finally, the approximate control given by Eq. (24) exceeds the upper limit, \( u_{\text{max}} \), given in Eq. (22) and, as a result, the NLP returns an approximate optimal control solution that is not a member of the admissible set of solutions for the original continuous optimal control problem described in Eqs. (15)–(18). Consequently, adding the switch time of the control as a variable results in a solution with a lower objective and an incorrect switch time, thus making it the case that the allowable search space in the two-interval problem is larger than what should be permissible under the continuous time constraints. Such a disparity of the search space and objective between the discretized problem and the continuous time problem is an example of \emph{Lavrentiev phenomenon} [45].

As previously stated by Ferriero [45], the way in which Lavrentiev phenomenon manifests itself is dependent on the numerical method utilized to solve the problem. Therefore, the search space of any particular numerical method must be understood in order to effectively close the Lavrentiev gap. In the next
section the search space of the standard LGR method for the example problem solved in the previous subsection will be analyzed.

4 Search Space Using Standard LGR Collocation

The search space of the continuous time optimal control problem from Section 2 is a collection of functions that satisfy the constraints from Eqs. (2)–(4). The search space of the NLP described in Section 2 is a collection of discrete points that satisfy the constraints of Eq. (12). However, for this research it is necessary to compare this discrete space with the continuous functions that constitute the search space of the continuous time optimal control problem. A continuous search space is constructed from the discrete space of the NLP by using the Lagrange polynomial state approximation to determine a continuous time control approximation. In this section, the two interval, two collocation point discretization from Section 3.3 is used to demonstrate how a continuous time representation of a discrete solution can be constructed, and how some of the possible solutions violate constraints of the continuous time problem.

The differentiation matrix $D$ of the optimal control problem of Section 3 using the chosen two-interval two-collocation-point LGR approximation is given as

$$
D = \begin{bmatrix}
D_{11}^{(1)} & D_{12}^{(1)} & D_{13}^{(1)} & 0 & 0 \\
D_{21}^{(1)} & D_{22}^{(1)} & D_{23}^{(1)} & 0 & 0 \\
0 & 0 & D_{11}^{(2)} & D_{12}^{(2)} & D_{13}^{(2)} \\
0 & 0 & D_{21}^{(2)} & D_{22}^{(2)} & D_{23}^{(2)} \\
\end{bmatrix}.
$$

(25)

Furthermore, the collocation constraints for the two-interval approximation of the dynamics given in Eq. (16) are given as

$$
\begin{bmatrix}
D & 0 \\
0 & D \\
\end{bmatrix}
\begin{bmatrix}
X \\
V \\
\end{bmatrix} - \frac{t_f}{2}
\begin{bmatrix}
V_{1:4} \\
U \\
\end{bmatrix} = 0,
$$

(26)

where $V_{1:4}$ is the column vector formed using the first four rows of the column vector $V$. Now, solving Eq. (16) gives $x(t_f) = \int_{t_0}^{t_f} v(t) + x_0$. Suppose now that $v(\tau)$ is approximated as a Lagrange polynomial of degree two [as given in Eq. (6)] in each mesh interval. Given that the boundary conditions are fixed values, suppose now that $V_2$, $V_3$, and $V_4$ are defined to be the coefficients of the Lagrange polynomial approximation of $v(\tau)$ at the following points, respectively: (1) the second LGR point in the first mesh interval (that is, the first interior LGR point in the first mesh interval); (2) the non-collocated point at the end of the first mesh interval (which is the same coefficient as that at the first point of the second interval); and (3) the second LGR point in the second mesh interval (that is, the first interior LGR point in the second mesh interval). Because $v(\tau)$ is approximated as a piecewise quadratic, varying these three coefficients results in polynomial approximations for the integral of $v(\tau)$ [that is, $x(\tau)$], and the derivative of $v(\tau)$ [that is $\dot{v}(\tau)$] along with a value for $t_f$. A feasible solution of the two-interval LGR NLP that approximates the optimal control problem defined in Eqs. (15–18) for any given values of $V_2$, $V_3$, $V_4$ is calculated by solving the
following linear system of four equations for the unknowns $X_2$, $X_3$, $X_4$, and $t_f$.

$$
\begin{bmatrix}
D_{12}^{(1)} & D_{13}^{(1)} & 0 \\
D_{22}^{(1)} & D_{23}^{(1)} & 0 \\
0 & D_{12}^{(2)} & D_{13}^{(2)} \\
0 & D_{22}^{(2)} & D_{23}^{(2)}
\end{bmatrix}
\begin{bmatrix}
X_2 \\
X_3 \\
X_4 \\
t_f
\end{bmatrix}
= 
\begin{bmatrix}
-D_{11}^{(1)}x_1 - 0.5V_1 \\
-D_{12}^{(1)}x_2 - 0.5V_2 \\
-D_{11}^{(2)}x_5 - 0.5V_3 \\
-D_{12}^{(2)}x_5 - 0.5V_4
\end{bmatrix}.
$$

(27)

To understand how the allowable values of the polynomial approximations of the state compare to feasible solutions of the corresponding continuous optimal control problem, it is necessary to analyze the associated polynomial approximation of the state. Let $(X^{(k)}(\tau), V^{(k)}(\tau))$ be Lagrange polynomial approximations of the state in mesh interval $k$ that satisfy the constraints of Eqs. (16)–(18) at the points defined by the LGR approximation (that is, the state bounds in Eqs. (17)–(18) are satisfied at the collocation points and any non-collocated points while the control bounds in Eqs. (17)–(18) are satisfied at all collocation points). Figure 3 shows all possible control functions, that is, all functions (which in this case are polynomials because the state approximation is a polynomial and the dynamics of the example in Section 3 are linear in the control) of the form

$$
U^{(k)}(\tau) = \dot{V}^{(k)}(\tau),
$$

(28)

that arise from state approximations $(X^{(k)}(\tau), V^{(k)}(\tau))$ that satisfy the constraints of the LGR NLP. It is seen from Fig. 3 that the possible control function that satisfy the LGR NLP constraints violate the bounds on the control as given in Eq. (18). Consequently, the set of possible solutions of the LGR discrete approximation produce control approximations that are infeasible with respect to the continuous constraints given in Eqs. (17)–(18). The goal of the next section is to modify the search space for the standard LGR method by introducing constraints into the NLP that allow the admissible solutions of the NLP to more closely represent the feasible solutions of the continuous time problem.

![Figure 3: Possible approximate control functions for the example given in Eqs. (20)–(23).](image)

In order to modify the search space in the LGR collocation method, suppose now that the functions $(X^{(k)}(\tau), V^{(k)}(\tau))$ defined previously are restricted such that the only possible control functions $U^{(k)}(\tau)
are those such that the state and control approximations, \((X^{(k)}(\tau), V^{(k)}(\tau), \text{and } U^{(k)}(\tau))\), are feasible with respect to the control bounds \((u_{\min}, u_{\max})\) given in Eq. (22). In other words, the only possible state approximations are those that are feasible with respect to the control bounds and simultaneously satisfy all other constraints in the LGR NLP. Using Eq. (28), a search space different from that of the standard LGR collocation method can now be constructed that provides those control function approximations that lead to state approximations \((X^{(k)}(\tau), V^{(k)}(\tau))\) that are feasible with respect to the constraints of the continuous time optimal control problem.

5 Modification of Standard LGR Method

Using the results of Section 4, additional constraints are now augmented to the standard collocation method presented in Section 2 in order to improve the approximation of the location of the nonsmoothness in the solution to the optimal control problem (thereby improving the accuracy of the solution itself). In particular collocation constraints are added at the end of each mesh interval, but such constraints are added to only those differential equations that are a function of control. In this manner, and as stated in Section 1, the approach developed in this research differ fundamentally from the approaches developed in Refs. [35] and [36].

5.1 New Decision Variables

The modified LGR method introduces the following two new decision variables in mesh interval \(k\). The first new variable is the independent variable (in this case, time) associated with the intersection of the mesh intervals. This value of time is denoted \(T^{(k)}\), \((k = 1, \ldots, K - 1)\). The second new variable is the approximation of the control at the end of each mesh interval. The value of this control approximation is denoted \(U^{(k)}_{N_{k+1}}\), \((k = 1, \ldots, K)\). The portion of the decision vector associated with the control in the modified LGR collocation method is then defined as

\[
\tilde{U}^{(k)} = \begin{bmatrix}
U^{(k)} \\
V^{(k)}_{N_{k+1}}
\end{bmatrix}.
\]

It is important to note that \(U^{(k)}_{N_{k+1}}\) and \(U^{(k+1)}_1\) are both associated with \(T_k\). In other words, the time must be continuous across mesh intervals but the control need not be continuous. Now, as opposed to formulating the modified LGR collocation method in terms of the variable \(T_k\), it is convenient to use the quantity \(\alpha^{(k)}\) where \(\alpha^{(k)}\) is defined as

\[\alpha^{(k)} = \frac{T_{k+1} - T_k}{2}, \quad (k = 1, \ldots, K - 1).\]

Finally, note that while \(T_k\), \((k = 1, \ldots, K - 1)\), is a variable in the modified LGR collocation method but is not a variable in the standard LGR method. More specifically, in the standard LGR method each mesh point is static and is not determined when solving the NLP given in Eqs. (11) and (12). On the other hand, \(U^{(k)}_{N_{k+1}}\) is not defined in the standard LGR method.
5.2 New Constraints

Given that additional variables are included in the modified LGR collocation method, constraints must be added in order to properly modify the search space. In particular, in the modified LGR collocation method collocation constraints are added at the end of each mesh interval using only those differential equations that are a function of the control. In order to gain insight as to why collocation constraints are added to only those differential equations that are a function of the control, consider the first differential equation in the example of Section 3 that is, consider the dynamics $\dot{x}(\tau) = v(\tau)$, where $x(\tau)$ and $v(\tau)$ are the two components of the state. Next, let $P^N$ be the space of all polynomials of degree $N$ on the domain $\tau \in [-1, +1]$. Let $x(\tau) \in P^N$ and $v(\tau) \in P^N$. Finally, let $(\tau_1, \ldots, \tau_N)$ be the LGR points and let $\tau_{N+1} = +1$. Suppose now that both $X(\tau)$ and $V(\tau)$ are Lagrange polynomial approximations of degree $N$ as given in Eq. (29). Then, the function $\dot{X}(\tau) - V(\tau)$ is a polynomial of degree $N$ and, thus, has $N$ roots. Consequently, the only possible way that the constraint $\dot{X}(\tau) - V(\tau)$ will be satisfied at the $N + 1$ points $(\tau_1, \ldots, \tau_{N+1})$ is if a polynomial of degree $N$ has $N + 1$ zeros. Any $N$ degree polynomial that has $N + 1$ zeros requires the polynomial to be zero everywhere. Thus, enforcing the constraint $\dot{X}(\tau) - V(\tau) = 0$ at $N + 1$ points will lead to an over-determined system which is why collocation constraints are not added at $\tau = +1$ for a differential equation that is not a function of control.

Next, consider the second differential equation in the example of Section 3 that is, consider the dynamics $\dot{v}(\tau) = u(\tau)$, where $v(\tau)$ is the second component of the state and $u(\tau)$ is the control. Suppose again that the state approximation $V(\tau)$ of $v(\tau)$ is a polynomial of degree $N$ in each of the two mesh intervals of a two-interval formulation of the optimal control problem given in Section 3. Finally, suppose that the constraint $\dot{v}(\tau) = u(\tau)$ is enforced at the $N$ LGR points plus the final point of the first mesh interval. Because $V(\tau)$ is a polynomial of degree $N$ and the differential equation is a function of control, it is possible to satisfy all $N + 1$ conditions

$$\dot{V}(\tau_i) - U_i = 0, \quad (i = 1, \ldots, N + 1)$$

(29)

in the first mesh interval because the control is a variable in Eq. (29). In other words, $U_{N+1}$ can be varied in order to satisfy Eq. (29) at the endpoint of the first interval. Moreover, when adding this collocation condition, it is also necessary to add the constraint that $u_{\min} \leq U_{N+1} \leq u_{\max}$ in order to ensure that the control at the end of the first mesh interval satisfies the limits on the control.

The preceding argument leads to a modification of the LGR collocation method for the case where the solution may be nonsmooth. A collocation condition similar to that given in Eq. (29) is included along with a constraint that enforces all control bounds at the end of the mesh interval. This modification leads to formation of the modified LGR differentiation matrix

$$\tilde{D}^{(k)} = \begin{bmatrix} D_{1:Nk}^{(k)} & D_{Nk+1}^{(k)} \\ E^{(k)} & E_0^{(k)} \end{bmatrix},$$

(30)

in the mesh intervals where the additional collocation constraints are included for those differential equations that are a function of the control. It is noted in Eq. (30) that $[D_{1:N} \ D_{N+1}] \in \mathbb{R}^{N \times (N+1)}$ is the standard
The example of Section 3 is now revisited using the modified LGR collocation method. Figure 4 exhibits the impact of the additional collocation constraint from Eq. (31) has on the search space of the example problem. Fig. 4a demonstrates that each admissible set for control now falls between the allowable control limits \((u_{\text{min}}, u_{\text{max}}) = (-1, 1)\). Next, to examine the effect that the modified LGR method has on the solution of the NLP for the example in Section 3 Fig. 4b shows the objective of the modified LGR NLP as a function of the switch time, \(\tau_s\), where it is assumed that the switch time is fixed. At the optimal switch time \(\tau_s^*\), the objective of both the original and modified LGR methods is identical. Note, however, that when for \(\tau_s < \tau_s^*\), the optimal objective of the standard LGR method is smaller than the modified LGR method. In fact, Fig. 4b shows that the optimal objective for the modified LGR method occurs when \(\tau_s = \tau_s^*\). This last result indicates that the modified LGR method reduces the allowable search space such that the solution of the NLP leads to a state approximation that is closer to the solution of the continuous optimal control problem. Figure 5 shows the control solution obtained by solving for the control as a function of time using the Lagrange polynomial approximation of the state obtained using the modified LGR collocation method. It is seen that, not only does the control function lie within its allowable limits \((u_{\text{min}}, u_{\text{max}}) = (-1, 1)\), but the switch time obtained using the modified LGR collocation method matches the switch time of the solution of the continuous optimal control problem.
Figure 4: Admissible controls for modified LGR collocation method, comparison of optimal objective for both standard and modified LGR collocation methods as a function of switch time, $\tau_s$, and optimal control obtained using modified LGR collocation method.

Figure 5: Optimal control for the example defined by Eqs. (15)–(18) using the modified LGR method.

7 Transformed Adjoint System of Modified LGR Method

This section derives the adjoint system of the modified LGR collocation method based on the optimal control problem given in Eqs. (1)–(4) as described in Section 2. The first-order optimality conditions for the continuous time problem described in Eqs. (1)–(4) are given as

$$
\dot{\lambda} = -\frac{\partial L}{\partial x} - \lambda \left[ \frac{\partial f}{\partial x} \right]^T,
$$

$$
\dot{\tilde{\lambda}} = -\frac{\partial L}{\partial v} - \lambda - \tilde{\lambda} \left[ \frac{\partial f}{\partial v} \right]^T,
$$

(34) (35)
\begin{align}
0 &= \frac{\partial L}{\partial u} + \hat{\lambda} \frac{\partial f}{\partial u}, \quad \text{(36)} \\
\lambda(-1) &= -\frac{\partial M}{\partial x(-1)} + \psi \left[ \frac{\partial b}{\partial x(-1)} \right]^T, \quad \text{(37)} \\
\hat{\lambda}(-1) &= -\frac{\partial M}{\partial v(-1)} + \psi \left[ \frac{\partial b}{\partial v(-1)} \right]^T, \quad \text{(38)} \\
\lambda(+1) &= \frac{\partial M}{\partial x(+1)} - \psi \left[ \frac{\partial b}{\partial x(+1)} \right]^T, \quad \text{(39)} \\
\hat{\lambda}(+1) &= \frac{\partial M}{\partial v(+1)} - \psi \left[ \frac{\partial b}{\partial v(+1)} \right]^T, \quad \text{(40)}
\end{align}

where the gradient of a scalar function \( f : \mathbb{R}^n \rightarrow \mathbb{R} \) and the Jacobian of a vector function of a vector \( g : \mathbb{R}^n \rightarrow \mathbb{R}^m \) are defined, respectively, as

\[
\frac{\partial f}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n} \end{bmatrix}, \quad \frac{\partial \mathbf{g}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial g_1}{\partial x_1} \\ \vdots \\ \frac{\partial g_m}{\partial x_n} \end{bmatrix},
\]

where it is noted again that \( \mathbf{x} \) is a row vector (that is, \( \mathbf{x} = [x_1 \ \cdots \ x_n] \)) and that the function \( g(\mathbf{x}) \) is also a row vector (that is, \( g(\mathbf{x}) = [g_1(\mathbf{x}) \ \cdots \ g_m(\mathbf{x})] \)). The goal of this section is to derive the first-order optimality conditions, also known as the Karush-Kuhn-Tucker (KKT) conditions, of the modified LGR collocation method. Then, using these first-order optimality conditions, a transformation is derived that relates the dual variables of the modified LGR collocation method to the costates of the continuous optimal control problem.

First the NLP associated with the LGR collocation method as described in Section 2 will be re-written in a multiple-interval formulation [29] with the additional constraints and variables from Section 5. The time domain \( \tau \in [-1, +1] \) is now partitioned into \( K \) mesh intervals \([T_{k-1}, T_k], k = 1, \ldots, K\) such that \( T_0 = -1 \) and \( T_K = +1 \). Let \( t_i^{(k)} \in [T_{k-1}, T_k], i = 1, \ldots, N_k \) be the \( N_k \) LGR quadrature points on the interval \([T_{k-1}, T_k]\).

Additionally, state continuity is enforced across mesh interval such that \( X_{N_k+1}^{k-1} = X_1^{(k)} \) and \( V_{N_k+1}^{k-1} = V_1^{(k)} \). Now the multiple-interval NLP formulation of the modified LGR collocation method can be written as

\[
\begin{align}
\text{minimize } J &= M(X_i^{(k)}, V_i^{(k)}, X_{N_k+1}^{(k)}, V_{N_k+1}^{(k)}, t_0, t_f) + \sum_{k=1}^{K} \alpha^{(k)} \frac{t_f - t_0}{2} \sum_{i=1}^{N_k} w_i^{(k)} L(X_i^{(k)}, V_i^{(k)}, U_i^{(k)}) \\
\text{subject to } \quad D^{(k)} X^{(k)} - \alpha^{(k)} \frac{t_f - t_0}{2} V_{1:N_k} = 0, \\
\tilde{D}^{(k)} V^{(k)} - \alpha^{(k)} \frac{t_f - t_0}{2} b(X_i^{(k)}, V_i^{(k)}, U_i^{(k)}) = 0, \\
b(X_1^{(1)}, V_1^{(1)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) = 0, \\
\sum_{k=1}^{K} \alpha^{(k)} - 1 = 0,
\end{align}
\]
where $\tilde{D}^{(k)}$ is the modified LGR differentiation matrix as given in Eq. (30) and alpha is ratio of the change in time of the $k^{th}$ mesh interval to the change in time across the entire problem and is expressed as

$$\alpha^{(k)} = \frac{T_k - T_{k-1}}{2}.$$  \hfill (46)

Now the first-order optimality conditions of the discrete system described in Eqs. (41)–(44) are derived. First the augmented objective function is written as

$$J_a = \mathcal{M}(X_1^{(1)}, V_1^{(1)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) + \sum_{k=1}^{K} \alpha^{(k)} f_{tf} - t_0 \frac{2}{2} \sum_{i=1}^{N_k} \mathcal{L}(X_i^{(k)}, V_i^{(k)}, U_i^{(k)})$$

$$- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left( A_i^{(k)} \right) D_{i,1:N_k}^{(k)} X_i^{(k)} - \alpha^{(k)} f_{tf} - t_0 \frac{2}{2} V_i^{(k)}$$

$$- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left( \tilde{A}_i^{(k)} \right) D_{i,1:N_k+1}^{(k)} V_i^{(k)} - \alpha^{(k)} f_{tf} - t_0 \frac{2}{2} f(X_i^{(k)}, V_i^{(k)}, U_i^{(k)})$$

$$- \Psi b^T (X_1^{(1)}, V_1^{(1)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) - \beta \left( \sum_{k=1}^{K} \alpha^{(k)} - 1 \right),$$  \hfill (47)

where $A^{(k)} \in \mathbb{R}^{N_k \times n}, \tilde{A}^{(k)} \in \mathbb{R}^{(N_k+1) \times n}, \Psi \in \mathbb{R}^b, \beta \in \mathbb{R}$ and $\langle \cdot, \cdot \rangle$ denotes the standard inner product between two vectors. Rewriting Eq. (47) so that the final row of the state matrix is separated from the first $N_k$ rows gives

$$J_a = \mathcal{M}(X_1^{(k)}, V_1^{(k)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) + \sum_{k=1}^{K} \alpha^{(k)} f_{tf} - t_0 \frac{2}{2} \sum_{i=1}^{N_k} \mathcal{L}(X_i^{(k)}, V_i^{(k)}, U_i^{(k)})$$

$$- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left( A_i^{(k)} \right) D_{i,1:N_k}^{(k)} X_i^{(k)} + D_{i,N_k+1}^{(k)} X_i^{(k)} - \alpha^{(k)} f_{tf} - t_0 \frac{2}{2} V_i^{(k)}$$

$$- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left( \tilde{A}_i^{(k)} \right) D_{i,1:N_k}^{(k)} V_i^{(k)} + D_{i,N_k+1}^{(k)} V_i^{(k)} - \alpha^{(k)} f_{tf} - t_0 \frac{2}{2} f(X_i^{(k)}, V_i^{(k)}, U_i^{(k)})$$

$$- \sum_{k=1}^{K} \left( A_{N_k+1}^{(k)} \right) E_{0}^{(k)} V_{N_k+1}^{(k)} + E_{1}^{(k)} V_{N_k+1}^{(k)} + \sum_{k=1}^{K} \left( A_{N_k+1}^{(k)} \right) f_{tf} - t_0 \frac{2}{2} f(X_{N_k+1}^{(k)}, V_{N_k+1}^{(k)}, U_{N_k+1}^{(k)})$$

$$- \Psi b^T (X_1^{(1)}, V_1^{(1)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) - \beta \left( \sum_{k=1}^{K} \alpha^{(k)} - 1 \right),$$  \hfill (48)

Next, the following theorem is introduced that will allow the terms involving $f(X_{N_k+1}^{(k)}, V_{N_k+1}^{(k)}, U_{N_k+1}^{(k)})$, and $E^{(k)}$ in Eq. (48) to be written as functions of $X_{1:N_k}^{(k)}, V_{1:N_k}^{(k)}, X_{N_k+1}^{(k)}$, and $D^{(k)}_{N_k+1}$.

**Theorem 1.** Consider a polynomial $f(\tau)$ on the interval $\tau \in [-1, 1]$ that is of degree at most $N - 1$. Let $\tau \in \mathbb{R}^{(N+1)}$ such that $(\tau_1, \ldots, \tau_N)$ are the Legendre-Gauss-Radau points on $\tau \in [-1, 1]$ and $\tau_{N+1} = +1$. If the Lagrange basis polynomial associated with $\tau_{N+1} = +1$ is given as

$$\ell_{N+1}(\tau) = \prod_{1 \leq j \leq N} \frac{\tau - \tau_j}{\tau_{N+1} - \tau_j},$$  \hfill (49)
then

\[
\int_{-1}^{+1} f(\tau) \dot{\ell}_{N+1}(\tau) d\tau = f(+1).
\]  

(50)

Proof. Integrating the left-hand side of Eq. (50) by parts yields

\[
f(\tau) \ell_{N+1}(\tau) \bigg|_{-1}^{+1} - \int_{-1}^{+1} f(\tau) \dot{\ell}_{N+1}.\]

(51)

Because \( \dot{f}(\tau) \) is a polynomial of degree at most \( N - 2 \) and \( \ell_{N+1}(\tau) \) is a polynomial of at most degree \( N \), then the integrand in Eq. (51) is at most degree \( 2N - 2 \) and the integral can be evaluated using LGR quadrature as

\[
\int_{-1}^{+1} \dot{f}(\tau) \ell_{N+1}(\tau) d\tau = \sum_{i=1}^{N} w_i \dot{f}(\tau_i) \ell_{N+1}(\tau_i),
\]

(52)

where \( w_i \) is the \( i^{th} \) LGR quadrature weight. Recall that \( \ell_i(\tau_i) = 1 \) and \( \ell_i(\tau_j) = 0 \) when \( i \neq j \), then Eq. (52) is zero and Eq. (51) reduces to

\[
f(\tau) \ell_{N+1}(\tau) \bigg|_{-1}^{+1} = f(+1) \ell_{N+1}(+1)
\]

\[- f(-1) \ell_{N+1}(-1)
\]

\[= f(+1),\]

(53)

which completes the proof. \( \square \)

Equation (50) allows the vector \( E^{(k)} \) of \( \dot{D}^{(k)} \) to be related to \( D^{(k)}_{N+1} \) of \( \dot{D}^{(k)} \). The elements of \( E^{(k)} \) are defined as

\[
E^{(k)}_j = \dot{\ell}_j(+1), \quad (j = 1, \ldots, N).
\]

(54)

Letting \( \dot{\ell}_j(\tau) = f(\tau) \) from Equation (50), \( E^{(k)}_j \) can then be expressed as

\[
E^{(k)}_j = \int_{-1}^{+1} \dot{\ell}_j(\tau) \dot{\ell}_{N+1}(\tau) d\tau \quad (j = 1, \ldots, N).
\]

(55)

Because \( \dot{\ell}_j(\tau) \dot{\ell}_{N+1}(\tau) \) is a polynomial of degree most \( 2N_k - 2 \), Eq. (55) can be replaced exactly with an LGR quadrature as

\[
E^{(k)}_j = \sum_{i=1}^{N_k} w^{(k)}_i \dot{\ell}_j(\tau^{(k)}_i) \dot{\ell}_{N+1}(\tau^{(k)}_i), \quad (j = 1, \ldots, N_k).
\]

(56)

Note that \( \dot{\ell}_{N_k+1}(\tau^{(k)}_i) \) is the \( i^{th} \) element of \( D^{(k)}_{N+1} \). Using the definition of the \( \dot{D}^{(k)} \) matrix and the relationship from Eq. (50) gives

\[
D^{(k)T}_{N+1} W^{(k)} D^{(k)}_{1:N_k} = E^{(k)},
\]

(57)

\[
D^{(k)T}_{N+1} W^{(k)} f^{(k)}_{1:N_k} = f^{(k)}_{N+1},
\]

(58)

where \( W^{(k)} \in \mathbb{R}^{N_k \times N_k} \) is defined as

\[
W^{(k)} = \begin{bmatrix}
  w^{(k)}_1 & 0 & \ldots & 0 \\
  0 & w^{(k)}_2 & \ldots & 0 \\
  \vdots & \ddots & \ddots & \vdots \\
  0 & \ldots & 0 & w^{(k)}_{N_k}
\end{bmatrix},
\]

(59)
and \( w_i^{(k)} \), \( i = 1, \ldots, N_k \) are the LGR quadrature weights. Equations (57–58) allows Eq. (48) to be rewritten as

\[
J_a = \mathcal{M}(X_1^{(1)}, V_1^{(1)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) + \frac{K}{2} \sum_{k=1}^{N_k} \alpha^{(k)} \sum_{i=1}^{N_k} w_i^{(k)} \mathcal{L}(X_i^{(k)}, V_i^{(k)}, U_i^{(k)})
\]

\[
- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left\langle A_i^{(k)}, \left( D_{i,N_k+1}^{(k)} X_{1:N_k}^{(k)} + D_{i,N_k+1}^{(k)} X_{N_k+1}^{(k)} - \alpha_i^{(k)} \frac{t_f - t_0}{2} V_i^{(k)} \right) \right\rangle 
\]

\[
- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left\langle A_i^{(k)}, \left( D_{1:N_k}^{(k)} V_{1:N_k}^{(k)} - \alpha_i^{(k)} \frac{t_f - t_0}{2} f(X_i^{(k)}, V_i^{(k)}, U_i^{(k)}) \right) \right\rangle 
\]

\[
- \sum_{k=1}^{K} \sum_{i=1}^{N_k} \left\langle A_i^{(k)}, D_{i,N_k+1}^{(k)} V_{N_k+1}^{(k)} \right\rangle 
\]

\[
- \sum_{k=1}^{K} \left\langle A_{N_k+1}^{(k)} D_{N_k+1}^{(k)} W^{(k)} D_{1:N_k}^{(k)} V_{1:N_k}^{(k)} + E_0^{(k)} V_i^{(k)} \right\rangle 
\]

\[
+ \sum_{k=1}^{K} \left\langle A_{N_k+1}^{(k)} + \alpha_i^{(k)} \frac{t_f - t_0}{2} \left( D_{N_k+1}^{(k)} W^{(k)} f(X_i^{(k)}, V_i^{(k)}, U_i^{(k)}) \right) \right\rangle 
\]

\[- \Psi b^T(X_1^{(1)}, V_1^{(1)}, X_{N_k+1}^{(K)}, V_{N_k+1}^{(K)}, t_0, t_f) \]

\[- \beta \left( \sum_{k=1}^{K} \alpha^{(k)} - 1 \right). \]

To simplify the following derivation, the function arguments will be removed from the equations. For instance, \( \mathcal{L}(X^{(k)}, V^{(k)}, U^{(k)}) \) will be expressed as \( \mathcal{L}^{(k)} \). The KKT conditions are derived by taking the partial derivatives \( J_a \) with respect to \( X^{(k)}, V^{(k)}, U^{(k)}, A^{(k)}, \tilde{A}^{(k)}, \Psi^{(k)} \), \( t_0, t_f \) and \( \alpha^{(k)} \) and setting them equal to zero. Note that Eqs. (42–44) arise from setting \( \partial J_a / \partial A^{(k)} \), \( \partial J_a / \partial \tilde{A}^{(k)} \), \( \partial J_a / \partial \Psi \) respectively to zero and the remaining derivatives are written as

\[
D_i^{(k)^T} A^{(k)} = \alpha^{(k)} \frac{t_f - t_0}{2} \nabla x_i^{(k)} \left( w_i^{(k)} \mathcal{L}_i^{(k)} + \left\langle \tilde{A}_i^{(k)} + \tilde{A}_{N_k+1}^{(k)} D_{i,N_k+1}^{(k)} w_i^{(k)}, f_i^{(k)} \right\rangle \right) 
\]

\[- \delta_{1i}(-\nabla x_i^{(1)} M + \nabla x_i^{(0)} \Psi b^T), \]

\[
D_{N_k+1}^{(k)^T} A = \nabla x_{N_k+1}^{(k)} M - \nabla x_{N_k+1}^{(k)} \Psi b^T, \]

\[
D_i^{(k)^T} \left( \tilde{A}_{1:N_k}^{(k)} + A_{N_k+1}^{(k)} D_{i,N_k+1}^{(k)} w_i^{(k)} \right) = \alpha^{(k)} \frac{t_f - t_0}{2} \nabla v_i^{(k)} \left( w_i^{(k)} \mathcal{L}_i^{(k)} + \left\langle \tilde{A}_i^{(k)} + \tilde{A}_{N_k+1}^{(k)} D_{i,N_k+1}^{(k)} w_i^{(k)}, f_i^{(k)} \right\rangle \right) 
\]

\[- \delta_{1i}(-\nabla v_i^{(1)} M + \nabla v_i^{(1)} \Psi b^T), \]

\[
D_{N_k+1}^{(k)^T} A_{1:N_k}^{(k)} + E_0^{(k)} A_{N_k+1}^{(k)} = \nabla v_{N_k+1}^{(k)} M - \nabla v_{N_k+1}^{(k)} \Psi b^T, \]

\[
0 = \alpha^{(k)} \frac{t_f - t_0}{2} \nabla u_i^{(k)} \left( w_i^{(k)} \mathcal{L}_i^{(k)} - \left\langle A_{1:N_k}^{(k)} + A_{N_k+1}^{(k)} D_{i,N_k+1}^{(k)} w_i^{(k)}, f_{1:N_k}^{(k)} \right\rangle \right), \]

\((k = 1, \ldots, K; i = 1, \ldots, N_k), \)
\[ 0 = \sum_{k=1}^{K} -\alpha(k) \sum_{i=1}^{N_k} w_i^{(k)} L_i^{(k)} + \sum_{i=1}^{N_k} \left\langle \Lambda_i^{(k)}, \frac{-\alpha(k)}{2} \mathbf{V}_i \right\rangle + \sum_{i=1}^{N_k} \left\langle \tilde{\Lambda}_i^{(k)}, \frac{-\alpha(k)}{2} \mathbf{r}_i \right\rangle \]  

\[ + \left\langle \tilde{\Lambda}_{N_k+1}^{(k)}, \frac{-\alpha(k)}{2} \left( \mathbf{D}_{N_k+1}^{(k)T} \mathbf{W}^{(k)} \mathbf{f}^{(k)} \right) \right\rangle + \nabla_{t_0} \left( \mathbf{M} - \Psi_b^T \right), \]  

\[ 0 = \sum_{k=1}^{K} \alpha(k) \sum_{i=1}^{N_k} w_i^{(k)} L_i^{(k)} + \sum_{i=1}^{N_k} \left\langle \Lambda_i^{(k)}, \alpha(k) \mathbf{V}_i \right\rangle + \sum_{i=1}^{N_k} \left\langle \tilde{\Lambda}_i^{(k)}, \alpha(k) \mathbf{r}_i \right\rangle \]  

\[ + \left\langle \tilde{\Lambda}_{N_k+1}^{(k)}, \alpha(k) \left( \mathbf{D}_{N_k+1}^{(k)T} \mathbf{W}^{(k)} \mathbf{f}^{(k)} \right) \right\rangle + \nabla_{t_f} \left( \mathbf{M} - \Psi_b^T \right), \]  

\[ 0 = \frac{t_f - t_0}{2} \sum_{i=1}^{N_k} w_i^{(k)} L_i^{(k)} + \sum_{i=1}^{N_k} \left\langle \Lambda_i^{(k)}, \frac{t_f - t_0}{2} \mathbf{V}_i \right\rangle + \sum_{i=1}^{N_k} \left\langle \tilde{\Lambda}_i^{(k)}, \frac{t_f - t_0}{2} \mathbf{r}_i \right\rangle \]  

\[ + \left\langle \tilde{\Lambda}_{N_k+1}^{(k)}, \frac{t_f - t_0}{2} \left( \mathbf{D}_{N_k+1}^{(k)T} \mathbf{W}^{(k)} \mathbf{f}^{(k)} \right) \right\rangle - \beta \quad (k = 1, \ldots, K), \]  

where \( \delta_{ij} \) is the Kronecker delta function defined as

\[ \delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \]  

The KKT conditions given in equation (68) are unique to the modified LGR method and is not required for an extremal solution of the standard LGR NLP transcription. Now propose the change of variables

\[ \lambda_i^{(k)} = \frac{\Lambda_i^{(k)}}{w_i^{(k)}}, \quad (70) \]

\[ \lambda_{N_k+1}^{(k)} = \mathbf{D}_{N_k+1}^{(k)T} \Lambda_1:N_k, \quad (71) \]

\[ \psi_i = \Psi_i, \quad (72) \]

\[ \tilde{\lambda}_i^{(k)} = \frac{\tilde{\Lambda}_i^{(k)}}{w_i^{(k)}} + \tilde{\Lambda}_{N_k+1}^{(k)} D_{i,N_k+1}, \quad (73) \]

\[ \tilde{\lambda}_{N_k+1}^{(k)} = \mathbf{D}_{N_k+1}^{(k)T} \tilde{\Lambda}_1:N_k + \tilde{\Lambda}_{N_k+1}^{(k)} E_0^{(k)}, \quad (74) \]

Note that Eqs. (70)–(72) are the same transformations used for the standard LGR method. Finally, define \( \mathbf{D}^{(k)\dagger} \in R^{N_k \times N_k} \) such that

\[ D_{ij}^{(k)\dagger} = -D_{11}^{(k)} - \frac{1}{w_1^{(k)}} \quad (75) \]

\[ D_{ij}^{(k)\dagger} = \frac{w_i^{(k)}}{w_1^{(k)}} D_{ji}^{(k)} \quad \text{otherwise}, \quad (76) \]

for \( i = j = 1, 2, \ldots, N_k \). Note that \( \mathbf{D}^{(k)\dagger} \) is the same matrix derived by Refs. [12, 13] where it was shown that \( \mathbf{D}^{(k)\dagger} \) is the differentiation matrix for the space of polynomials of degree at most \( N_k - 1 \). Now the KKT conditions can be rewritten as

\[ \mathbf{D}_i^{(k)\dagger} \lambda_{1:N_k}^{(k)} = -\alpha(k) \frac{t_f - t_0}{2} \nabla_{x_i^{(k)}} \left( \left\langle \tilde{\Lambda}_i^{(k)}, \mathbf{r}_i^{(k)} \right\rangle + \mathbf{r}_i^{(k)} \right) \]  

\[ + \frac{\delta_{1i}}{w_1^{(1)}} \left( -\nabla_{X_1^{(1)}} \left( \mathbf{M} - \Psi_b^T \right) - \lambda_{1}^{(1)} \right), \]  

(77)
\[ D^{(k)}_i \tilde{\lambda}^{(k)}_{1:N_k} = -\alpha^{(k)} \frac{t_f - t_0}{2} \nabla V^{(k)}_{i,N_k} \left( \left\langle \tilde{\lambda}^{(k)}_i, \tilde{r}^{(k)}_i \right\rangle + \mathcal{L}^{(k)}_i \right) - \lambda^{(k)}_i \]

\[ + \frac{\delta i}{w_1} \left( -\nabla V^{(i)}_{i} \left( \mathcal{M} - \psi \beta \right) - \tilde{\lambda}^{(1)}_1 \right), \]

\[ 0 = \alpha^{(k)} \frac{t_f - t_0}{2} \nabla U^{(k)}_{i,N_k} \left( \mathcal{L}^{(k)}_i - \left\langle \tilde{\lambda}^{(k)}_i, \tilde{r}^{(k)}_i \right\rangle \right), \]

\[(i = 1, \ldots, N_k, \quad k = 1, \ldots, K), \]

\[ \lambda^{(K)}_{N_k+1} = \nabla x^{(K)}_{N_k+1} \left( \mathcal{M} - \psi \beta \right), \]

\[ \lambda^{(K)}_{N_k+1} = \nabla v^{(K)}_{N_k+1} \left( \mathcal{M} - \psi \beta \right), \]

\[-\nabla t_0 \left( \mathcal{M} - \psi \beta \right) = \sum_{k=1}^{K} -\alpha^{(k)} \sum_{i=1}^{N_k} H_i^{(k)} w_i^{(k)}, \]

\[-\nabla t_f \left( \mathcal{M} - \psi \beta \right) = \sum_{k=1}^{K} \alpha^{(k)} \sum_{i=1}^{N_k} H_i^{(k)} w_i^{(k)}, \]

where \( H_i^{(k)} = \mathcal{L}_i^{(k)} + \lambda_i^{(k)} v_i^{(k)} + \tilde{\lambda}_i^{(k)} f_i^{(k)} \) is the approximation the Hamiltonian, \( \mathcal{H} \) in interval \( k \). Equations (37)–(38) allow the terms in the second lines of Eqs. (77)–(78) to vanish which results in Eqs. (77)–(83) becoming discrete representations of the continuous time first-order optimality conditions from Eqs. (34)–(40).

### 8 Weierstrass-Erdmann Conditions

If the optimal control is discontinuous, additional optimality conditions called the Weierstrass-Erdmann conditions [28] must be satisfied. One of the Weierstrass-Erdmann conditions states that the Hamiltonian must be continuous at the location of a control discontinuity. The Hamiltonian for the optimal control problem defined in Eqs. (20)–(23) can be approximated as

\[ \mathcal{H}^{(k)}(\tau_i^{(k)}) \approx H_i^{(k)} = \mathcal{L}_i^{(k)} + \lambda_i^{(k)} v_i^{(k)} + \tilde{\lambda}_i^{(k)} f_i^{(k)}, \]

where \( \tau_i^{(k)} \in [T_{k-1}, T_k], \quad i = 1, \ldots, N_k \) and \( k = 1, \ldots, K \) are the \( N \) LGR points in the \( k \)th mesh interval. The Weierstrass-Erdmann condition on the Hamiltonian can be written as [28]

\[ \mathcal{H}(\tau_s^+) = \mathcal{H}(\tau_s^-), \]

where \( \tau_s^+ \) is the switch time approached by the right side and \( \tau_s^- \) is the switch time approached by the left side.

The analysis that follows will demonstrate that the transformed adjoint system of the modified LGR collocation method satisfies a discrete representation of the Weierstrass-Erdmann condition given in Eq. (85). First, the transformations given in Eqs. (70)–(74) together with the definition of the Hamiltonian given in Eq. (84), Eq. (68) simplifies to

\[ \beta = \frac{t_f - t_0}{2} \sum_{i=1}^{N_k} w_i^{(k)} H_i^{(k)}, \quad (k = 1, \ldots, K), \]
where $\beta$ is the Lagrange multiplier defined in Eq. (47) associated with the constraint given in Eq. (33). Next, multiplying Eq. (86) by $\alpha^{(k)}$ gives

$$\alpha^{(k)} \beta = \frac{t_f - t_0}{2} \alpha^{(k)} \sum_{i=1}^{N_k} w_j^{(k)} H_i^{(k)}, \quad (k = 1, \ldots, K).$$

(87)

Then, because $H$ is not an explicit function of time it follows that $H_i^{(k)}$ is constant in each mesh interval. Moreover, the right-hand side of Eq. (87) is LGR quadrature approximation of the integral of the Hamiltonian over the interval $[T_{k-1}, T_k]$. Consequently, using the definition of $\alpha^{(k)}$ from Eq. (46), Eq. (87) can be rewritten as

$$\frac{\Delta T_k}{t_f - t_0} \beta = H^{(k)} \Delta T_k, \quad (k = 1, \ldots, K),$$

(88)

where $\Delta T_k = T_k - T_{k-1}$. Equation (88) then reduces to

$$\frac{\beta}{t_f - t_0} = H^{(k)}, \quad (k = 1, \ldots, K).$$

(89)

The implication of Eq. (89) is that the Hamiltonian must be the same value in each mesh interval. Therefore, Eq. (88) can only be satisfied if the Hamiltonian is constant on the time interval $[-1, 1]$. The transformed adjoint system of the standard LGR collocation method adjoint mapping scheme requires only that the Hamiltonian is constant within a mesh interval, but does not require that the Hamiltonian be constant across the entire time interval. On the other hand, the modified LGR collocation mesh ensures that the Hamiltonian is constant across the entire time interval. The following section provides an example that demonstrates the accuracy of the costate estimation method developed in Section 7 and compares the results of the modified LGR collocation method with the results obtained using the standard LGR collocation method.

9 Example

In this section the costate estimation method for the modified LGR collocation method is demonstrated on the example problem given in Eqs. (15)–(18) Section 3. For comparison, the exact switch point was hard coded into the standard LGR method. The dual variables returned by the NLP solver are shown in Fig. 6a and 6b. Figure 6a shows that the dual variables returned for $\dot{x}$ approximation are exactly the same. Figure 6b shows a difference in the dual variables associated with the approximation $\dot{V}$ of $\dot{v}$, with the two dual variables of the modified LGR method located at the switch time ($\tau = 0$) arising from the additional collocation conditions associated with those differential equations that are a function of the control.

Figures 7a and 7b shows the costate approximations obtained using the standard LGR method and the modified LGR method. Both methods return the correct value for $\lambda(t)$. Note, however, that the estimate for $\dot{\lambda}(t)$ is not correct when the standard LGR method is implemented with the switch time fixed at its exact value. The fact that the approximation of $\dot{\lambda}(t)$ is incorrect when using the exact switch time in the standard LGR method implies that the location of the switch time computed by the standard LGR method will also be incorrect.
(a) Dual variable, $\Lambda$, for example problem using both the standard and modified LGR collocation methods.

(b) Dual variable, $\tilde{\Lambda}$, for example problem using both the standard and modified LGR collocation methods.

Figure 6: Costate estimates for the example problem using both the standard and modified LGR collocation methods.

Figures 8a and 8b demonstrate further the problem when using the standard LGR method when the switch time fixed at its exact value. While the integral from $-1$ to $+1$ in Fig. 8a is correct and each interval has a continuous and constant Hamiltonian, the integral from $-1$ to $\tau_s$ and from $\tau_s$ to $+1$ is incorrect. The clear discontinuity in the Hamiltonian from Fig. 8a shows that the Weierstrass-Erdmann conditions from Eq. (85) are not satisfied by the standard LGR method. Furthermore, the discontinuity in Fig. 8a is a result of the incorrect costate that is returned from the standard LGR method as seen in Fig. 7b. Figure 7b shows that the $\tilde{\lambda}(\tau_s) \neq 0$ for the standard LGR method, so not only are the Weierstrass-Erdmann conditions not satisfied, but neither are the standard necessary conditions for optimality. Figure 8b demonstrates that the additional constraint from Eq. (89) enforces continuity throughout the Hamiltonian thus satisfying the Weierstrass-Erdmann conditions of Eq. (85).

10 Comparison with Methods of Refs. [35] and [36]

In this section the modified LGR method developed in this paper is compared against the methods developed in Refs. [35] and [36]. In particular, the methods of Refs. [35] and [36] are used as the basis of this comparison because these methods also use collocation at Gaussian quadrature points and were developed for solving optimal control problems whose solutions are nonsmooth. Section 10.1 provides a comparison of the method of this paper with the work of Ref. [35], while Section 10.2 provides a comparison of the method of this paper with the work of Ref. [36].
Figure 7: Costate estimates for the example problem using both the standard and modified LGR collocation methods.

Figure 8: Hamiltonian for both the standard and modified LGR collocation method.

10.1 Comparison with Method of Ref. [35]

Reference [35] presents a method that employs collocation at Legendre-Gauss (LG) points, where the LG points include neither the initial point nor the terminal point of a mesh interval. Specifically, the method of Ref. [35] divides the time interval into multiple domains called super-elements where each super-element is a collocation of mesh intervals. Then, a variable that defines the time point at the junction between two adjacent super-elements is introduced. The new variable is then treated as an additional optimization parameter and is determined in the process of solving the NLP on the given super-element mesh. The idea
behind this approach is that the new variable provides an estimate of the location of any nonsmoothness in the solution. Similar to the method of Ref. [35], the method presented in this paper also introduces a new variable that is designed to identify the location of nonsmoothness in the solution. It is noted, however, that the method of this paper is fundamentally different from that of Ref. [35] in that the method of this paper also introduces a new control variable at the end of a mesh interval whereas the method of Ref. [35] does not include a new control variable. In particular, Section 5 shows that adding a variable that defines the control at the end of a mesh interval closes the Lavrentiev gap and produces a solution such that the continuous control approximation lies within the control limits.

10.2 Comparison with Method of Ref. [36]

Reference [36] presents a method that employs collocation at Legendre-Gauss-Lobatto (LGL) points, where the LGL points include both the initial and terminal point of a mesh interval. Then, in a manner similar to that of Ref. [35], the method of Ref. [36] divides the time interval into segments, performs LGL collocation within each segment, and introduces a variable that defines the location of a possible discontinuity in the control. Similar to the method of Ref. [36], the modified LGR method of this paper also collocates the dynamics at both the initial and terminal points of a mesh interval. The approach of this paper, however, differs fundamentally in several aspects from the method of Ref. [36]. In particular, in the method of Ref. [36] the collocation point at the end of a mesh interval is one of the LGL quadrature points. On the other hand, in the method of this paper the collocation point at the end of a mesh interval is not a quadrature point. Next, in the method of Ref. [36] collocation is performed on all of the differential equations. In the method of this paper, however, collocation is performed at the end of a mesh interval only on those differential equations that depend upon the control. Next, because the method of Ref. [36] employs collocation at LGL points using a square and singular differentiation matrix, the method of Ref. [36] is not a Gauss quadrature integrator. In the method of this paper, however, the matrix formed from the first \( N_k \) rows and columns \( 2, \ldots, N_k + 1 \) of the matrix \( \tilde{D} \) [see Eq. (31)] can be inverted to produce the Legendre-Gauss-Radau integration matrix [12]. As a result, similar to the standard LGR method [12], the method of this paper is also a Gauss quadrature integrator [22]. Second, the singular differentiation matrix employed in Ref. [36] leads to a transformed adjoint system that contains a nonzero null space with an oscillatory behavior [22], and it was shown that this nonzero null space leads to a costate estimate itself that may be inaccurate (see the example at the end of Ref. [22]). On the other hand, as derived in Section 7 the method of this paper leads to a transformed adjoint system that does not have a null space and produces an accurate costate estimate when the solution is nonsmooth. Finally, it was shown in Section 8 that the method of this paper satisfies the Weierstrass-Erdmann conditions.
11 Conclusions

A new method has been developed for solving optimal control problems whose solutions are nonsmooth. The standard LGR collocation method has been modified to include two variables and two constraints at the end of a mesh interval. These new variables are the time associated with the intersection of mesh intervals and the value of the control at the end of the each mesh interval. The two additional constraints are a collocation condition on each differential equation that is a function of control and an inequality constraint on the control at the endpoint of each mesh interval. These additional constraints modify the search space of the nonlinear programming problem such that an accurate approximation to the location of the nonsmoothness is obtained. A transformation of the Lagrange multipliers of the NLP to the costate of the optimal control problem has then been developed and the resulting transformed adjoint system of the modified Legendre-Gauss-Radau method has then been derived. Furthermore, it has been shown that the costate estimate satisfies the Weierstrass-Erdmann optimality conditions. Finally, the method developed in this paper has been demonstrated on an example whose solution is nonsmooth.

Acknowledgments

Support for this research from the U.S. Office of Naval Research under grants N00014-15-1-2048 and N00014-19-1-2543 and from the U.S. National Science Foundation under grants CBET-1404767, DMS-1522629, DMS-1819002, and CMMI-1563225, is gratefully acknowledged.

References

[1] Betts, J. T., Practical Methods for Optimal Control and Estimation Using Nonlinear Programming, SIAM Press, Philadelphia, 2010.

[2] Gill, P. E., Murray, W., and Saunders, M. A., “SNOPT: An SQP algorithm for large-scale constrained optimization,” SIAM review, Vol. 47, No. 1, 2005, pp. 99–131.

[3] Biegler, L. T. and Zavala, V. M., “Large-Scale Nonlinear Programming Using IPOPT: An Integrating Framework for Enterprise-Wide Optimization,” Computers and Chemical Engineering, Vol. 33, No. 3, March 2008, pp. 575–582.

[4] Jain, D. and Tsitras, P., “Trajectory Optimization Using Multiresolution Techniques,” Journal of Guidance, Control, and Dynamics, Vol. 31, No. 5, September-October 2008, pp. 1424–1436.

[5] Zhao, Y. and Tsitras, P., “Density Functions for Mesh Refinement in Numerical Optimal Control,” Journal of Guidance, Control, and Dynamics, Vol. 34, No. 1, January–February 2011, pp. 271–277.

[6] Elnagar, G., Kazemi, M. A., 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.

[7] Elnagar, G. N. and Razzaghi, M., “Short communication: A collocationtype method for linear quadratic optimal control problems,” Optimal Control Applications and Methods, Vol. 18, No. 3, 1997, pp. 227–235.

[8] 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.
[9] Fahroo, F. and Ross, I. M., “Direct Trajectory Optimization by a Chebyshev Pseudospectral Method,” *Journal of Guidance, Control, and Dynamics*, Vol. 25, No. 1, 2002, pp. 160–166.

[10] 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, 2006, pp. 1435–1440.

[11] 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.

[12] 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*, Vol. 49, No. 2, June 2011, pp. 335–358.

[13] Garg, D., Hager, W. W., and Rao, A. V., “Pseudospectral methods for solving infinite-horizon optimal control problems,” *Automatica*, Vol. 47, No. 4, 2011, pp. 829–837.

[14] 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*, Vol. 32, No. 4, July–August 2011, pp. 476–502.

[15] Darby, C. L., Hager, W. W., and Rao, A. V., “Direct Trajectory Optimization Using a Variable Low-Order Adaptive Pseudospectral Method,” *Journal of Spacecraft and Rockets*, Vol. 48, No. 3, May–June 2011, pp. 433–445.

[16] Francolin, C. C., Hager, W. W., and Rao, A. V., “Costate Approximation in Optimal Control Using Integral Gaussian Quadrature Collocation Methods,” *Optimal Control Applications and Methods*, Vol. 36, No. 4, July–August 2015, pp. 381–397.

[17] Patterson, M. A., Hager, W. W., and Rao, A. V., “A ph Mesh Refinement Method for Optimal Control,” *Optimal Control Applications and Methods*, Vol. 36, No. 4, July–August 2015, pp. 398–421.

[18] Canuto, C., Hussaini, M. Y., Quarteroni, A. M., and Thomas Jr, A., *Spectral methods in fluid dynamics*, Springer Science & Business Media, 2012.

[19] Fornberg, B., *A practical guide to pseudospectral methods*, Vol. 1, Cambridge university press, 1998.

[20] Trefethen, L. N., *Spectral methods in MATLAB*, Vol. 10, Siam, 2000.

[21] Rao, A. V., Benson, D. A., Darby, C., Patterson, M. A., Francolin, C., 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 (TOMS)*, Vol. 37, No. 2, 2010, pp. 22.

[22] 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, November 2010, pp. 1843–1851.

[23] Hager, W. W., Hou, H., and Rao, A. V., “Convergence Rate for a Radau Collocation Method Applied to Unconstrained Optimal Control,” 2015, arXiv.org/abs/1508.03783.

[24] Hager, W. W., Hou, H., and Rao, A. V., “Convergence Rate for a Gauss Collocation Method Applied to Unconstrained Optimal Control,” *Journal of Optimization Theory and Applications*, Vol. 169, No. 3, June 2016, pp. 801 – 824.

[25] Hager, W. W., Hou, H., and Rao, A. V., “Lebesgue Constants Arising in a Class of Collocation Methods,” *IMA Journal of Numerical Analysis*, Vol. 37, No. 4, October 2017, pp. 1884–1901.
[26] Hager, W. W., Liu, J., Mohapatra, S., Rao, A. V., and Wang, X.-S., “Convergence rate for a Gauss collocation method applied to constrained optimal control,” *SIAM Journal on Control and Optimization*, Vol. 56, 2018, pp. 1386–1411.

[27] Hager, W. W., Hou, H., Mohapatra, S., Rao, A. V., and Wang, X.-S., “Convergence rate for a Radau $hp$-collocation method applied to constrained optimal control,” *Computational Optimization and Applications*, Vol. 74, 2019, pp. 274–314.

[28] Bryson, A. E. and Ho, Y., *Applied Optimal Control: Optimization, Estimation, and Control*, Hemisphere Publishing Corporation, 1975.

[29] Darby, C. L., *HP-Pseudospectral Method for Solving Continuous-Time Nonlinear Optimal Control Problems*, Ph.D. thesis, University of Florida, 2011.

[30] Liu, F., Hager, W. W., and Rao, A. V., “Adaptive Mesh Refinement for Optimal Control Using Nonsmoothness Detection and Mesh Size Reduction,” *Journal of the Franklin Institute*, Vol. 352, No. 10, 2015, pp. 4081–4106.

[31] Liu, F., Hager, W. W., and Rao, A. V., “Adaptive mesh refinement method for optimal control using decay rates of Legendre polynomial coefficients,” *IEEE Transactions on Control Systems Technology*, Vol. 26, No. 4, 2018, pp. 1475–1483.

[32] Guerra, M. and Sarychev, A., “Measuring singularity of generalized minimizers for control-affine problems,” *Journal of Dynamical and Control Systems*, Vol. 15, No. 2, 2009, pp. 177–221.

[33] Li, S., Zhao, R., and Zhang, Q., “Optimization method for solving bang-bang and singular control problems,” *Journal of Control Theory and Applications*, Vol. 10, No. 4, 2012, pp. 559–564.

[34] Martinon, P., Bonnans, F., Laurent-Varin, J., and Trélat, E., “Numerical Study of Optimal Trajectories with Singular Arcs for an Ariane 5 Launcher,” *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 1, 2009, pp. 51–55.

[35] Cuthrell, J. E. and Biegler, L. T., “Simultaneous Optimization and Solution Methods for Batch Reactor Control Profiles,” *Computers and Chemical Engineering*, Vol. 13, No. 1–2, January–February 1987, pp. 49–62.

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

[37] Lavrentiev, M., “MA Lavrent’ev,” *Sur quelques problemes du calcul des variations*, 1927, pp. 7–28.

[38] Ball, J. M. and Knowles, G., “A numerical method for detecting singular minimizers,” *Numerische Mathematik*, Vol. 51, No. 2, 1987, pp. 181–197.

[39] Eide, J. and Rao, A. V., “Lavrentiev Phenomenon in $hp$ Gaussian Quadrature Collocation Methods for Optimal Control,” 2016 *AIAA/AAS Astrodynamics Specialist Conference*, AIAA Paper 2016–5575, Long Beach, California, 13–16 September 2016.

[40] Eide, J. D., Hager, W. W., and Rao, A. V., “Modified Radau Collocation Method for Solving Optimal Control Problems with Nonsmooth Solutions. Part I: Lavrentiev Phenomenon and the Search Space,” 2018 *IEEE Conference on Decision and Control (CDC)*, 17–19 December 2018, pp. 1644–1650.

[41] Eide, J. D., Hager, W. W., and Rao, A. V., “Modified Radau Collocation Method for Solving Optimal Control Problems with Nonsmooth Solutions. Part II: Costate Estimation and the Transformed Adjoint System,” 2018 *IEEE Conference on Decision and Control (CDC)*, 17–19 December 2018, pp. 1651–1656.

[42] Manià, B., “Soppa un esempio di Lavrentieff,” *Boll. Unione Mat. Ital.*, Vol. 13, 1934, pp. 147–153.

[43] Hager, W. W., “Runge-Kutta Methods in Optimal Control and the Transformed Adjoint System,” *Numerische Mathematik*, Vol. 87, 2000, pp. 247–282.
[44] Hou, H., Hager, W. W., and Rao, A. V., “Convergence of a Gauss Pseudospectral Method for Optimal Control,” AIAA Guidance, Navigation, and Control Conference and Exhibit. Minnesota: American Institute of Aeronautics and Astronautics, Vol. 8, 2012, pp. 1–9.

[45] Ferriero, A., The Lavrentiev phenomenon in the Calculus of Variations, Phd thesis, Universita degli Studi di Milano-Bicocca, 2004.