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

Joseph D. Eide[1] · William W. Hager[2,3] · Anil V. Rao[1,4]

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. The two additional variables are the time at the interface between two mesh intervals and the control at the end of each mesh interval. The two additional constraints are a collocation condition for those differential equations that depend upon the control and an inequality constraint on the control at the endpoint of each mesh interval. The 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 one of the Weierstrass–Erdmann optimality conditions. Finally, the method developed in this paper is demonstrated on an example whose solution is nonsmooth.

Keywords Optimal control · Gaussian quadrature collocation · Lavrentiev phenomenon · Nonsmooth optimal control

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

Extended author information available on the last page of the article

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 interval. Convergence in an h method is then achieved by increasing the number of mesh intervals [1,4,5]. More recently, a great deal of research as been done in the class of direct *Gaussian quadrature orthogonal collocation* methods [6–13]. 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 [14–16]. The most well-developed Gaussian quadrature methods are those that employ either Legendre–Gauss (LG) points [17,18], Legendre–Gauss–Radau (LGR) points [19–22], or Legendre–Gauss–Lobatto (LGL) points [6]. In addition, a convergence theory has recently been developed using Gaussian quadrature collocation. Research on this theory had 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 [23–28]. In particular, it is shown in Refs. [23–28] that the convergence rate is exponentially fast as a function of the polynomial degree and is a polynomial function of the mesh interval width.

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 [29]. 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 discretized using a sequence of meshes such that the last mesh satisfies a specified solution accuracy tolerance. In the context of Gaussian quadrature collocation, hp mesh refinement methods [9,11–13,19–22,30] 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. [23–27] 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 [29,31–33]. One approach for estimating the location of nonsmoothness is to introduce a variable called a break point or knot [34,35] 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. This extra degree of freedom can be addressed by introducing additional constraints into the problem. Research performed in Ref. [36–38] introduced a bilevel or nested approach to solving nonsmooth optimal control problems. The bilevel approach formulates two separate NLP problems referred to as an inner problem and an outer problem. The inner problem is the transcribed optimal control problem, whereas the outer problem determines the properties of the mesh which is used by the inner problem and insures additional optimality conditions associated with nonsmooth optimal control problems are satisfied. Similar to the method in Ref. [35], the inner problem places explicit assumptions on the control functions. Reference [39] 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. A more detailed discussion that compares the method developed in this paper with the methods of Refs. [34–39] is provided in Sect. 8.

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. [34] 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 and introducing a constraint that will satisfy the equations of motion at the nonsmoothness but will not place explicit assumptions on the optimal control function. Specifically, it is shown in this paper that the incorrect nonsmoothness location is obtained due to Lavrentiev phenomenon [40]. Lavrentiev phenomenon occurs in a practical situation when it is desired to minimize a numerical approximation of a continuous (functional) optimization problem. In particular, a continuous optimization problem may be subject to Lavrentiev phenomenon whenever a numerical approximation of a functional leads to an optimal objective value that is either strictly greater than or strictly less than the optimal value of the functional [41–44]. Simple examples of optimization problems that possess Lavrentiev phenomenon are given in Ref. [45], and the concept of Lavrentiev phenomenon has been extended to optimal control through the Lavrentiev gap [31]. 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 are provided in Refs. [42–44]. 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. [34,39], and [36–38]. The key difference between the approach of this paper and that of Ref. [34] 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. [34] introduces no such additional collocation constraints. Next, the key difference between the approach of this paper and the work of Ref. [39] is that the work of Ref. [39] 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. [39] uses Legendre–Gauss–Lobatto which employs a square and singular differentiation matrix. On the other hand, the approach developed in this paper employs Legendre–Gauss–Radau collocation where the differentiation matrix is rectangular. Moreover, it has been shown previously that Legendre–Gauss–Radau is a Gaussian quadrature integration method [22]. Finally, the key difference between the method of this paper and the methods of Refs. [36–38] is that the methods of [36–38] parameterize the control as a function of time, and this parameterization is used to approximate the control at the end of each mesh interval. The method of this paper, however, introduces a variable that defines the control at the end of a mesh interval and adds collocation conditions at the end of the mesh interval using only those differential equations that are a function of the control.

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 boundaries and the corresponding value of the control at the end of the mesh interval. The additional constraints are 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 [20–22]. Using this costate estimation method, the transformed adjoint system [20–22,46] 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 and therefore does not require additional constraints to enforce the Weierstrass–Erdmann conditions. It is also noted that the method developed in this paper may be applicable in a wide range of aerospace applications. Such applications include optimal spacecraft orbital transfer, aircraft performance optimization, and hypersonic flight. Solutions to optimal control problems in these aerospace application areas often have nonsmooth solutions (for example, bang–bang thrust profiles or fast actuator dynamics in surface controllers on board an aircraft).

The remainder of this paper is organized as follows. Section 2 provides the notations and conventions used in this paper. Section 3 presents the Bolza optimal control problem. Section 4 presents the standard Legendre–Gauss–Radau (LGR) collocation method for discretizing optimal control problems. Section 5 provides a description of Lavrentiev phenomenon, a discussion of the Lavrentiev gap that arises when using LGR collocation to solve an optimal control problem whose solution is nonsmooth, and an analysis of the search space using LGR collocation. Section 6 presents the modified LGR collocation method for solving optimal control problems with nonsmooth solutions. Section 7 derives the transformed adjoint system along with the Weierstrass–Erdmann conditions that arise from modified LGR collocation and demonstrates the accuracy of the modified LGR collocation method costate estimate. Next, Sect. 8 provides a comparison of the method developed in this paper with the methods developed in Refs. [34–39]. Finally, Sect. 9 provides conclusions on this research.

2 Notation and Conventions

In this paper, the following notation and conventions will be used. First, the independent variable is denoted τ. Therefore, the notation $x(\tau)$ denotes a dependence of the quantity x on τ. Next, all vectors will be denoted as *row* vectors. Therefore, if $\mathbf{x}(\tau) \in \mathbb{R}^n$ is a vector function of τ, then $\mathbf{x}(\tau)$ is given as

$$\mathbf{x}(\tau) = [x_1(\tau), x_2(\tau), \ldots, x_n(\tau)]. \tag{1}$$

Suppose now that $\mathbf{x}(\tau)$ is approximated using a basis of Lagrange polynomials $\ell_j(\tau), \ (j = 1, \ldots, N+1)$ as

$$\mathbf{x}(\tau) \approx \hat{\mathbf{x}}(\tau) = \sum_{j=1}^{N+1} \mathbf{X}_j \ell_j(\tau), \quad \ell_j(\tau) = \prod_{\substack{l=1 \\ l \neq j}}^{N+1} \frac{\tau - \tau_l}{\tau_j - \tau_l}, \tag{2}$$

where $(\tau_1, \ldots, \tau_{N+1})$ are the support points of $\ell_j(\tau), \ (j = 1, \ldots, N+1)$. It is known that the Lagrange polynomials $\ell_j(\tau), \ (j = 1, \ldots, N+1)$ satisfy the property

$$\ell_j(\tau_i) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \tag{3}$$

which implies that

$$\mathbf{X}(\tau_i) = \mathbf{X}_i, \quad (i = 1, \ldots, N + 1). \tag{4}$$

Using aforementioned row vector conventions and function approximations, in this paper, the notation $\mathbf{X}_{i:j}$ is a matrix whose rows are the values $(\mathbf{X}_i, \ldots, \mathbf{X}_j)$, that is,

$$\mathbf{X}_{i:j} = \begin{bmatrix} \mathbf{X}_i \\ \mathbf{X}_{i+1} \\ \vdots \\ \mathbf{X}_j \end{bmatrix}. \tag{5}$$

Furthermore, the notation $\mathbf{A}^T$ denotes the transpose of a matrix $\mathbf{A}$. The inner product between two matrices $\mathbf{A}$ and $\mathbf{B}$ of the same size is then denoted $\langle \mathbf{A}, \mathbf{B} \rangle$ and is defined as

$$\langle \mathbf{A}, \mathbf{B} \rangle = \text{trace } \mathbf{A}^T \mathbf{B}. \tag{6}$$

Note that when $\mathbf{A}$ and $\mathbf{B}$ are row vectors, $\langle \mathbf{A}, \mathbf{B} \rangle$ is the standard inner product.

Next, differentiation matrices are used throughout this paper. The following conventions will be adopted for the elements of a differentiation matrix $\mathbf{D}$:

$$
\begin{aligned}
\mathbf{D}_{(i,j)} \quad &= \text{element in row } i \text{ and column } j, \\
\mathbf{D}_{(:,i)} \quad &= \text{elements in all rows and column } i, \\
\mathbf{D}_{(i,:)} \quad &= \text{elements all columns and row } i, \\
\mathbf{D}_{(i:j,k:l)} &= \text{elements in rows } i \text{ through } j \text{ and columns } k \text{ through } l.
\end{aligned}
$$

Finally, the following conventions are adopted for functions and their first derivatives (gradients or Jacobians). First, if $\mathbf{f} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ is a function of the vector $\mathbf{x} \in \mathbb{R}^n$, then $\mathbf{f}(\mathbf{x})$ is given as

$$\mathbf{f}(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \ldots, f_m(\mathbf{x})]. \tag{7}$$

Furthermore, the notation $\nabla_{\mathbf{x}} \mathbf{f}(\mathbf{x})$ is defined as

$$\nabla \mathbf{f}(\mathbf{x}) = \frac{\partial \mathbf{f}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial \mathbf{x}} \\ \frac{\partial f_2}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial f_m}{\partial \mathbf{x}} \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_2}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}. \tag{8}$$

Then, using the definitions provided in Eqs. (7) and (8), if $g : \mathbb{R}^{n \times m} \to \mathbb{R}$ is a scalar function of the $m \times n$ matrix $\mathbf{X} \in \mathbb{R}^{m \times n}$, then the gradient of $g(\mathbf{X})$ with respect to $\mathbf{X}$, denoted $\nabla_{\mathbf{X}} g(\mathbf{X})$, is defined as

$$
\nabla_{\mathbf{X}} g(\mathbf{X}) = \begin{bmatrix} \frac{\partial g}{\partial X_{11}} & \cdots & \frac{\partial g}{\partial X_{1n}} \\ \frac{\partial g}{\partial X_{21}} & \cdots & \frac{\partial g}{\partial X_{2n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial g}{\partial X_{m1}} & \cdots & \frac{\partial g}{\partial X_{mn}} \end{bmatrix}. \tag{9}
$$

3 Bolza Optimal Control Problem

Without loss of generality, consider the following optimal control problem in Bolza form. Minimize the objective functional

$$
\mathcal{J} = \mathcal{M}(\mathbf{x}(-1), \mathbf{v}(-1), \mathbf{x}(+1), \mathbf{v}(+1), t_0, t_f) + \frac{t_f - t_0}{2} \int_{-1}^{+1} \mathcal{L}(\mathbf{x}(t), \mathbf{v}(t), \mathbf{u}(t)) \mathrm{d}t, \tag{10}
$$

subject to the dynamic constraints

$$
\begin{aligned}
\frac{\mathrm{d}\mathbf{x}(t)}{\mathrm{d}t} &= \frac{t_f - t_0}{2} \mathbf{f}_x(\mathbf{x}(t), \mathbf{v}(t)), \\
\frac{\mathrm{d}\mathbf{v}(t)}{\mathrm{d}t} &= \frac{t_f - t_0}{2} \mathbf{f}_v(\mathbf{x}(t), \mathbf{v}(t), \mathbf{u}(t)),
\end{aligned} \tag{11}
$$

the control inequality path constraints

$$
\mathbf{c}(\mathbf{u}(t)) \le \mathbf{0}, \tag{12}
$$

and the boundary conditions

$$
\mathbf{b}(\mathbf{x}(-1), \mathbf{v}(-1), \mathbf{x}(+1), \mathbf{v}(+1), t_0, t_f) = \mathbf{0}. \tag{13}
$$

It is noted in Eqs. (10)–(13) that $\mathbf{x}(t) \in \mathbb{R}^{n_x}$, $\mathbf{v}(t) \in \mathbb{R}^{n_v}$, and, together, $(\mathbf{x}(t), \mathbf{v}(t)) \in \mathbb{R}^n$ is the state (where $n = n_x + n_v$), $\mathbf{u}(t) \in \mathbb{R}^{n_u}$ is the control, $\mathbf{f}_x : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_x}$, $\mathbf{f}_v : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_v}$, $\mathbf{c} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_u} \to \mathbb{R}^{n_c}$, $\mathbf{b} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}^{n_b}$, $\mathcal{M} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \to \mathbb{R}$, and $\mathcal{L} : \mathbb{R}^{n_x} \times \mathbb{R}^{n_v} \times \mathbb{R}^{n_u} \to \mathbb{R}$. It is seen from from the optimal control problem in Eq. (10)–(13) that the dynamics are decomposed into those differential equations that depend upon the control and those differential equations that do not depend upon the control. This decomposition is done deliberately because the modified Legendre–Gauss–Radau collocation method developed in this paper exploits this separation. It is noted that no generality is lost with such a decomposition because $(n_x, n_v) = (0, n)$ is a special case of the dynamics given in Eq. (11).

Consider now the following partitioning of the independent variable $t \in [-1, +1]$ into a mesh consisting of $K + 1$ mesh points $-1 = T_0 < T_1 < T_2 < \ldots < T_K = +1$, where $\mathcal{I}_k = [T_{k-1}, T_k]$ corresponds to mesh interval $k \in [1, \ldots, K]$. Then, the Bolza optimal control problem of Sect. 3 can be expressed in multiple-interval form as follows. Minimize the objective functional

$$
\begin{aligned}
\mathcal{J} = \mathcal{M}(\mathbf{x}^{(1)}(T_0), \mathbf{v}^{(1)}(T_0), \mathbf{x}^{(K)}(T_K), \mathbf{v}^{(K)}(T_K), t_0, t_f) \\
+ \frac{t_f - t_0}{2} \sum_{k=1}^{K} \int_{T_{k-1}}^{T_k} \mathcal{L}(\mathbf{x}^{(k)}(t), \mathbf{v}^{(k)}(t), \mathbf{u}^{(k)}(t)) dt,
\end{aligned}
\tag{14}
$$

subject to the dynamic constraints

$$
\begin{aligned}
\frac{d\mathbf{x}^{(k)}(t)}{dt} &= \frac{t_f - t_0}{2} \mathbf{f}_x(\mathbf{x}^{(k)}(t), \mathbf{v}^{(k)}(t)), \\
\frac{d\mathbf{v}^{(k)}(t)}{dt} &= \frac{t_f - t_0}{2} \mathbf{f}_v(\mathbf{x}^{(k)}(t), \mathbf{v}^{(k)}(t), \mathbf{u}^{(k)}(t)),
\end{aligned}
\tag{15}
$$

the control inequality path constraints

$$
\mathbf{c}(\mathbf{u}^{(k)}(t)) \leq \mathbf{0}, \quad (k = 1, \ldots, K),
\tag{16}
$$

the boundary conditions

$$
\mathbf{b}(\mathbf{x}^{(1)}(T_0), \mathbf{v}^{(1)}(T_0), \mathbf{x}^{(K)}(T_K), \mathbf{v}^{(K)}(T_K), t_0, t_f) = \mathbf{0},
\tag{17}
$$

and the state continuity constraint

$$
\left(\mathbf{x}^{(k)}(T_k), \mathbf{v}^{(k)}(T_k) \right) = \left(\mathbf{x}^{(k+1)}(T_k), \mathbf{v}^{(k+1)}(T_k) \right), \quad (k = 1, \ldots, K - 1),
\tag{18}
$$

at the boundaries of the interior mesh intervals. It is noted that Eq. (18) ensures continuity in the constraint across the domain $t \in [-1, +1]$ as is assumed in the original formulation of the Bolza optimal control problem stated in Sect. 3.

The multiple-interval form of the Bolza optimal control problem given in Eqs. (14)–(18) is now transformed to the independent variable $\tau \in [-1, +1]$ on each mesh interval $\mathcal{I}_k$, $(k = 1, \ldots, K)$. First, it is seen that $t \in [T_{k-1}, T_k]$ can be related to $\tau \in [-1, +1]$ as

$$
t = \frac{T_k - T_{k-1}}{2} \tau + \frac{T_k + T_{k-1}}{2}
\tag{19}
$$

which further implies that

$$
\frac{dt}{d\tau} = \frac{T_k - T_{k-1}}{2} \equiv \alpha_k, \quad (k = 1, \ldots, K).
\tag{20}
$$

Consequently, the multiple-interval Bolza optimal control problem given in Eqs. (14)–(18) can be written in terms of the variable τ as follows. Minimize the objective

functional

$$\mathcal{J} = \mathcal{M}(\mathbf{x}^{(1)}(-1), \mathbf{v}^{(1)}(-1), \mathbf{x}^{(K)}(+1), \mathbf{v}^{(K)}(+1), t_0, t_f)$$
$$+ \frac{t_f - t_0}{2} \sum_{k=1}^{K} \int_{-1}^{+1} \alpha_k \mathcal{L}(\mathbf{x}^{(k)}(\tau), \mathbf{v}^{(k)}(\tau), \mathbf{u}^{(k)}(\tau)) \mathrm{d}\tau, \tag{21}$$

subject to the dynamic constraints

$$\begin{aligned} \frac{\mathrm{d}\mathbf{x}^{(k)}(\tau)}{\mathrm{d}\tau} &\equiv \dot{\mathbf{x}}^{(k)}(\tau) = \frac{t_f - t_0}{2} \alpha_k \mathbf{f}_x(\mathbf{x}^{(k)}(\tau), \mathbf{v}^{(k)}(\tau)), \\ \frac{\mathrm{d}\mathbf{v}^{(k)}(\tau)}{\mathrm{d}\tau} &\equiv \dot{\mathbf{v}}^{(k)}(\tau) = \frac{t_f - t_0}{2} \alpha_k \mathbf{f}_v(\mathbf{x}^{(k)}(\tau), \mathbf{v}^{(k)}(\tau), \mathbf{u}^{(k)}(\tau)), \end{aligned} \quad (k = 1, \dots, K), \tag{22}$$

the control inequality path constraints

$$\mathbf{c}(\mathbf{u}^{(k)}(\tau)) \le \mathbf{0}, \quad (k = 1, \dots, K), \tag{23}$$

the boundary conditions

$$\mathbf{b}(\mathbf{x}^{(1)}(-1), \mathbf{v}^{(1)}(-1), \mathbf{x}^{(K)}(+1), \mathbf{v}^{(K)}(+1), t_0, t_f) = \mathbf{0}, \tag{24}$$

and the state continuity constraint

$$\left(\mathbf{x}^{(k)}(+1), \mathbf{x}^{(k)}(+1)\right) = \left(\mathbf{x}^{(k+1)}(-1), \mathbf{x}^{(k+1)}(-1)\right), \quad (k = 1, \dots, K - 1). \tag{25}$$

4 Legendre–Gauss–Radau Collocation

The Legendre–Gauss–Radau (LGR) collocation method approximates the multiple-interval form of the Bolza optimal control problem defined in Sect. 3. First, it is assumed that the number of collocation points is the same in each mesh interval and is denoted N. Next, let τ_i, $(i = 1, \dots, N)$ be the N Legendre–Gauss–Radau collocation points [47] on the interval $[-1, +1)$ and that $\tau_{N+1} = +1$ is a noncollocated point. Then, in every mesh interval $k \in [1, \dots, K]$, the state $(\mathbf{x}^{(k)}(\tau), \mathbf{v}^{(k)}(\tau))$ is approximated as

$$\begin{aligned} \mathbf{x}^{(k)}(\tau) &\approx \hat{\mathbf{x}}^{(k)}(\tau) = \sum_{j=1}^{N+1} \mathbf{X}_j^{(k)} \ell_j(\tau), \\ \mathbf{v}^{(k)}(\tau) &\approx \hat{\mathbf{v}}^{(k)}(\tau) = \sum_{j=1}^{N+1} \mathbf{V}_j^{(k)} \ell_j(\tau), \end{aligned} \tag{26}$$

where $\ell_j(\tau)$ are the Lagrange polynomials

$$\ell_j(\tau) = \prod_{\substack{l=1 \\ j \ne l}}^{N+1} \frac{\tau - \tau_l}{\tau_j - \tau_l}, \quad (j = 1, \dots, N + 1) \tag{27}$$

whose support points are $(\tau_1, \ldots, \tau_{N+1})$. Differentiating $\mathbf{x}^{(k)}(\tau)$ and $\mathbf{v}^{(k)}(\tau)$ in Eq. (26) gives

$$\begin{aligned}
\dot{\mathbf{x}}^{(k)}(\tau) &\approx \dot{\hat{\mathbf{x}}}^{(k)}(\tau) = \sum_{j=1}^{N+1} \mathbf{X}_j^{(k)} \dot{\ell}_j(\tau), \\
\dot{\mathbf{v}}^{(k)}(\tau) &\approx \dot{\hat{\mathbf{v}}}^{(k)}(\tau) = \sum_{j=1}^{N+1} \mathbf{V}_j^{(k)} \dot{\ell}_j(\tau).
\end{aligned} \tag{28}$$

Evaluating the functions $\dot{\mathbf{x}}^{(k)}(\tau)$ and $\dot{\mathbf{v}}^{(k)}(\tau)$ at $\tau = \tau_i$ gives

$$\begin{aligned}
\dot{\mathbf{x}}^{(k)}(\tau_i) &\approx \dot{\hat{\mathbf{x}}}^{(k)}(\tau_i) = \sum_{j=1}^{N+1} \mathbf{X}_j^{(k)} \dot{\ell}_j(\tau_i) = \sum_{j=1}^{N+1} \mathbf{D}_{(i,j)} \mathbf{X}_j^{(k)}, \\
\dot{\mathbf{v}}^{(k)}(\tau_i) &\approx \dot{\hat{\mathbf{v}}}^{(k)}(\tau_i) = \sum_{j=1}^{N+1} \mathbf{V}_j^{(k)} \dot{\ell}_j(\tau_i) = \sum_{j=1}^{N+1} \mathbf{D}_{(i,j)} \mathbf{V}_j^{(k)},
\end{aligned} \tag{29}$$

where the coefficients $\mathbf{D}_{(i,j)}$, $(i = 1, \ldots, N;\ j = 1, \ldots, N+1)$ form the $N \times (N+1)$ *LGR differentiation matrix* $\mathbf{D}$. Next, the matrices $\mathbf{X}^{(k)} \in \mathbb{R}^{(N+1) \times n_x}$ and $\mathbf{V}^{(k)} \in \mathbb{R}^{(N+1) \times n_v}$ correspond row-wise to the state approximations at $(\tau_1, \ldots, \tau_{N+1})$, while the matrix $\mathbf{U}^{(k)} \in \mathbb{R}^{N \times n_u}$ corresponds row-wise to the approximations of the control at $(\tau_1, \ldots, \tau_N)$. The LGR approximation of the state leads to the following nonlinear programming problem (NLP) that approximates the optimal control problem given in Eqs. (10)–(13). Minimize the objective function

$$J = \mathcal{M}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) + \frac{t_f - t_0}{2} \sum_{k=1}^{K} \sum_{i=1}^{N} \alpha_k w_i \mathcal{L}(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}), \tag{30}$$

subject to

$$\begin{aligned}
\mathbf{D}_{(i,:)} \mathbf{X}^{(k)} - \tfrac{t_f - t_0}{2} \alpha_k \mathbf{f}_x\left(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}\right) &= \mathbf{0}, \\
\mathbf{D}_{(i,:)} \mathbf{V}^{(k)} - \tfrac{t_f - t_0}{2} \alpha_k \mathbf{f}_v(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) &= \mathbf{0}, \quad (k = 1, \ldots, K), \\
\mathbf{c}(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) &\leq \mathbf{0}, \\
\mathbf{b}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) &\leq \mathbf{0}, \\
\left(\mathbf{X}_{N+1}^{(k)}, \mathbf{V}_{N+1}^{(k)}\right) &= \left(\mathbf{X}_1^{(k+1)}, \mathbf{V}_1^{(k+1)}\right), \quad (k = 1, \ldots, K-1),
\end{aligned} \tag{31}$$

where $i \in (1, \ldots, N)$. It is noted in Eq. (30) that w_i, $(i = 1, \ldots, N)$ are the LGR quadrature weights. Equations (30) and (31) will be referred to as the *Legendre–Gauss–Radau collocation method*.

5 LGR Collocation and Lavrentiev Phenomenon

This section provides an overview of Lavrentiev phenomenon. First, Sect. 5.1 provides a discussion of the concept of Lavrentiev phenomenon and how Lavrentiev phenomenon manifests itself when using LGR collocation. Next, Sect. 5.2 provides a discussion of the Lavrentiev gap that arises when solving an optimal control problem whose solution is nonsmooth using LGR collocation. Finally, Sect. 5.3 provides an analysis of the search space using LGR collocation on an example optimal control problem whose solution is nonsmooth.

5.1 Lavrentiev Phenomenon

The LGR collocation method described in Sect. 4 is a finite element method that approximates the optimal control problem described in Sect. 3 with a finite-dimensional nonlinear programming problem (NLP). Reference [41] described cases where the optimal solution of the finite-dimensional approximation produces a pseudo-minimizer that differs from the true optimal solution. The behavior described in Ref. [41] is called *Lavrentiev phenomenon* [40,41] and is important to understand and address when solving an optimal control problem using LGR collocation. To gain an understanding of Lavrentiev phenomenon, consider the classical least action calculus of variations problem of the form

$$\min J(x) = \int_a^b L(x(t), \dot{x}(t), t) dt \text{ subject to } (x(a), x(b)) = (x_0, x_f). \qquad (32)$$

Suppose now that $\mathcal{A}(a, b)$ and $\mathcal{W}(a, b)$ are, respectively, the space of absolutely continuous functions and Lipschitz continuous functions on the interval $t \in [a, b]$. Furthermore, consider particular instances where the minimizer $x^*(t)$ lies in $\mathcal{A}(a, b)$ [41]. For such cases, the minimizer $x^*(t)$ of $J(x)$ has an unbounded derivative at certain points [41], and these singularities may prevent the minimizer from satisfying the classical first-order Euler–Lagrange necessary optimality conditions

$$\frac{\partial L}{\partial x} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{x}} = 0, \qquad (33)$$

where in general, the weak form of the Euler–Lagrange equations that is usually satisfied.

Now, in general, it is not possible to solve a calculus of variations problem analytically. Consequently, the integral in Eq. (32) must be approximated numerically via quadrature using a finite-element method, and this quadrature approximation leads to a finite-dimensional nonlinear programming problem (NLP) that must be solved using nonlinear optimization solvers [2,3]. As it turns out, when the minimizer $x^*(t)$ lies in $\mathcal{A}(a, b)$ collocation methods typically fail in computing both the correct minimizer $x^*(t)$ and the correct minimizing value of the integral $I(x)$. To illustrate the failure of the finite element method, consider the following problem [41] of minimizing over $\mathcal{A}(0, 1)$ the integral

$$J(x) = \int_0^1 (x^3(t) - t)^2 \dot{x}^6(t) dt, \quad (x(0), x(+1)) = (0, +1), \qquad (34)$$

where $x^*(t) = t^{1/3} \in \mathcal{A}(0, 1)$ is the unique minimizer and $I(x^*) = 0$. This last fact, namely that $x^*(t)$ lies in $\mathcal{A}(0, 1)$, can be connected to the following result from Maniá [45]:

$$\inf_{x \in \mathcal{W}(0,1)} J(x) > \inf_{x \in \mathcal{A}(0,1)} J(x) = 0. \qquad (35)$$
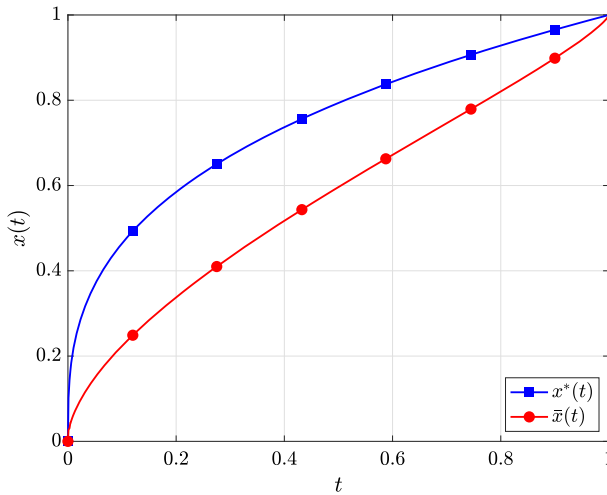
Fig. 1 Minimizer $x^*(t) = t^{1/3}$ of Eq. (36) alongside hp LGR ($N = 4$) collocation pseudo-minimizer $\bar{x}(t)$

The property defined by Eq. (35) is called *Lavrentiev phenomenon* and shows that the minimizer over $\mathcal{A}(0, 1)$ differs from the minimizer over $\mathcal{W}(0, 1)$. Thus, if the minimizer is absolutely continuous, while the optimization search is performed over the space of Lipschitz continuous functions, the result of the optimization will be a pseudo-minimizer $\bar{x} \neq x^*$ [41].

The preceding discussion leads into the fact that Lavrentiev phenomenon can create misleading results when employing numerical optimization with a finite-element method. To see the effect that Lavrentiev phenomenon can have when using a finite element method, consider the following equivalent formulation of the problem in Eq. (34) as the Lagrange optimal control problem

$$\min J(x, u) = \int_0^{+1} (x^3(t) - t)^2 u^6(t) \mathrm{d}t \quad \text{subject to} \begin{cases} \dot{x}(t) & = u(t), \\ (x(0), x(+1)) & = (0, 1). \end{cases}$$
$$(36)$$

Figure 1 shows the exact solution $x^*(t)$ alongside the solution obtained using the multiple-interval LGR collocation method [9–13,19–22] described in Sect. 4 using $N = 4$ LGR collocation points in each mesh interval and the NLP solver IPOPT [3]. Similar to the result obtained in Ref. [41] using midpoint rule integration, Fig. 1 shows that the LGR approximation does not match the optimal solution $x^*(t)$. In fact, consistent with the discussion in Ref. [41], the LGR approximation converges to pseudo-minimizer that differs from $x^*(t)$. Over the years, the concept of Lavrentiev phenomenon has been expanded beyond those that involve the space of absolutely continuous and Lipschitz continuous functions [48]. For instance, Guerra [31] 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.

The modification of the LGR collocation method described in Sect. 4 developed in this paper is motivated by the preceding discussion of Lavrentiev phenomenon. In

particular, it was shown in this section that it is essential to perform the optimization over the appropriate search space. The method developed in this paper focuses on identifying the correct search space when solving an optimal control problem whose solution is nonsmooth. The remainder of this paper focuses on the aforementioned modification of the LGR collocation method [9–13,19–22] described in Sect. 4.

5.2 Lavrentiev Gap

The main point of Sect. 5.1 is that the search space of the optimization problem affects the solution obtained. If an incorrect search space is used, an incorrect solution may be obtained and the corresponding objective may be either smaller or larger than the optimal objective. The difference between the incorrect and correct search spaces is called the *Lavrentiev gap*. If the incorrect search space is strictly larger than the correct search space, then the Lavrentiev gap is said to be negative. On the other hand, if the incorrect search space is strictly smaller than the correct search space, then the Lavrentiev gap is said to be positive.

This research focuses on the solution of optimal control problems with nonsmooth solutions. For such problems, it is desired to improve the accuracy in the numerical solution by adjusting the mesh points to coincide with the locations of nonsmoothness in the solution. In particular, if the mesh points are positioned at the exact locations of nonsmoothness and the solution is smooth on the interior of each mesh interval, then the numerical approximation of the optimal control problem would be smooth.

Typically, the locations of nonsmoothness in the solution of an optimal control problem are not known a priori. One strategy for computing a numerical approximation of the solution to the optimal control problem is introduce variables in the optimization that correspond to the locations of nonsmoothness and then adjust the values of these variables to improve the accuracy of the approximation. As shown in Sect. 5.3, the idea of adding variables that correspond to the locations of nonsmoothness may lead to an incorrect result because the search space may be larger than the correct search space. In such a case, the Lavrentiev gap is negative and convergence to an objective value smaller than the optimal value occurs. To close the gap, additional constraints are required to reduce the size of the search space. In this paper, it is shown that adding a new collocation constraint closes the Lavrentiev gap.

5.3 Analysis of Search Space Using LGR Collocation

To show the occurrence of Lavrentiev phenomenon [41,45] as described in Sect. 5.1 and the Lavrentiev gap as described in Sect. 5.2, in this section, the search space associated with the LGR collocation method developed in Sect. 4 is analyzed using an example whose solution contains a bang–bang optimal control. The results obtained studying this bang-bang optimal control problem then sets the stage for the modified LGR collocation method developed in Sect. 6.

5.4 Motivating Example

Consider the following optimal control problem:

$$\min t_f \text{ subject to } \begin{cases} (\dot{x}(t), \dot{v}(t)) = \dfrac{t_f}{2}(v(t), u(t)), \\ u(t) \in (u_{\min}, u_{\max}) = (-1, +1), \\ (x(-1), x(+1), v(-1), v(+1)) = (x_0, v_0, x_f, v_f) = (10, 0, 0, 0). \end{cases}$$

(37)

The optimal solution to the optimal control problem given in Eq. (37) is

$$(x^*(t), v^*(t), u^*(t)) = \begin{cases} \left(-x_0 \dfrac{(t+1)^2}{2} + x_0, -\sqrt{x_0}(t+1), -1\right), & t \le t_s, \\ \left(+x_0 \dfrac{(t-1)^2}{2}, +\sqrt{x_0}(t-1), +1\right), & t > t_s, \end{cases}$$

(38)

where $t_s^* = 0$ and $t_f^* = 2\sqrt{x_0} \approx 6.32456$. It is seen that the $x^*(t)$ is piecewise quadratic while $u^*(t)$ is bang–bang with a single switch at $t = t_s^* = 0$. Given that the state is piecewise quadratic, it should be possible to obtain the exact solution to this example using two intervals.

5.5 Two-Interval Reformulation of Example: Lavrentiev Gap

Consider now the following two-interval reformulation of the example given in Sect. 5.4:

$$\min t_f \text{ subject to } \begin{cases} (\dot{x}^{(k)}(\tau), \dot{v}^{(k)}(\tau)) = \dfrac{t_f}{2}\alpha_k\left(v^{(k)}(\tau), u^{(k)}(\tau)\right), \\ u^{(k)}(\tau) \in (u_{\min}, u_{\max}) = (-1, +1), \\ (x^{(1)}(-1), x^{(2)}(+1), v^{(1)}(-1), v^{(2)}(+1)) = (10, 0, 0, 0). \end{cases} \quad (k = 1, 2)$$

(39)

where $\alpha_k = (T_k - T_{k-1})/2$, $(k = 1, 2)$ as given in Eq. (20) and T_1 is a variable in the problem formulation of Eq. (39) and represents the time at the boundary between the two mesh intervals $\mathcal{I}_1$ and $\mathcal{I}_2$. Suppose now that the LGR collocation method is used to approximate the two-interval optimal control problem of Eq. (39). 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 the exact solution using $N = 2$ collocation points in each subinterval with T_1 included as an optimization variable.

Now define the *approximate control* as

$$\hat{u}^{(k)}(\tau) = \frac{2}{t_f}\frac{1}{\alpha_k}\dot{\hat{v}}^{(k)}(\tau), \quad (k = 1, 2),$$
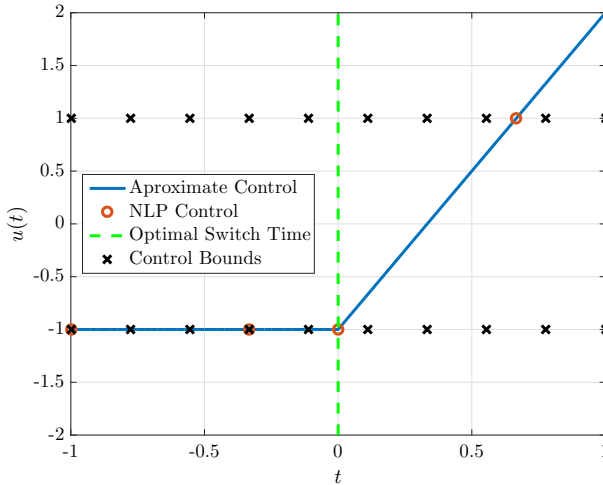
(40)

Fig. 2 Control obtained for two-interval formulation of example given in Eq. (39)

where $\hat{v}^{(k)}(\tau)$ is the $N = 2$ Lagrange polynomial approximation of the state $v^{(k)}(\tau)$. Figure 2 shows the NLP control values $U_i^{(k)}$, $(i, k = 1, 2)$ obtained from solving the NLP using $N = 2$ LGR points in each of the two mesh intervals alongside the approximate control given in Eq. (40). It is seen that the approximate control bears little resemblance to the known bang–bang structure of the optimal control. Next, the NLP solver returns a value $T_1 \approx -0.3$ which is in significant error from the known optimal value $T_1^* = 0$. Furthermore, in the second mesh interval, the approximate control $U^{(2)}(\tau)$ exceeds the upper limit $u_{\max} = +1$ given in the continuous optimal control problem of Eq. (37). Finally, the NLP objective is approximately 6.0 which is less than the optimal objective $2\sqrt{x_0} \approx 6.32456$ of the continuous optimal control problem. Consequently, including the variable T_1 as part of the two-interval formulation results in a misleading solution with regard to the control structure, the objective, and the value of T_1. As a result, the Lavrentiev gap in the formulation of Eq. (39) is positive which implies that the search space is too large. The reason that the search space is too large is because in the discrete problem the control constraint is imposed only at the LGR collocation points $(\tau_1, \ldots, \tau_N)$. As a result, at the final point $\tau_{N+1} = +1$, the approximate control given in Eq. (40) can violate the constraint as shown in Fig. 3.

6 Modified Legendre–Gauss–Radau Collocation

Using the results of Sect. 5, additional constraints are now augmented to the standard collocation method presented in Sect. 4 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
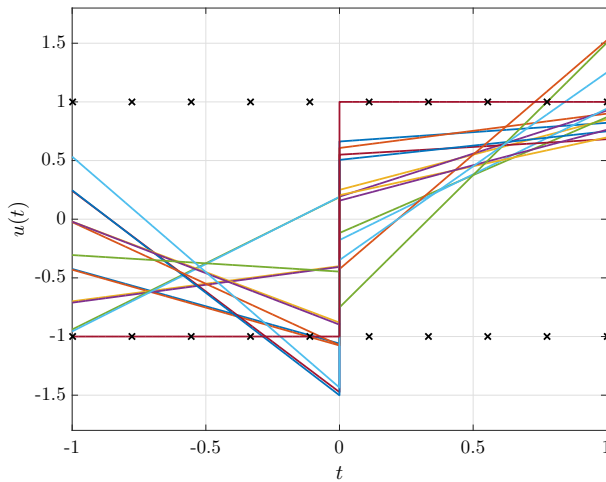
Fig. 3 Possible approximate control functions for the two-interval LGR approximation given in Eq. (39) of the continuous optimal control problem given in Eq. (37)

in Sect. 1, the approach developed in this research differs fundamentally from the approaches developed in Refs. [34] and [39]. To simplify the following discussion, it will be assumed that each mesh interval for any given state discretization contains the same number of collocation points. Therefore, the differentiation matrix for any given state approximation is identical in each mesh interval.

6.1 New Decision Variables

The modified LGR method introduces the following two new decision variables at the end of each mesh interval $\mathcal{I}_k$, $k = (1, \ldots, K)$. The first new variables are the interior mesh points 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 $\mathbf{U}_{N+1}^{(k)}$, $(k = 1, \ldots, K)$. The portion of the decision vector associated with the control in the modified LGR collocation method is then defined as

$$\tilde{\mathbf{U}}^{(k)} = \begin{bmatrix} \mathbf{U}^{(k)} \\ \mathbf{U}_{N+1}^{(k)} \end{bmatrix}.$$

It is important to note that $\mathbf{U}_{N+1}^{(k)}$ and $\mathbf{U}_1^{(k+1)}$ correspond to the same time point T_k. In other words, $\mathbf{U}_{N+1}^{(k)}$ and $\mathbf{U}_1^{(k+1)}$ correspond to the control at T_k^- and T_k^+, respectively. This last point highlights the fact that the control need not be continuous at a mesh point. Reiterating, the two new variables in the modified LGR collocation method are the time at the end of each mesh interval, T_k, $(k = 1, \ldots, K - 1)$, and the control at the end of each mesh interval, $\mathbf{U}_{N+1}^{(k)}$, $(k = 1, \ldots, K - 1)$.

6.2 New Constraints

Given that variables have been added at the end of every mesh interval as described in Sect. 6.1, additional constraints must also be included in order to make the appropriate modifications to the search space. In particular, collocation constraints are added at the end of each mesh interval using those differential equations that are a function of the control. To understand why these new collocation constraints are included, consider the second differential equation $\dot{v}(\tau) = \frac{t_f}{2}\alpha_k u(\tau)$ in the two-interval formulation of the example given in Eq. (39) of Sect. 5.5. Furthermore, suppose that $\hat{v}(\tau)$ is the Lagrange polynomial approximation of $v(\tau)$ and is a polynomial of degree N in each of the two mesh intervals of the problem formulation given in Sect. 5.4. Finally, suppose that the constraint $\dot{v}(\tau) = \frac{t_f}{2}\alpha_k u(\tau)$ is enforced at the N LGR points plus the final point of every mesh interval. Because $\hat{v}(\tau)$ is a polynomial of degree N in each mesh interval and the differential equation depends upon the control, it is possible to satisfy the $N + 1$ conditions

$$\dot{\hat{v}}^{(k)}(\tau_i) - \alpha_k \frac{t_f}{2} U_i^{(k)} = 0, \quad (i = 1, \ldots, N + 1; \; k = 1, 2) \tag{41}$$

in each mesh interval because the control is a variable in Eq. (41). In other words, $U_{N+1}^{(k)}$ can be varied in order to satisfy Eq. (41) at the endpoint of the first interval. Moreover, when adding this collocation condition, it is also necessary to add the constraint that $u_{\min} \leq U_{N+1}^{(k)} \leq u_{\max}$ in order to ensure that the control at the end of every 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. (41) is included along with a constraint that enforces all control bounds at the end of the mesh interval. Adding a collocation condition at the end of a mesh interval results in a modified LGR differentiation matrix of the form

$$\tilde{\mathbf{D}} = \begin{bmatrix} \mathbf{D} \\ \left[\dot{\ell}_1(\tau_{N+1}), \ldots, \dot{\ell}_{N+1}(\tau_{N+1}) \right] \end{bmatrix} \in \mathbb{R}^{(N+1) \times (N+1)}, \tag{42}$$

where noted that $\tilde{\mathbf{D}}$ is a matrix of size $(N + 1) \times (N + 1)$ and the last row of $\tilde{\mathbf{D}}$ is given as

$$\tilde{\mathbf{D}}_{(N+1,1:N+1)} = \left[\dot{\ell}_1(\tau_{N+1}), \ldots, \dot{\ell}_{N+1}(\tau_{N+1}) \right] \in \mathbb{R}^{N+1}. \tag{43}$$

It is important to note that the matrix $\tilde{\mathbf{D}}$ is used to collocate those differential equations that depend upon the control. Furthermore, it is noted that the matrix $\mathbf{D}$ in Eq. (42) is the standard LGR differentiation matrix as given in Sect. 4 [20–22]. Including the new collocation constraint, Eq. (31) is replaced with

$$\begin{aligned} \mathbf{D}_{(i,:)}\mathbf{X}^{(k)} - \alpha_k \tfrac{t_f - t_0}{2} \mathbf{f}_x \left(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)} \right) &= \mathbf{0}, \quad (i = 1, \ldots, N), \\ \tilde{\mathbf{D}}_{(i,:)}\mathbf{V}^{(k)} - \alpha_k \tfrac{t_f - t_0}{2} \mathbf{f}_v \left(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \tilde{\mathbf{U}}_i^{(k)} \right) &= \mathbf{0}, \quad (i = 1, \ldots, N + 1). \end{aligned} \tag{44}$$
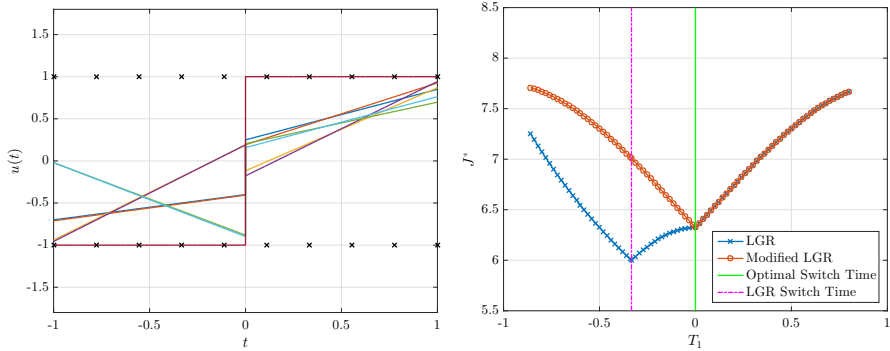
(a) Admissible controls for two-interval formulation of example given in Eqs. (39) using modified LGR collocation.

(b) Optimal objective, $J^* = t_f^*$, vs. T_1 for two-interval formulation of example given in Eqs. (39).

Fig. 4 Admissible controls and optimal control for modified LGR collocation and comparison of optimal objective, J^* versus switch time T_1 for standard and modified LGR collocation

Observe that consistent with the explanation provided earlier in this section, the first constraint in Eq. (44) is *not* a function of control and, as a result, is identical to the first constraint given in Eq. (31).

Additional constraints are added for the new α_k decision variable. These additional constraints are

$$\alpha_k > 0, \quad (k = 1, \dots, K),$$ (45)

$$\sum_{k=1}^{K} \alpha_k - 1 = 0.$$ (46)

These two constraints ensure that each element α_k, $(k = 1, \dots, K)$, is always positive and that the sum is equal to unity. The objective function given in Eq. (30), together with the constraints in Eqs. (44)–(46), is referred to as the *modified Legendre–Gauss–Radau collocation method*.

6.3 Search Space of Modified LGR Method

The example of Sect. 5.4 is now revisited using the modified LGR collocation method. Figure 4 exhibits the impact of the additional collocation constraint from Eq. (44) has on the search space of the example problem. Figure 4a demonstrates that each admissible set for control now falls between the allowable control limits $(u_{\min}, u_{\max}) = (-1, +1)$. Next to examine the effect that the modified LGR method has on the solution of the NLP for the example in Sect. 5.4, Fig. 4b shows the objective of the modified LGR NLP as a function of the switch time, T_1, where it is assumed that the switch time is fixed. At the optimal switch time T_1^*, the objective of both the original and modified LGR methods is identical. Note, however, that when for $T_1 < T_1^*$, the optimal objective of the standard LGR method is *smaller* than the mod-
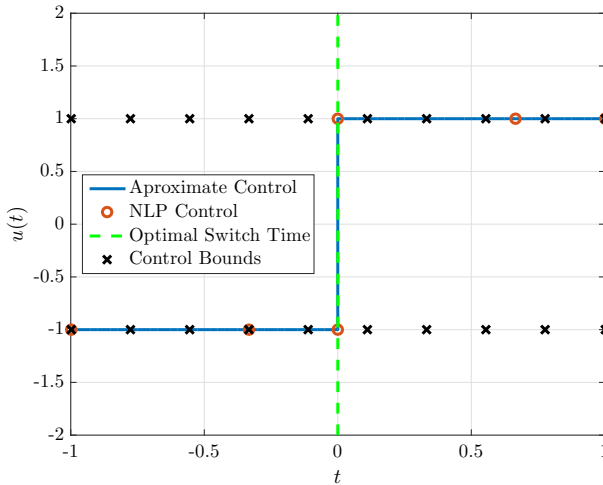
Fig. 5 Optimal control for the example defined by Eq. (37) using the modified LGR method

ified LGR method. In fact, Fig. 4b shows that the optimal objective for the modified LGR method occurs when $T_1 < T_1^*$. 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 $(u_{\min}, u_{\max}) = (-1, +1)$, but the switch time obtained using the modified LGR collocation method matches the switch time of the solution of the continuous optimal control problem.

7 Transformed Adjoint System and Weierstrass–Erdmann Conditions

This section derives the adjoint system of the modified LGR collocation method based on the optimal control problem given in Eqs. (10), (11), and (13). In order to simplify the derivation, the state and control inequality path constraint given in Eq. (16) is dropped. The first-order optimality conditions for the continuous optimal control problem are given as

$$\dot{\boldsymbol{\lambda}}_x = -\frac{\partial \mathcal{L}}{\partial \mathbf{x}} - \boldsymbol{\lambda}_x \left[\frac{\partial \mathbf{f}_x}{\partial \mathbf{x}}\right]^T - \boldsymbol{\lambda}_v \left[\frac{\partial \mathbf{f}_v}{\partial \mathbf{x}}\right]^T, \tag{47}$$

$$\dot{\boldsymbol{\lambda}}_v = -\frac{\partial \mathcal{L}}{\partial \mathbf{v}} - \boldsymbol{\lambda}_x \left[\frac{\partial \mathbf{f}_x}{\partial \mathbf{v}}\right]^T - \boldsymbol{\lambda}_v \left[\frac{\partial \mathbf{f}_v}{\partial \mathbf{v}}\right]^T, \tag{48}$$

$$0 = \frac{\partial \mathcal{L}}{\partial \mathbf{u}} + \lambda_v \left[\frac{\partial \mathbf{f}_v}{\partial \mathbf{u}} \right]^T, \tag{49}$$

$$\lambda_x(-1) = -\frac{\partial \mathcal{M}}{\partial \mathbf{x}(-1)} + \boldsymbol{\psi} \left[\frac{\partial \mathbf{b}}{\partial \mathbf{x}(-1)} \right]^T, \tag{50}$$

$$\lambda_v(-1) = -\frac{\partial \mathcal{M}}{\partial \mathbf{v}(-1)} + \boldsymbol{\psi} \left[\frac{\partial \mathbf{b}}{\partial \mathbf{v}(-1)} \right]^T, \tag{51}$$

$$\lambda_x(+1) = \frac{\partial \mathcal{M}}{\partial \mathbf{x}(+1)} - \boldsymbol{\psi} \left[\frac{\partial \mathbf{b}}{\partial \mathbf{x}(+1)} \right]^T, \tag{52}$$

$$\lambda_v(+1) = \frac{\partial \mathcal{M}}{\partial \mathbf{v}(+1)} - \boldsymbol{\psi} \left[\frac{\partial \mathbf{b}}{\partial \mathbf{v}(+1)} \right]^T, \tag{53}$$

where $\lambda_x(\tau) \in \mathbb{R}^{n_x}$ and $\lambda_v(\tau) \in \mathbb{R}^{n_v}$. 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.

7.1 Derivation of Transformed Adjoint System

The derivation of the transformed adjoint system for the modified LGR collocation method proceeds as follows. First, the Lagrangian associated with the modified LGR collocation constraints of Eq. (44) is given as

$$
\begin{aligned}
J_a &= \mathcal{M}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) + \sum_{k=1}^{K} \alpha_k \frac{t_f - t_0}{2} \sum_{i=1}^{N} w_i \mathcal{L}(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) \\
&\quad - \sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \boldsymbol{\Lambda}_{x_i}^{(k)}, \mathbf{D}_{(i,1:N+1)} \mathbf{X}^{(k)} - \alpha_k \frac{t_f - t_0}{2} \mathbf{f}_x(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}) \right\rangle \\
&\quad - \sum_{k=1}^{K} \sum_{i=1}^{N+1} \left\langle \boldsymbol{\Lambda}_{v_i}^{(k)}, \tilde{\mathbf{D}}_{(i,1:N+1)} \mathbf{V}^{(k)} - \alpha_k \frac{t_f - t_0}{2} \mathbf{f}_v(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) \right\rangle \\
&\quad - \boldsymbol{\Psi} \mathbf{b}^T(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) - \beta \left(\sum_{k=1}^{K} \alpha_k - 1 \right),
\end{aligned}
\tag{54}
$$

where $\boldsymbol{\Lambda}_x^{(k)} \in \mathbb{R}^{N \times n_x}$, $\boldsymbol{\Lambda}_v^{(k)} \in \mathbb{R}^{(N+1) \times n_v}$, $\boldsymbol{\Psi} \in \mathbb{R}^b$, $\beta \in \mathbb{R}$ and $\langle \cdot, \cdot \rangle$ denotes the standard inner product between two vectors. Furthermore, $\boldsymbol{\Lambda}_{x_i}^{(k)}$ and $\boldsymbol{\Lambda}_{v_i}^{(k)}$ denoted the ith rows of $\boldsymbol{\Lambda}_x^{(k)}$ and $\boldsymbol{\Lambda}_v^{(k)}$, respectively. Rewriting Eq. (54) so that the final row of the state matrix is separated from the first N rows gives

$$
\begin{aligned}
J_a = {} & \mathcal{M}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) + \sum_{k=1}^{K} \alpha_k \frac{t_f - t_0}{2} \sum_{i=1}^{N} w_i \mathcal{L}(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) \\
& - \sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \mathbf{D}_{(i,1:N+1)}\mathbf{X}_{1:N}^{(k)} + \mathbf{D}_{(i,N+1)}\mathbf{X}_{N+1}^{(k)} - \alpha_k \frac{t_f - t_0}{2}\mathbf{f}_x(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}) \right\rangle \\
& - \sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \mathbf{\Lambda}_{v_i}^{(k)}, \mathbf{D}_{(i,1:N)}\mathbf{V}_{1:N}^{(k)} + \mathbf{D}_{(i,N+1)}\mathbf{V}_{N+1}^{(k)} - \alpha_k \frac{t_f - t_0}{2}\mathbf{f}_v(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) \right\rangle \\
& - \sum_{k=1}^{K} \left\langle \mathbf{\Lambda}_{v_{N+1}}^{(k)}, \tilde{\mathbf{D}}_{(N+1,1:N)}\mathbf{V}_{1:N}^{(k)} + \tilde{\mathbf{D}}_{(N+1,N+1)}\mathbf{V}_{N+1}^{(k)} \right\rangle \\
& + \sum_{k=1}^{K} \left\langle \mathbf{\Lambda}_{v_{N+1}}^{(k)}, \alpha_k \frac{t_f - t_0}{2}\mathbf{f}_v(\mathbf{X}_{N+1}^{(k)}, \mathbf{V}_{N+1}^{(k)}, \mathbf{U}_{N+1}^{(k)}) \right\rangle \\
& - \mathbf{\Psi}\mathbf{b}^T(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) \\
& - \beta \left(\sum_{k=1}^{K} \alpha_k - 1 \right),
\end{aligned}
\tag{55}
$$

Next, the following theorem is introduced that will allow the terms involving $\mathbf{f}_v(\mathbf{X}_{N+1}^{(k)}, \mathbf{V}_{N+1}^{(k)}, \mathbf{U}_{N+1}^{(k)})$, and $\tilde{\mathbf{D}}_{(N+1,:)}$ in Eq. (55) to be written as functions of $\mathbf{X}_{1:N}^{(k)}, \mathbf{V}_{1:N}^{(k)}, \mathbf{U}_{1:N}^{(k)}$, and $\mathbf{D}_{(:,N+1)}$.

Theorem 1 *Let $f(\tau)$ be a polynomial of degree at most $N - 1$ on the interval $\tau \in [-1, 1]$. Furthermore, let $(\tau_1, \ldots, \tau_N)$ be the Legendre–Gauss–Radau points on the interval $[-1, 1)$ and let $\tau_{N+1} = +1$. Then, if $\ell_j(\tau)$ are the Lagrange polynomials given in Eq. (27), it is the case that*

$$
\int_{-1}^{+1} f(\tau)\dot{\ell}_{N+1}(\tau)\mathrm{d}\tau = f(+1).
\tag{56}
$$

Proof From Eq. (27), the Lagrange polynomial $\ell_{N+1}(\tau)$ is given as

$$
\ell_{N+1}(\tau) = \prod_{l=1}^{N} \frac{\tau - \tau_l}{\tau_{N+1} - \tau_l}.
\tag{57}
$$

Then, the left-hand side of Eq. (56) can be integrated by parts as

$$
\int_{-1}^{+1} f(\tau)\dot{\ell}_{N+1}(\tau)\mathrm{d}\tau = f(\tau)\ell_{N+1}(\tau)\Big|_{-1}^{+1} - \int_{-1}^{+1} \dot{f}(\tau)\ell_{N+1}(\tau)\mathrm{d}\tau.
\tag{58}
$$

Because $f(\tau)$ is a polynomial of degree at most $N - 1$, it follows that $\dot{f}(\tau)$ is a polynomial of degree at most $N - 2$. Furthermore, because $\ell_{N+1}(\tau)$ is a polynomial

of at most degree N, then the integrand in Eq. (58) is at most degree $2N - 2$ and the integral can be evaluated exactly using LGR quadrature as

$$\int_{-1}^{+1} \dot{f}(\tau)\ell_{N+1}(\tau)d\tau = \sum_{i=1}^{N} w_i \dot{f}(\tau_i)\ell_{N+1}(\tau_i), \tag{59}$$

where w_i is the ith LGR quadrature weight. Then, from Eq. (3), every term $\ell_{N+1}(\tau_i) = 0$, $(i \neq N + 1)$ is zero which implies that

$$\int_{-1}^{+1} f(\tau)\dot{\ell}_{N+1}(\tau)d\tau = f(\tau)\ell_{N+1}(\tau)\Big|_{-1}^{+1} \tag{60}$$

Consequently, Eq. (58) reduces to

$$\int_{-1}^{+1} f(\tau)\dot{\ell}_{N+1}(\tau)d\tau = f(\tau)\ell_{N+1}(\tau)\Big|_{-1}^{+1} = f(+1)\ell_{N+1}(+1) - f(-1)\ell_{N+1}(-1) = f(+1). \tag{61}$$

□

The result of Theorem 1 enables expressing the elements $\tilde{\mathbf{D}}_{(N+1,j)}$, $(j = 1, \ldots, N)$ in Eq. (42) in terms of $\mathbf{D}_{(:,N+1)}$ and $\mathbf{D}_{(:,1:N)}$. First, the N elements of $\tilde{\mathbf{D}}_{(N+1,j)}$, $(j = 1, \ldots, N)$ are defined as

$$\tilde{\mathbf{D}}_{(N+1,j)} = \dot{\ell}_j(+1), \quad (j = 1, \ldots, N). \tag{62}$$

Then, replacing $f(\tau)$ in Eq. (56) with $\dot{\ell}_j(\tau)$, $(j = 1, \ldots, N)$, the quantities $\tilde{\mathbf{D}}_{(N+1,j)}$, $(j = 1, \ldots, N)$ are given as

$$\tilde{\mathbf{D}}_{(N+1,j)} = \int_{-1}^{+1} \dot{\ell}_j(\tau)\dot{\ell}_{N+1}(\tau)d\tau, \quad (j = 1, \ldots, N). \tag{63}$$

Because $\dot{\ell}_j(\tau)\dot{\ell}_{N+1}(\tau)$ is a polynomial of degree most $2N-2$, Eq. (63) can be replaced exactly with an LGR quadrature as

$$\tilde{\mathbf{D}}_{(N+1,j)} = \sum_{i=1}^{N} w_i \dot{\ell}_j(\tau_i)\dot{\ell}_{N+1}(\tau_i), \quad (j = 1, \ldots, N). \tag{64}$$

Noting that $\mathbf{D}_{(i,N+1)} = \dot{\ell}_{N+1}(\tau_i)$ and that $\mathbf{D}_{(i,j)} = \dot{\ell}_j(\tau_i)$, Eq. (64) can be written as

$$\tilde{\mathbf{D}}_{(N+1,j)} = \sum_{i=1}^{N} w_i \mathbf{D}_{(i,j)}\mathbf{D}_{(i,N+1)} = \mathbf{D}_{(:,N+1)}^T \mathbf{W} \mathbf{D}_{(:,j)}, \quad (j = 1, \ldots, N), \tag{65}$$

where $\mathbf{W} = \text{diag}(w_1, \ldots, w_N)$ is a diagonal matrix of LGR quadrature weights $(w_1, \ldots, w_N)$. The N quantities $\tilde{\mathbf{D}}_{(N+1,1:N)}$ given in Eq. (65) can be written in a single equation as

$$\tilde{\mathbf{D}}_{(N+1,1:N)} = \mathbf{D}_{(:,N+1)}^T \mathbf{W} \mathbf{D}_{(:,1:N)}. \tag{66}$$

Now suppose that $(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)})$, $(i = 1, \ldots, N+1)$ satisfy the constraints given in Eq. (44). Then, in the case $i = N+1$

$$\alpha_k \frac{t_f - t_0}{2} \mathbf{f}_v \left(\mathbf{X}_{N+1}^{(k)}, \mathbf{V}_{N+1}^{(k)}, \mathbf{U}_{N+1}^{(k)} \right) = \sum_{j=1}^{N+1} \dot{\ell}_j(\tau_{N+1}) \mathbf{V}_j^{(k)}, \tag{67}$$

where $\tau_{N+1} = +1$. Now, let $f(\tau)$ in Theorem 1 be chosen as the vector function

$$\mathbf{F}(\tau) = \sum_{j=1}^{N+1} \dot{\ell}_j(\tau) \mathbf{V}_J^{(k)}, \tag{68}$$

where it is noted in Eq. (68) that $\mathbf{F}(\tau)$ is a polynomial of degree at most $N - 1$. Then, the result Eq. (59) gives

$$\mathbf{F}(+1) = \int_{-1}^{+1} \dot{\ell}_{N+1}(\tau) \mathbf{F}(\tau) \mathrm{d}\tau = \sum_{j=1}^{N+1} \ell_j(+1) \mathbf{V}_j^{(k)}. \tag{69}$$

Next, LGR quadrature is exact for a polynomial of degree at most $2N - 2$ and because $(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)})$, $(i = 1, \ldots, N+1)$ satisfy the constraints in Eq. (44). Therefore, the integral in Eq. (69) can be replaced with

$$\int_{-1}^{+1} \dot{\ell}_{N+1}(\tau) \mathbf{F}(\tau) \mathrm{d}\tau = \sum_{i=1}^{N} w_i \dot{\ell}_{N+1}(\tau_i) \mathbf{F}(\tau_i), \tag{70}$$

which implies that

$$\sum_{j=1}^{N+1} \ell_j(+1) \mathbf{V}_j^{(k)} = \sum_{i=1}^{N} w_i \dot{\ell}_{N+1}(\tau_i) \mathbf{F}(\tau_i) = \alpha_k \frac{t_f - t_0}{2} \sum_{i=1}^{N} w_i \dot{\ell}_{N+1}(\tau_i) \mathbf{f}_v \left(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)} \right). \tag{71}$$

Combining Eqs. (67) and (71) gives

$$\mathbf{f}_v \left(\mathbf{X}_{N+1}^{(k)}, \mathbf{V}_{N+1}^{(k)}, \mathbf{U}_{N+1}^{(k)} \right) = \mathbf{D}_{(:,N+1)}^T \mathbf{W} \mathbf{f}_v \left(\mathbf{X}_{1:N}^{(k)}, \mathbf{V}_{1:N}^{(k)}, \mathbf{U}_{1:N}^{(k)} \right), \tag{72}$$

where

$$\mathbf{f}_v(\mathbf{X}_{1:N}^{(k)}, \mathbf{V}_{1:N}^{(k)}, \mathbf{U}_{1:N}^{(k)}) \equiv \begin{bmatrix} \mathbf{f}_v(\mathbf{X}_1^{(k)}, \mathbf{V}_1^{(k)}, \mathbf{U}_1^{(k)}) \\ \vdots \\ \mathbf{f}_v(\mathbf{X}_N^{(k)}, \mathbf{V}_N^{(k)}, \mathbf{U}_N^{(k)}) \end{bmatrix}.$$

Then, subsituting the identities given in Eqs. (66) and (72) into the Lagrangian of Eq. (55) gives

$$
\begin{aligned}
J_a = {}& \mathcal{M}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) + \sum_{k=1}^{K} \alpha_k \frac{t_f - t_0}{2} \sum_{i=1}^{N} w_i \mathcal{L}(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) \\
& - \sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \mathbf{D}_{(i,1:N)} \mathbf{X}_{1:N}^{(k)} + \mathbf{D}_{(i,N+1)} \mathbf{X}_{N+1}^{(k)} - \alpha_i^{(k)} \frac{t_f - t_0}{2} \mathbf{f}_x(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}) \right\rangle \\
& - \sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \mathbf{\Lambda}_{v_i}^{(k)}, \mathbf{D}_{(i,1:N)} \mathbf{V}_{1:N}^{(k)} - \alpha_k \frac{t_f - t_0}{2} \mathbf{f}_v(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}) \right\rangle \\
& - \sum_{k=1}^{K} \sum_{i=1}^{N} \left\langle \mathbf{\Lambda}_{v_i}^{(k)}, \mathbf{D}_{(i,N+1)} \mathbf{f}_{v_{N+1}}^{(k)} \right\rangle \\
& - \sum_{k=1}^{K} \left\langle \mathbf{\Lambda}_{v_{N+1}}^{(k)}, \mathbf{D}_{(:,N+1)}^{T} \mathbf{W} \mathbf{D}_{(1:N,:)} \mathbf{V}_{1:N}^{(k)} + \tilde{\mathbf{D}}_{(N+1,N+1)} \mathbf{V}_{N+1}^{(k)} \right\rangle \\
& + \sum_{k=1}^{K} \left\langle \mathbf{\Lambda}_{v_{N+1}}, \alpha_i^{(k)} \frac{t_f - t_0}{2} \left(\mathbf{D}_{(:,N+1)}^{T} \mathbf{W} \mathbf{f}_v(\mathbf{X}_{1:N}^{(k)}, \mathbf{V}_{1:N}^{(k)}, \mathbf{U}_{1:N}^{(k)}) \right) \right\rangle \\
& - \mathbf{\Psi} \mathbf{b}^{T}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) \\
& - \beta \left(\sum_{k=1}^{K} \alpha_k - 1 \right).
\end{aligned}
\tag{73}
$$

Now, to simplify the derivations that follow, the following substitutions will be made:

$$
\begin{aligned}
\mathcal{L}_i^{(k)} &= \mathcal{L}(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}), \\
\mathbf{f}_{x_i}^{(k)} &= \mathbf{f}_x(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}), \\
\mathbf{f}_{v_i}^{(k)} &= \mathbf{f}_v(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \mathbf{U}_i^{(k)}).
\end{aligned}
\tag{74}
$$

The KKT conditions are then derived by taking the partial derivatives J_a with respect to $\mathbf{X}^{(k)}, \mathbf{V}^{(k)}, \mathbf{U}^{(k)}, \mathbf{\Lambda}_x^{(k)}, \mathbf{\Lambda}_v^{(k)}, \mathbf{\Psi}^{(k)}, t_0, t_f$ and α_k and setting them equal to zero. These derivatives are given as follows:

$$
\mathbf{D}_{(i,:)} \mathbf{X}^{(k)} - \alpha_k \frac{t_f - t_0}{2} \mathbf{f}_x\left(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}\right) = \mathbf{0}, \quad (i = 1, \ldots, N),
\tag{75}
$$

$$
\tilde{\mathbf{D}}_{(i,:)} \mathbf{V}^{(k)} - \alpha^{(k)} \frac{t_f - t_0}{2} \mathbf{f}_v\left(\mathbf{X}_i^{(k)}, \mathbf{V}_i^{(k)}, \tilde{\mathbf{U}}_i^{(k)}\right) = \mathbf{0}, \quad (k = 1, \ldots, N+1),
\tag{76}
$$

$$
\mathbf{b}(\mathbf{X}_1^{(1)}, \mathbf{V}_1^{(1)}, \mathbf{X}_{N+1}^{(K)}, \mathbf{V}_{N+1}^{(K)}, t_0, t_f) = \mathbf{0},
\tag{77}
$$

$$
\sum_{k=1}^{K} \alpha_k - 1 = 0,
\tag{78}
$$

$$
\mathbf{D}_{(i,:)}^{T} \mathbf{\Lambda}_x^{(k)} = \alpha_k \frac{t_f - t_0}{2}
$$

$$\nabla_{\mathbf{X}}\left(w_i \mathcal{L}_i^{(k)} + \left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \mathbf{f}_{x_i}^{(k)}\right\rangle + \left\langle \mathbf{\Lambda}_{v_i}^{(k)} + \mathbf{\Lambda}_{v_{N+1}}^{(k)}\mathbf{D}_{(i,N+1)}w_i, \mathbf{f}_{v_i}^{(k)}\right\rangle\right)$$
$$- \delta_{1i}(-\nabla_{\mathbf{X}}\mathcal{M} + \nabla_{\mathbf{X}}\mathbf{\Psi}\mathbf{b}^T), \tag{79}$$

$$\mathbf{D}_{(:,N+1)}^T \mathbf{\Lambda}_x = \nabla_{\mathbf{X}}\mathcal{M} - \nabla_{\mathbf{X}}\mathbf{\Psi}\mathbf{b}^T, \tag{80}$$

$$\mathbf{D}_{(i,:)}^T \left(\mathbf{\Lambda}_{v_{1:N}}^{(k)} + \mathbf{\Lambda}_{v_{N+1}}^{(k)}\mathbf{D}_{(i,N+1)}w_i\right) = \alpha_k \frac{t_f - t_0}{2}\nabla_{\mathbf{V}}\left(w_i\mathcal{L}_i^{(k)} + \left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \mathbf{f}_{x_i}^{(k)}\right\rangle\right)$$
$$+ \alpha_k \frac{t_f - t_0}{2}\nabla_{\mathbf{V}}\left(\left\langle \mathbf{\Lambda}_{v_i}^{(k)} + \mathbf{\Lambda}_{v_{N+1}}^{(k)}\mathbf{D}_{(i,N+1)}w_i, \mathbf{f}_{v_i}^{(k)}\right\rangle\right)$$
$$+ \mathbf{\Lambda}_i^{(k)} - \delta_{1i}(-\nabla_{\mathbf{V}}\mathcal{M} + \nabla_{\mathbf{V}}\mathbf{\Psi}\mathbf{b}^T), \tag{81}$$

$$\mathbf{D}_{(:,N+1)}^T \mathbf{\Lambda}_{v_{1:N}} + \tilde{\mathbf{D}}_{(N+1,N+1)}\mathbf{\Lambda}_{v_{N+1}}^{(K)} = \nabla_{\mathbf{V}}\mathcal{M} - \nabla_{\mathbf{V}}\mathbf{\Psi}\mathbf{b}^T, \tag{82}$$

$$\mathbf{0} = \alpha_k \frac{t_f - t_0}{2}\nabla_{\mathbf{U}}\left(w_i\mathcal{L}_i^{(k)} - \left\langle \mathbf{\Lambda}_{v_{1:N}}^{(k)} + \mathbf{\Lambda}_{v_{N+1}}^{(k)}\mathbf{D}_{(i,N+1)}w_i, \mathbf{f}_{v_{1:N}}^{(k)}\right\rangle\right),$$
$$(k = 1, \ldots, K; \ i = 1, \ldots, N), \tag{83}$$

$$0 = \sum_{k=1}^{K}\frac{-\alpha_k}{2}\sum_{i=1}^{N}w_i\mathcal{L}_i^{(k)} + \sum_{i=1}^{N}\left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \frac{-\alpha_k}{2}\mathbf{f}_{x_i}^{(k)}\right\rangle + \sum_{i=1}^{N}\left\langle \mathbf{\Lambda}_{v_i}^{(k)}, \frac{-\alpha_k}{2}\mathbf{f}_{v_i}^{(k)}\right\rangle$$
$$+ \left\langle \mathbf{\Lambda}_{v_{N+1}}^{(k)}, \frac{-\alpha_k}{2}\left(\mathbf{D}_{N+1}^T\mathbf{W}\mathbf{f}^{(k)}\right)\right\rangle + \nabla_{t_0}\left(\mathcal{M} - \mathbf{\Psi}\mathbf{b}^T\right), \tag{84}$$

$$0 = \sum_{k=1}^{K}\frac{\alpha_k}{2}\sum_{i=1}^{N}w_i\mathcal{L}_i^{(k)} + \sum_{i=1}^{N}\left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \frac{\alpha_k}{2}\mathbf{f}_{x_i}^{(k)}\right\rangle + \sum_{i=1}^{N}\left\langle \mathbf{\Lambda}_{v_i}^{(k)}, \frac{\alpha_k}{2}\mathbf{f}_{v_i}^{(k)}\right\rangle$$
$$+ \left\langle \mathbf{\Lambda}_{v_{N+1}}^{(k)}, \frac{\alpha_k}{2}\left(\mathbf{D}_{N+1}^T\mathbf{W}\mathbf{f}_v^{(k)}\right)\right\rangle + \nabla_{t_f}\left(\mathcal{M} - \mathbf{\Psi}\mathbf{b}^T\right), \tag{85}$$

$$0 = \frac{t_f - t_0}{2}\sum_{i=1}^{N}w_i\mathcal{L}_i^{(k)} + \sum_{i=1}^{N}\left\langle \mathbf{\Lambda}_{x_i}^{(k)}, \frac{t_f - t_0}{2}\mathbf{f}_{x_i}^{(k)}\right\rangle + \sum_{i=1}^{N}\left\langle \mathbf{\Lambda}_{v_i}^{(k)}, \frac{t_f - t_0}{2}\mathbf{f}_{v_i}^{(k)}\right\rangle$$
$$+ \left\langle \mathbf{\Lambda}_{v_{N+1}}^{(k)}, \frac{t_f - t_0}{2}\left(\mathbf{D}_{N+1}^T\mathbf{W}\mathbf{f}_v^{(k)}\right)\right\rangle - \beta \quad (k = 1, \ldots, K), \tag{86}$$

where δ_{ij} is the Kronecker delta function defined as

$$\delta_{ij} = \begin{cases} 1, & i = j \\ 0, & i \neq j. \end{cases} \tag{87}$$

The KKT conditions given in equation (86) are unique to the modified LGR method and is not required for an extremal solution of the standard LGR NLP transcription. Now, consider the change of variables

$$\boldsymbol{\lambda}_{x_i}^{(k)} = \frac{\mathbf{\Lambda}_{x_i}^{(k)}}{w_i}, \tag{88}$$

$$\boldsymbol{\lambda}_{x_{N+1}}^{(k)} = \mathbf{D}_{(:,N+1)}^T\mathbf{\Lambda}_{x_{1:N}}^{(K)}, \tag{89}$$

$$\boldsymbol{\psi}_i = \mathbf{\Psi}_i, \tag{90}$$

$$\lambda_{v_i}^{(k)} = \frac{\Lambda_{v_i}^{(k)}}{w_i} + \Lambda_{v_{N+1}}^{(k)} \mathbf{D}_{(i,N+1)}, \tag{91}$$

$$\lambda_{v_{N+1}}^{(K)} = \mathbf{D}_{(:,N+1)}^T \Lambda_{v_{1:N}}^{(K)} + \Lambda_{v_{N+1}} \tilde{\mathbf{D}}_{(N+1,N+1)}. \tag{92}$$

Note that Eqs. (88)–(90) are the same transformations used for the standard LGR method. Finally, define $\mathbf{D}^\dagger \in R^{N \times N}$ such that

$$\mathbf{D}_{(1,1)}^\dagger = -\mathbf{D}_{(1,1)} - \frac{1}{w_1} \tag{93}$$

$$\mathbf{D}_{(i,j)}^\dagger = -\frac{w_j}{w_i} \mathbf{D}_{(j,i)} \quad \text{otherwise}, \tag{94}$$

for $i = j = 1, 2, \ldots, N$. Note that $\mathbf{D}^\dagger$ is the same matrix derived by Refs. [20,21] where it was shown that $\mathbf{D}^\dagger$ is the differentiation matrix for the space of polynomials of degree at most $N - 1$. Now, the KKT conditions can be rewritten as

$$\mathbf{D}_i^\dagger \lambda_{x_{1:N}}^{(k)} = -\alpha_k \frac{t_f - t_0}{2} \nabla_\mathbf{X} \left(\left\langle \lambda_{x_i}^{(k)}, \mathbf{f}_{x_i}^{(k)} \right\rangle + \left\langle \lambda_{v_i}^{(k)}, \mathbf{f}_{v_i}^{(k)} \right\rangle + \mathcal{L}_i^{(k)} \right)$$
$$+ \frac{\delta_{1i}}{w_1^{(1)}} \left(-\nabla_X \left(\mathcal{M} - \boldsymbol{\psi} \mathbf{b}^T \right) - \lambda_{x_1}^{(1)} \right), \tag{95}$$

$$\mathbf{D}_i^\dagger \lambda_{v_{1:N}}^{(k)} = -\alpha_k \frac{t_f - t_0}{2} \nabla_\mathbf{V} \left(\left\langle \lambda_{x_i}^{(k)}, \mathbf{f}_{x_i}^{(k)} \right\rangle + \left\langle \lambda_{v_i}^{(k)}, \mathbf{f}_{v_i}^{(k)} \right\rangle + \mathcal{L}_i^{(k)} \right)$$
$$+ \frac{\delta_{1i}}{w_1} \left(-\nabla_\mathbf{V} \left(\mathcal{M} - \boldsymbol{\psi} \mathbf{b}^T \right) - \lambda_{v_1}^{(1)} \right), \tag{96}$$

$$\mathbf{0} = \alpha_k \frac{t_f - t_0}{2} \nabla_\mathbf{U} \left(\mathcal{L}_i^{(k)} - \left\langle \lambda_{v_i}^{(k)}, \mathbf{f}_{v_i}^{(k)} \right\rangle \right),$$
$$(i = 1, \ldots, N, \quad k = 1, \ldots, K), \tag{97}$$

$$\lambda_{x_{N+1}}^{(K)} = \nabla_\mathbf{X} \left(\mathcal{M} - \boldsymbol{\psi} \mathbf{b}^T \right), \tag{98}$$

$$\lambda_{v_{N+1}}^{(K)} = \nabla_\mathbf{V} \left(\mathcal{M} - \boldsymbol{\psi} \mathbf{b}^T \right), \tag{99}$$

$$-\nabla_{t_0} \left(\mathcal{M} - \boldsymbol{\Psi} \mathbf{b}^T \right) = \sum_{k=1}^{K} -\alpha_k \sum_{i=1}^{N} H_i^{(k)} w_i, \tag{100}$$

$$-\nabla_{t_f} \left(\mathcal{M} - \boldsymbol{\Psi} \mathbf{b}^T \right) = \sum_{k=1}^{K} \alpha_k \sum_{i=1}^{N} H_i^{(k)} w_i, \tag{101}$$

where $H_i^{(k)} = \mathcal{L}_i^{(k)} + \lambda_{x_i}^{(k)} \mathbf{f}_{x_i}^{(k)T} + \lambda_{v_i}^{(k)} \mathbf{f}_{v_i}^{(k)T}$ is the approximation the Hamiltonian, $\mathcal{H}$ in interval k. Equations (50)–(51) allow the terms in the second lines of Eqs. (95)–(96) to vanish which results in Eqs. (95)–(101) becoming discrete representations of the continuous time first-order optimality conditions from Eqs. (47)–(53).

7.2 Weierstrass–Erdmann Conditions

If the optimal control is discontinuous, additional optimality conditions called the *Weierstrass–Erdmann conditions* [29] 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 can be approximated as

$$\mathcal{H}^{(k)}(\tau_i^{(k)}) \approx H_i^{(k)} = \mathcal{L}_i + \lambda_{x_i}^{(k)} \mathbf{f}_{x_i}^{(k)^\mathsf{T}} + \lambda_{v_i}^{(k)} \mathbf{f}_{v_i}^\mathsf{T}, \tag{102}$$

where $\tau_i^{(k)} \in [T_{k-1}, T_k]$, $i = 1, \ldots, N$ and $k = 1, \ldots, K$ are the N LGR points in the kth mesh interval. The Weierstrass–Erdmann condition on the Hamiltonian can be written as [29]

$$\mathcal{H}(T_k^-) = \mathcal{H}(T_k^+), \tag{103}$$

where $\mathcal{H}(T_k^-)$ and $\mathcal{H}(T_k^+)$ are the left and right limits of $\mathcal{H}$ at a point T_k of discontinuity in the control.

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. (103). First, the transformations given in Eqs. (88)–(92) together with the definition of the Hamiltonian given in Eq. (102), Eq. (86) simplifies to

$$\beta = \frac{t_f - t_0}{2} \sum_{i=1}^{N} w_i H_i^{(k)}, \quad (k = 1, \ldots, K). \tag{104}$$

where β is the Lagrange multiplier defined in Eq. (54) associated with the constraint given in Eq. (46). Next, the right-hand side of Eq. (104) is the LGR quadrature approximation to the integral of the Hamiltonian over the interval $[T_{k-1}, T_k]$.

The costate mapping associated with the transformed adjoint system of the standard LGR collocation method 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 Sect. 7 and compares the results of the modified LGR collocation method with the results obtained using the standard LGR collocation method.
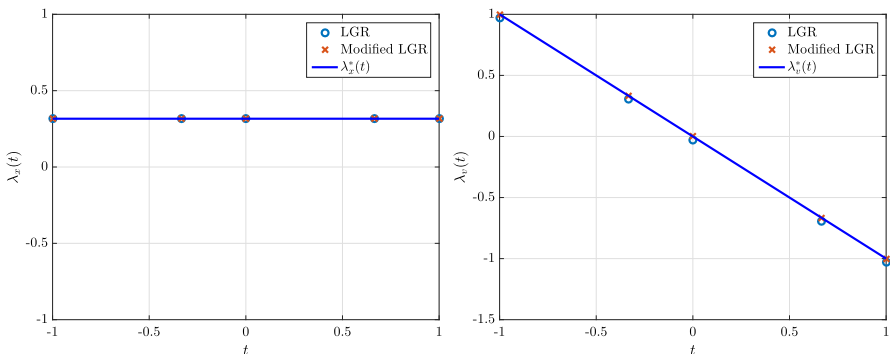
7.3 Example of Costate Estimate

In this section, the costate estimate arising from the modified LGR collocation method is demonstrated on the example problem given in Eq. (37) of Sect. 5.4. 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 b. Figure 6a shows that the dual variables returned for the $\dot{x}$ approximation are exactly the same. Figure 6b
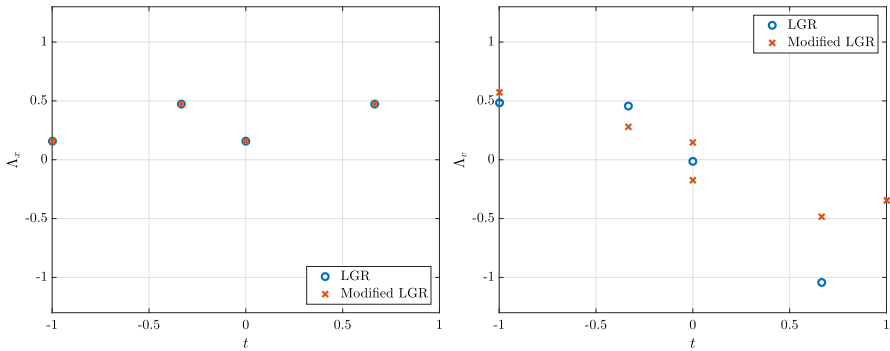
(a) Dual variable, Λ_x, for two-interval formulation of example given in Eq. (39) using both the standard and modified LGR collocation.

(b) Dual variable, Λ_v, for two-interval formulation of example given in Eq. (39) using both the standard and modified LGR collocation.

Fig. 6 Dual variables Λ_x and Λ_v for the example problem using both the standard and modified LGR collocation methods


(a) Costate estimate, $\lambda_x(t)$, for two-interval formulation of example given in Eq. (39) using both the standard and modified LGR collocation.

(b) Costate estimate, $\lambda_v(t)$, for two-interval formulation of example given in Eq. (39) using both the standard and modified LGR collocation.

Fig. 7 Costate estimates λ_x and λ_v for the example problem using both the standard and modified LGR collocation methods

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.

Figure 7a and b 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 $\lambda_v(t)$ 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 $\lambda_v(t)$ is incorrect when using the exact switch time in the
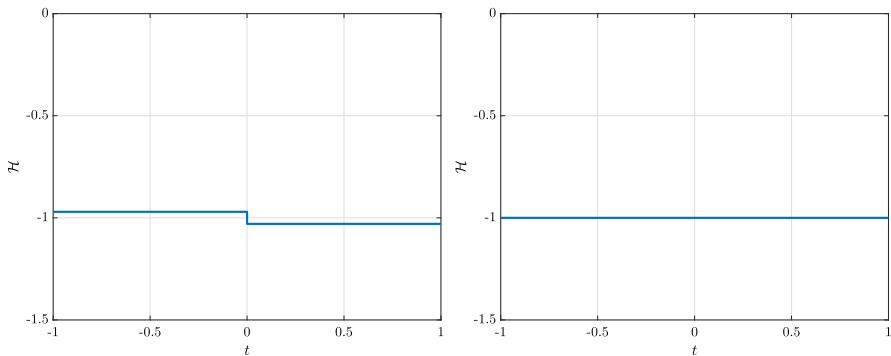
(a) Hamiltonian, $\mathcal{H}$, for the standard LGR method when the correct switch time, $T_1^* = 0$, is provided to the NLP solver.

(b) Hamiltonian, $\mathcal{H}$, for the modified LGR method when the switch time, $T_1^* = 0$, is determined by the NLP solver.

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

standard LGR method implies that the location of the switch time computed by the standard LGR method will also be incorrect.

Figure 8a and b demonstrates 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 T_1 and from T_1 to $+1$ is incorrect. The clear discontinuity in the Hamiltonian from Fig. 8a shows that the Weierstrass–Erdmann conditions from Eq. (103) 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 shown in Fig. 7. Figure 7b shows that the $\lambda_v(T_1) \neq 0$ for the standard LGR method, so not only are the Weierstrass–Erdmann conditions not satisfied, but neither are the standard necessary conditions for optimality. Finally, different from the discontinuous Hamiltonian shown Fig. 8a, b shows that the Hamiltonian obtained using the costate mapping developed in this paper is continuous.

8 Comparison with Methods of References [34–39]

In this section, the modified LGR method developed in this paper is compared against the methods developed in Refs. [34–39]. In particular, Sect. 8.1 provides a comparison of the method of this paper with the work of Ref. [34–38], while Sect. 8.2 provides a comparison of the method of this paper with the work of Ref. [39].

8.1 Comparison with Method of Refs. [34–38]

Reference [34] presents a method that employs collocation at Legendre–Gauss (LG) points. Specifically, the method of Ref. [34] 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. Reference [35] introduced an inequality constraint at the break point location which bounded the value of each control at the break point. Moreover, in order to estimate the value of the control at the breakpoint, the control was parameterized as a polynomial using the information from the interior control points. Next, Refs. [36–38] utilizes LG and LGR direct collocation methods in what is called a nested direct transcription method. In this nested collocation method, the optimal control problem is decomposed into an inner NLP and an outer NLP. The inner NLP is a transcription of the continuous optimal problem using either LG or LGR collocation with fixed mesh interval lengths. The outer NLP problem determines the widths of the mesh interval and satisfies additional optimality conditions (for example, the Weierstrass–Erdmann conditions or higher order optimality conditions if the solution contains a singular arc).

Similar to the methods of Ref. [34] and [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 the methods of Refs. [34] and [35] because the method of this paper not only introduces a variable that defines the time at the junction between two adjacent mesh intervals, but also introduces a new control variable at the end of a mesh interval. Next, in the methods of Refs. [36–38], the control is parameterized as a function of time and this parameterization is used to approximate the control at the end of each mesh interval. On the other hand, in the method of this paper the Weierstrass–Erdmann conditions are satisfied implicitly without requiring that additional constraints be imposed.

8.2 Comparison with Method of Ref. [39]

Reference [39] 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. [34], the method of Ref. [39] 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. [39], 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. [39]. In particular, in the method of Ref. [39], 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. [39] 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. [39] employs collocation at LGL points using a square and singular differentiation matrix, the method of Ref. [39] 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 $\tilde{\mathbf{D}}$ [see Eq. (44)] can be inverted to produce the Legendre–Gauss–Radau integration matrix [20]. As a result, similar to the standard LGR method [20], the method of this paper is also a Gauss quadrature integrator [22]. Second, the singular differentiation matrix employed in Ref. [39] 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 Sect. 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 is shown in Sect. 7.2 that the method of this paper satisfies the Weierstrass–Erdmann conditions.

9 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. Also, an example has been used throughout the paper to motivate the development of the method. Finally, it has been discussed that the method developed in this paper may be applicable in a wide range of aerospace applications where solutions may be nonsmooth including optimal spacecraft orbital transfer, aircraft performance optimization, and hypersonic flight.

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., Saunders, M.A.: SNOPT: an SQP algorithm for large-scale constrained optimization. SIAM Rev. **47**(1), 99–131 (2005)
3. Biegler, L.T., Zavala, V.M.: Large-scale nonlinear programming using IPOPT: an integrating framework for enterprise-wide optimization. Comput. Chem. Eng. **33**(3), 575–582 (2008)

4. Jain, D., Tsiotras, P.: Trajectory optimization using multiresolution techniques. J. Guid. Control Dyn. **31**(5), 1424–1436 (2008)
5. Zhao, Y., Tsiotras, P.: Density functions for mesh refinement in numerical optimal control. J. Guid. Control Dyn. **34**(1), 271–277 (2011)
6. Elnagar, G., Kazemi, M.A., Razzaghi, M.: The pseudospectral Legendre method for discretizing optimal control problems. IEEE Trans. Autom. Control **40**(10), 1793–1796 (1995)
7. Elnagar, G.N., Razzaghi, M.: Short communication: a collocation type method for linear quadratic optimal control problems. Optim. Control Appl. Methods **18**(3), 227–235 (1997)
8. Fahroo, F., Ross, I.M.: Costate estimation by a Legendre pseudospectral method. J. Guid. Control Dyn. **24**(2), 270–277 (2001)
9. Darby, C.L., Hager, W.W., Rao, A.V.: An hp-adaptive pseudospectral method for solving optimal control problems. Optim. Control Appl. Methods **32**(4), 476–502 (2011)
10. Darby, C.L., Hager, W.W., Rao, A.V.: Direct trajectory optimization using a variable low-order adaptive pseudospectral method. J. Spacecr. Rockets **48**(3), 433–445 (2011)
11. Patterson, M.A., Hager, W.W., Rao, A.V.: A ph mesh refinement method for optimal control. Optim. Control Appl. Methods **36**(4), 398–421 (2015)
12. Liu, F., Hager, W.W., Rao, A.V.: Adaptive mesh refinement for optimal control using nonsmoothness detection and mesh size reduction. J. Frankl. Inst. **352**(10), 4081–4106 (2015)
13. Liu, F., Hager, W.W., Rao, A.V.: Adaptive mesh refinement method for optimal control using decay rates of Legendre polynomial coefficients. IEEE Trans. Control Syst. Technol. **26**(4), 1475–1483 (2018)
14. Canuto, C., Hussaini, M.Y., Quarteroni, A.M., Thomas Jr., A.: Spectral Methods in Fluid Dynamics. Springer, Berlin (2012)
15. Fornberg, B.: A Practical Guide to Pseudospectral Methods, vol. 1. Cambridge University Press, Cambridge (1998)
16. Trefethen, L.N.: Spectral Methods in MATLAB, vol. 10. SIAM, Philadelphia (2000)
17. Benson, D.A., Huntington, G.T., Thorvaldsen, T.P., Rao, A.V.: Direct trajectory optimization and costate estimation via an orthogonal collocation method. J. Guid. Control Dyn. **29**(6), 1435–1440 (2006)
18. Rao, A.V., Benson, D.A., Darby, C., Patterson, M.A., Francolin, C., Sanders, I., Huntington, G.T.: Algorithm 902: Gpops, a matlab software for solving multiple-phase optimal control problems using the Gauss pseudospectral method. ACM Trans. Math. Softw. **37**(2), 22 (2010)
19. Kameswaran, S., Biegler, L.T.: Convergence rates for direct transcription of optimal control problems using collocation at Radau points. Comput. Optim. Appl. **41**(1), 81–126 (2008)
20. Garg, D., Patterson, M.A., Darby, C.L., Francolin, C., Huntington, G.T., Hager, W.W., Rao, A.V.: Direct trajectory optimization and costate estimation of finite-horizon and infinite-horizon optimal control problems via a Radau pseudospectral method. Comput. Optim. Appl. **49**(2), 335–358 (2011)
21. Garg, D., Hager, W.W., Rao, A.V.: Pseudospectral methods for solving infinite-horizon optimal control problems. Automatica **47**(4), 829–837 (2011)
22. Garg, D., Patterson, M.A., Hager, W.W., Rao, A.V., Benson, D.A., Huntington, G.T.: A unified framework for the numerical solution of optimal control problems using pseudospectral methods. Automatica **46**(11), 1843–1851 (2010)
23. Hager, W.W., Hou, H., Rao, A.V.: Convergence rate for a Radau collocation method applied to unconstrained optimal control (2015). arXiv:1508.03783
24. Hager, W.W., Hou, H., Rao, A.V.: Convergence rate for a Gauss collocation method applied to unconstrained optimal control. J. Optim. Theory Appl. **169**(3), 801–824 (2016)
25. Hager, W.W., Hou, H., Rao, A.V.: Lebesgue constants arising in a class of collocation methods. IMA J. Numer. Anal. **37**(4), 1884–1901 (2017)
26. Hager, W.W., Liu, J., Mohapatra, S., Rao, A.V., Wang, X.-S.: Convergence rate for a Gauss collocation method applied to constrained optimal control. SIAM J. Control Optim. **56**, 1386–1411 (2018)
27. Hager, W.W., Hou, H., Mohapatra, S., Rao, A.V., Wang, X.-S.: Convergence rate for a Radau hp-collocation method applied to constrained optimal control. Comput. Optim. Appl. **74**, 274–314 (2019)
28. Chen, W., Du, W., Hager, W.W., Yang, L.: Bounds for integration matrices that arise in Gauss and Radau collocation. Comput. Optim. Appl. **74**(1), 259–273 (2019)
29. Bryson, A.E., Ho, Y.: Applied Optimal Control: Optimization, Estimation, and Control. Hemisphere Publishing Corporation, Washington (1975)
30. Darby, C.L.: HP-Pseudospectral Method for Solving Continuous-Time Nonlinear Optimal Control Problems. Ph.D. thesis, University of Florida (2011)

31. Guerra, M., Sarychev, A.: Measuring singularity of generalized minimizers for control-affine problems. J. Dyn. Control Syst. **15**(2), 177–221 (2009)
32. Li, S., Zhao, R., Zhang, Q.: Optimization method for solving bang–bang and singular control problems. J. Control Theory Appl. **10**(4), 559–564 (2012)
33. Martinon, P., Bonnans, F., Laurent-Varin, J., Trélat, E.: Numerical study of optimal trajectories with singular arcs for an Ariane 5 launcher. J. Guid. Control Dyn. **32**(1), 51–55 (2009)
34. Cuthrell, J.E., Biegler, L.T.: On the optimization of differential-algebraic process systems. AIChE J. **33**(8), 1257–1270 (1987)
35. Cuthrell, J.E., Biegler, L.T.: Simultaneous optimization and solution methods for batch reactor control profiles. Comput. Chem. Eng. **13**(1–2), 49–62 (1989)
36. Chen, W., Shao, Z., Biegler, L.T.: A bilevel NLP sensitivitybased decomposition for dynamic optimization with moving finite elements. AIChE J. **60**(3), 966–979 (2014)
37. Chen, W., Biegler, L.T.: Nested direct transcription optimization for singular optimal control problems. AIChE J. **62**(10), 3611–3627 (2016)
38. Chen, W., Ren, Y., Zhang, G., Biegler, L.T.: A simultaneous approach for singular optimal control based on partial moving grid. AIChE J. **65**(6), e16584 (2019)
39. Ross, I.M., Fahroo, F.: Pseudospectral knotting methods for solving nonsmooth optimal control problems. J. Guid. Control Dyn. **27**(3), 397–405 (2004)
40. Lavrentiev, M.: Sur quelques problemes du calcul des variations. Ann. Mat. Pura Appl. **4**, 7–28 (1927)
41. Ball, J.M., Knowles, G.: A numerical method for detecting singular minimizers. Numer. Math. **51**(2), 181–197 (1987)
42. Eide, J. , Rao, A.V.: Lavrentiev phenomenon in *hp* Gaussian quadrature collocation methods for optimal control. In: 2016 AIAA/AAS Astrodynamics Specialist Conference, AIAA Paper 2016–5575, Long Beach, California, 13–16 Sept (2016)
43. Eide, J.D., Hager, W.W., Rao, A.V.: Modified radau collocation method for solving optimal control problems with nonsmooth solutions. Part I: Lavrentiev Phenomenon and the Search Space. In: 2018 IEEE Conference on Decision and Control (CDC), 17–19 Dec (2018), pp. 1644–1650
44. Eide, J.D., Hager, W.W., Rao, A.V.: Modified Radau collocation method for solving optimal control problems with nonsmooth solutions. Part II: costate estimation and the transformed adjoint system. In: 2018 IEEE Conference on Decision and Control (CDC), 17–19 Dec (2018), pp. 1651–1656
45. Manià, B.: Soppa un esempio di Lavrentieff. Boll. Unione Mat. Ital. **13**, 147–153 (1934)
46. Hager, W.W.: Runge–Kutta methods in optimal control and the transformed adjoint system. Numer. Math. **87**, 247–282 (2000)
47. Abramowitz, M., Stegun, I.: Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. Dover Publications, New York (1965)
48. Ferriero, A.: The Lavrentiev phenomenon in the calculus of variations. Ph.d. thesis, Universita degli Studi di Milano-Bicocca (2004)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Joseph D. Eide[1] · William W. Hager[2,3] · Anil V. Rao[1,4]

✉ Anil V. Rao
anilvrao@ufl.edu
https://www.anilvrao.com

Joseph D. Eide
gatoreide@ufl.edu

William W. Hager
hager@ufl.edu
http://people.clas.ufl.edu/hager

1 Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL
   32611, USA

2 Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

3 Society for Industrial and Applied Mathematics, Philadelphia, USA

4 AIAA, Reston, USA