2018 IEEE Conference on Decision and Control (CDC)
Miami Beach, FL, USA, Dec. 17-19, 2018

Modified Radau Collocation Method for Solving Optimal Control Problems with Nonsmooth Solutions Part II: Costate Estimation and the Transformed Adjoint System

Joseph D. Eide,† William W. Hager,‡ and Anil V. Rao§

Abstract— **A modified Legendre-Gauss-Radau collocation method is developed for solving optimal control problems whose solutions contain a nonsmooth optimal control. The method includes an additional variable that defines the location of nonsmoothness. In addition, collocation constraints are added at the end of a mesh interval that defines the location of nonsmoothness in the solution on each differential equation that is a function of control along with a control constraint at the endpoint of this same mesh interval. The transformed adjoint system for the modified Legendre-Gauss-Radau collocation method along with a relationship between the Lagrange multipliers of the nonlinear programming problem and a discrete approximation of the costate of the optimal control problem is then derived. Finally, it is shown via example that the new method provides an accurate approximation of the costate.**

I. Introduction

Over the past few decades, direct collocation methods have become a popular tool to computationally solve nonlinear optimal control problems. In a direct collocation method, the state and control of a continuous-time optimal control problem are discretized at a set of points along a given time interval. The infinite-dimensional optimal control problem is then transcribed to a finite-dimensional nonlinear programming problem (NLP) which can be solved using established NLP solvers [1], [2]. In recent years, a significant amount of research has focused direct *Gaussian quadrature orthogonal collocation methods* [3], [4], [5], [6] where the dynamics are collocated at points associated with a Gaussian quadrature. Most commonly used Gaussian quadrature methods employ Legendre-Gauss (LG) points [3], Legendre Gauss Radau (LGR) points [4], [5], [6], [7], and Legendre-Gauss-Lobatto (LGL) points [8]. In addition, in recent years a convergence theory has been developed using Gaussian quadrature collocation. This theory has led to recent work where it has been shown that, under certain assumptions of the smoothness of solution 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 [9], [10], [11], [12], [13].

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 [14]. The difficulty in solving problems with a nonsmooth optimal control lies in determining the location of the nonsmoothness. For example, dynamical systems where the control appears linearly or problems that have state inequality path constraints often have solutions where the control and state may be nonsmooth. One approach to handling nonsmoothness is to employ a mesh refinement method where the optimal control problem is partitioned into a mesh and a mesh that meets a specified solution accuracy tolerance is obtained iteratively. In the context of Gaussian quadrature collocation, so called hp-adaptive mesh refinement methods [15], [16], [17], [18], [19] have been developed more recently 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).

Now, as it turns out, for optimal control problems where the solution is nonsmooth the convergence theory developed in Refs. [9], [10], [11], [12], [13] 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. [10] are not met are those where the control appears linearly in the problem formulation [14], [20], [21], [22]. One approach for estimating the location of nonsmoothness in solutions to optimal control problems is to introduce a variable called a breakpoint [23] that defines the location of a point of nonsmoothness and to include this variable in the NLP. The key issue that arises by introducing a breakpoint 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. Next, Ref. [24] introduced a variable that defines the switch time and collocate the dynamics at both the end of a mesh interval and the start of the subsequent mesh interval using Legendre-Gauss-Lobatto collocation. Note, however, that the LGL method used in Ref. [24] employs a square and singular differentiation matrix. Therefore, unlike the approach of Ref. [23], which used Legendre-Gauss collocation, the scheme used in Ref. [24] is not a Gauss quadrature integrator.

The goal of this research is to develop a method that accurately determines the locations of nonsmoothness in

†Ph.D. Student, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, Florida 32611-6250. Email: gatoreide@ufl.edu

‡Professor, Department of Mathematics, University of Florida, Gainesville, FL 32611-6250. E-mail: hager@ufl.edu.

§Associate Professor, Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250. E-mail: anilvrao@ufl.edu. Corresponding Author

978-1-5386-1395-5/18/$31.00 ©2018 IEEE

the numerical solution of optimal control problems using Gaussian quadrature collocation. In this paper, which is Part II of a two-part sequence, the transformed adjoint system is derived that relates the Lagrange multipliers of the Karush-Kuhn-Tucker (KKT) system associated with the modified Legendre-Gauss-Radau (LGR) collocation method developed in Ref. [25] to the costate of the optimal control problem. Recall that the modified LGR collocation method developed in Ref. [25] includes collocation conditions adds, to those differential equations that are a function of control, collocation conditions at the end of a mesh interval that may be the point of nonsmoothness in the solution of an optimal control problem. As a result, the transformed adjoint system of the modified LGR method differs from the transformed adjoint system of the standard LGR collocation method [4], [5], [6]. Using this modified transformed adjoint system, a costate estimate for the modified LGR collocation method is developed.

This paper is organized as follows. Section II provides a description of the modified LGR collocation method in a manner similar to that given in Ref. [25]. Section III derives the Karush-Kuhn-Tucker optimality conditions for the modified LGR collocation method. In addition, Section III provides a costate estimate that relates the dual variables of the modified LGR collocation method to the costate of the continuous-time optimal control problem. Section IV demonstrates the accuracy of the costate estimate derived in Section III on an example. Finally, Section V provides conclusions on this work.

II. Legendre-Gauss-Radau Collocation

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

$$\begin{aligned} \mathcal{J} = {} & \mathcal{M}(\mathbf{x}(-1), \mathbf{v}(-1), \mathbf{x}(+1), \mathbf{v}(+1)) \\ & + \int_{-1}^{+1} \mathcal{L}(\mathbf{x}(\tau), \mathbf{v}(\tau), \mathbf{u}(\tau)) d\tau, \end{aligned} \tag{1}$$

subject to the dynamic constraints

$$\begin{aligned} \dot{\mathbf{x}}(\tau) &= \mathbf{v}(\tau), \\ \dot{\mathbf{v}}(\tau) &= \mathbf{f}(\mathbf{x}, (\tau), \mathbf{v}(\tau), \mathbf{u}(\tau)) \end{aligned} \tag{2}$$

and boundary conditions

$$\mathbf{b}(\mathbf{x}(-1), \mathbf{v}(-1), \mathbf{x}(+1), \mathbf{v}(+1)) = \mathbf{0}, \tag{3}$$

where $(\mathbf{x}(\tau), \mathbf{v}(\tau)) \in \mathbb{R}^{2n}$ is the state (such that $\mathbf{x}(\tau) \in \mathbb{R}^n$ and $\mathbf{v}(\tau) \in \mathbb{R}^n$), $\mathbf{u}(\tau) \in \mathbb{R}^m$ is the control, $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $\mathbf{b} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^b$,$\mathcal{M} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$, and $\mathcal{L} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$.

For convenience with the mathematical development that follows, the state or control at a value τ is considered a row vector. For example, $\mathbf{x}(\tau)$ and $\mathbf{u}(\tau)$ are defined as row vectors, respectively, as

$$\begin{aligned} \mathbf{x}(\tau) &= \left[\begin{array}{c} x_1(\tau) \cdots x_n(\tau) \end{array} \right] \in \mathbb{R}^n, \\ \mathbf{u}(\tau) &= \left[\begin{array}{c} u_1(\tau) \cdots u_m(\tau) \end{array} \right] \in \mathbb{R}^m. \end{aligned} \tag{4}$$

All other vector quantities are defined in a similar manner to that shown for $\mathbf{x}(\tau)$ and $\mathbf{u}(\tau)$ given in Eq. (4).

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

$$\ell_i(\tau) = \prod_{\substack{j=1 \\ j \neq i}}^{N+1} \frac{\tau - \tau_j}{\tau_i - \tau_j}, \qquad i = 1, \ldots, N+1.$$

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

$$\begin{aligned} x_j(\tau) &\approx X_j(\tau) = \textstyle\sum_{i=1}^{N+1} X_{ij} \ell_i(\tau), \\ v_j(\tau) &\approx V_j(\tau) = \textstyle\sum_{i=1}^{N+1} V_{ij} \ell_i(\tau), \end{aligned} \tag{5}$$

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

$$\begin{aligned} \dot{x}_j(\tau) &\approx \dot{X}_j(\tau) = \textstyle\sum_{i=1}^{N+1} X_{ij} \dot{\ell}_i(\tau_k) = \sum_{i=1}^{N+1} D_{ik} X_{ij}, \\ \dot{v}_j(\tau) &\approx \dot{V}_j(\tau) = \textstyle\sum_{i=1}^{N+1} V_{ij} \dot{\ell}_i(\tau_k) = \sum_{i=1}^{N+1} D_{ik} V_{ij}. \end{aligned} \tag{6}$$

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

$$\mathbf{D} = \left[\begin{array}{cccc} \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_{N+1} \end{array} \right] = \left[\begin{array}{cc} \mathbf{D}_{1:N} & \mathbf{D}_{N+1} \end{array} \right], \tag{7}$$

where $\mathbf{D}_i$ denotes the i^{th} column of $\mathbf{D}$, $\mathbf{D}_{1:N} \in \mathbb{R}^{N \times N}$ is an $N \times N$ matrix formed from the first N columns of $\mathbf{D}$, and $\mathbf{D}_{N+1}$ is the last column of $\mathbf{D}$ [4], [5], [6]. Thus, unlike the state and control, which are treated as row vectors at an instant of time, in this exposition the differentiation matrix is dealt with column-wise. Using the row vector convention for the state and control. the notation, the matrices $\mathbf{X} \in \mathbb{R}^{(N+1) \times n}$ and $\mathbf{V} \in \mathbb{R}^{(N+1) \times n}$ correspond row-wise to the state approximations at times $(\tau_1, \ldots, \tau_{N+1})$, while the matrix $\mathbf{U} \in \mathbb{R}^{N \times m}$ corresponds row-wise to the approximations of the control at times $(\tau_1, \ldots, \tau_N)$. Therefore, the matrices $\mathbf{X}$, $\mathbf{V}$, and $\mathbf{U}$ are given respectively

$$\begin{aligned} \mathbf{X} &= \left[\begin{array}{c} \mathbf{X}_1 \\ \vdots \\ \mathbf{X}_{N+1} \end{array} \right] \equiv \mathbf{X}_{1:N+1}, \\ \mathbf{V} &= \left[\begin{array}{c} \mathbf{V}_1 \\ \vdots \\ \mathbf{V}_{N+1} \end{array} \right] \equiv \mathbf{V}_{1:N+1}, \\ \mathbf{U} &= \left[\begin{array}{c} \mathbf{U}_1 \\ \vdots \\ \mathbf{U}_N \end{array} \right] \equiv \mathbf{U}_{1:N}, \end{aligned} \tag{8}$$

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

$$\dot{\mathbf{X}}(\tau_k) = [\mathbf{D}\mathbf{X}]_k \quad , \quad \dot{\mathbf{V}}(\tau_k) = [\mathbf{D}\mathbf{V}]_k . \tag{9}$$

It is noted that the state approximation is exact if the state is a polynomial of degree at most N. The LGR approximation of the state leads to the following nonlinear programming problem (NLP) that approximates the optimal control problem given in Eqs. (1)–(3):

$$\begin{aligned} \text{minimize } J = {} & \mathcal{M}(\mathbf{X}_1, \mathbf{V}_1, \mathbf{X}_{N+1}, \mathbf{V}_{N+1}) \\ & + \sum_{i=1}^{N} w_i \mathcal{L}(\mathbf{X}_i, \mathbf{V}_i, \mathbf{U}_i), \end{aligned} \tag{10}$$

subject to

$$\begin{aligned}\mathbf{DX}-\mathbf{V}_{1:N}&=\mathbf{0},\\ \mathbf{DV}-\mathbf{f}(\mathbf{X}_{1:N},\mathbf{V}_{1:N},\mathbf{U}_{1:N})&=\mathbf{0},\\ \mathbf{b}(\mathbf{X}_1,\mathbf{V}_1,\mathbf{X}_{N+1},\mathbf{V}_{N+1})&\leq\mathbf{0},\end{aligned}\tag{11}$$

where w_i, $(i=1,\ldots,N)$ are the LGR quadrature weights (and produce an exact integral if the integrand is a polynomial of degree at most $2N-2$). Equations (10) and (11) will be referred to as the *standard Legendre-Gauss-Radau collocation method*.

In this paper, a modification of the standard LGR collocation method [25] is employed the case where the solution of the optimal control problem may be nonsmooth. The modified LGR collocation method includes an additional variable that defines the location of the nonsmoothness along with collocation conditions at the end of a mesh interval that identifies the nonsmoothness in the solution. Note that, as described in Ref. [25], collocation conditions are added only to those differential equations that are a function of the control along with a constraint that enforces all state and control bounds at the end of the mesh interval that corresponds to the location of the nonsmoothness. The modified LGR collocation method employs the modified LGR differentiation matrix

$$\tilde{\mathbf{D}}=\begin{bmatrix}\mathbf{D}_{1:N} & \mathbf{D}_{N+1}\\ \mathbf{E} & E_0\end{bmatrix},\tag{12}$$

in the mesh intervals where the additional collocation constraints are included for those differential equations that are a function of the control. It is noted in Eq. (12) that $[\mathbf{D}_{1:N}\ \ \mathbf{D}_{N+1}]\in\mathbb{R}^{N\times(N+1)}$ is the standard $N\times(N+1)$ LGR differentiation matrix [4], [5], [6]. Finally, the last row of the $\tilde{\mathbf{D}}$ matrix consists of $\mathbf{E}\in\mathbb{R}^{1\times N}$ and $E_0\in\mathbb{R}$, where this last row corresponds to the fact that a collocation point has been added in the modified LGR method. With the inclusion of the new collocation constraint, the collocation equations given Eq. (11) are modified as

$$\begin{array}{rcl}\mathbf{DX}-\mathbf{V}_{1:N} & = & \mathbf{0},\\ \tilde{\mathbf{D}}\mathbf{V}-\mathbf{f}(\mathbf{X},\mathbf{V},\tilde{\mathbf{U}}) & = & \mathbf{0},\end{array}\tag{13}$$

where

$$\tilde{\mathbf{U}}=\begin{bmatrix}\mathbf{U}\\ \mathbf{U}_{N+1}\end{bmatrix}$$

and $\mathbf{U}_{N+1}$ is the value of the control at $\tau=+1$. The cost function given in Eq. (10), together with the constraints in Eq. (13), is referred to as the *modified Legendre-Gauss-Radau collocation method*.

III. Transformed Adjoint System for LGR Collocation

This section derives the adjoint system for the modified LGR collocation method based on the optimal control problem given in Eqs. (1)–(3) as described in Section II. The first-order optimality conditions for the continuous time problem described in Eqs. (1)–(3) are given as

$$\dot{\boldsymbol{\lambda}}=-\frac{\partial\mathcal{L}}{\partial\mathbf{x}}-\tilde{\boldsymbol{\lambda}}\left[\frac{\partial\mathbf{f}}{\partial\mathbf{x}}\right]^{\mathsf{T}},\tag{14}$$

$$\dot{\tilde{\boldsymbol{\lambda}}}=-\frac{\partial\mathcal{L}}{\partial\mathbf{v}}-\boldsymbol{\lambda}-\tilde{\boldsymbol{\lambda}}\left[\frac{\partial\mathbf{f}}{\partial\mathbf{v}}\right]^{\mathsf{T}},\tag{15}$$

$$0=\frac{\partial\mathcal{L}}{\partial\mathbf{u}}+\tilde{\boldsymbol{\lambda}}\frac{\partial\mathbf{f}}{\partial\mathbf{u}},\tag{16}$$

$$\boldsymbol{\lambda}(-1)=-\frac{\partial\mathcal{M}}{\partial\mathbf{x}(-1)}+\boldsymbol{\psi}\left[\frac{\partial\mathbf{b}}{\partial\mathbf{x}(-1)}\right]^{\mathsf{T}},\tag{17}$$

$$\tilde{\boldsymbol{\lambda}}(-1)=-\frac{\partial\mathcal{M}}{\partial\mathbf{x}(-1)}+\boldsymbol{\psi}\left[\frac{\partial\mathbf{b}}{\partial\mathbf{v}(-1)}\right]^{\mathsf{T}}.\tag{18}$$

$$\boldsymbol{\lambda}(+1)=\frac{\partial\mathcal{M}}{\partial\mathbf{x}(+1)}-\boldsymbol{\psi}\left[\frac{\partial\mathbf{b}}{\partial\mathbf{x}(+1)}\right]^{\mathsf{T}},\tag{19}$$

$$\tilde{\boldsymbol{\lambda}}(+1)=\frac{\partial\mathcal{M}}{\partial\mathbf{x}(+1)}-\boldsymbol{\psi}\left[\frac{\partial\mathbf{b}}{\partial\mathbf{v}(+1)}\right]^{\mathsf{T}},\tag{20}$$

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

$$\begin{aligned}\frac{\partial f}{\partial\mathbf{x}}&=\begin{bmatrix}\frac{\partial f}{\partial x_1} & \cdots & \frac{\partial f}{\partial x_n}\end{bmatrix},\\ \frac{\partial\mathbf{g}}{\partial\mathbf{x}}&=\begin{bmatrix}\frac{\partial\mathbf{g}}{\partial x_1}\\ \vdots\\ \frac{\partial\mathbf{g}^{\mathsf{T}}}{\partial x_n}\end{bmatrix}.\end{aligned}$$

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-time optimal control problem.

Using the NLP associated with the modified LGR collocation method as described in Section II, consider the augmented cost function

$$\text{minimize } J=\mathcal{M}(\mathbf{X}_1,\mathbf{V}_1,\mathbf{X}_{N+1},\mathbf{V}_{N+1})\tag{21}$$

$$+\sum_{i=1}^{N}w_i\mathcal{L}(\mathbf{X}_i,\mathbf{V}_i,\mathbf{U}_i)\tag{22}$$

subject to

$$\mathbf{DX}-\mathbf{V}_{1:N}=\mathbf{0}\tag{23}$$

$$\tilde{\mathbf{D}}\mathbf{V}-\mathbf{f}(\mathbf{X},\mathbf{V},\tilde{\mathbf{U}})=\mathbf{0}\tag{24}$$

$$\mathbf{b}(\mathbf{X}_1,\mathbf{V}_1,\mathbf{X}_{N+1},\mathbf{V}_{N+1})=\mathbf{0}\tag{25}$$

where $\tilde{\mathbf{D}}$ is the modified LGR differentiation matrix as given in Eq. (12). Now the first-order optimality conditions of the discrete system described in Eqs. (21)–(25) are derived. First the augmented cost function is written as

$$\begin{aligned}J_a=&\sum_{i=1}^{N}w_i\mathcal{L}(\mathbf{X}_i,\mathbf{V}_i,\mathbf{U}_i)\\ &-\sum_{i=1}^{N}\langle\boldsymbol{\Lambda}_i,\mathbf{D}_{i,1:N}\mathbf{X}-\mathbf{V}_i\rangle\\ &-\sum_{i=1}^{N+1}\left\langle\tilde{\boldsymbol{\Lambda}}_i,\tilde{\mathbf{D}}_{i,1:N+1}\mathbf{V}_i-\mathbf{f}(\mathbf{X}_i,\mathbf{V}_i,\mathbf{U}_i)\right\rangle\\ &-\boldsymbol{\Psi}\mathbf{b}^{\mathsf{T}}(\mathbf{X}_1,\mathbf{V}_1,\mathbf{X}_{N+1},\mathbf{V}_{N+1})\\ &+\mathcal{M}(\mathbf{X}_1,\mathbf{V}_1,\mathbf{X}_{N+1},\mathbf{V}_{N+1}),\end{aligned}\tag{26}$$

where $\boldsymbol{\Lambda}\in\mathbb{R}^{N\times n}$, $\tilde{\boldsymbol{\Lambda}}\in\mathbb{R}^{(N+1)\times n}$, $\boldsymbol{\Psi}\in\mathbb{R}^{N_b}$ and $\langle\cdot,\cdot\rangle$ denotes the standard inner product between two vectors. Equation (26) is now rewritten to separate the final row of the state matrix from the first N rows

$$
\begin{aligned}
J_a = &\sum_{i=1}^{N} w_i \mathcal{L}(\mathbf{X}_i, \mathbf{V}_i, \mathbf{U}_i) \\
&- \sum_{i=1}^{N} \langle \mathbf{\Lambda}_i, \mathbf{D}_{i,1:N}\mathbf{X}_{1:N} + D_{i,N+1}\mathbf{X}_{N+1} - \mathbf{V}_i \rangle \\
&- \sum_{i=1}^{N} \left\langle \tilde{\mathbf{\Lambda}}_i, \mathbf{D}_{i,1:N}\mathbf{V}_{1:N} + D_{i,N+1}\mathbf{V}_{N+1} - \mathbf{f}(\mathbf{X}_i, \mathbf{V}_i, \mathbf{U}_i) \right\rangle \\
&- \left\langle \tilde{\mathbf{\Lambda}}_{N+1}, \mathbf{E}\mathbf{V}_{1:N} + E_0 \mathbf{V}_{N+1} \right\rangle \\
&- \left\langle \tilde{\mathbf{\Lambda}}_{N+1}, \mathbf{f}(\mathbf{X}_{N+1}, \mathbf{V}_{N+1}, \mathbf{U}_{N+1}) \right\rangle \\
&- \mathbf{\Psi}\mathbf{b}^{\mathsf{T}}(\mathbf{X}_1, \mathbf{V}_1, \mathbf{X}_{N+1}, \mathbf{V}_{N+1}) \\
&+ \mathcal{M}(\mathbf{X}_1, \mathbf{V}_1, \mathbf{X}_{N+1}, \mathbf{V}_{N+1}).
\end{aligned} \tag{27}
$$

Now the following theorem is introduced which will allow the $f(\mathbf{X}_{N+1}, \mathbf{V}_{N+1}, \mathbf{U}_{N+1})$, and $\mathbf{E}$ terms of Eq. (27) to be written as a function of $\mathbf{X}_{1:N}$,$\mathbf{V}_{1:N}$,$\mathbf{X}_{1:N}$, and $\mathbf{D}_{N+1}$.

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

$$
\ell_{N+1}(\tau) = \prod_{1 \le j \le N} \frac{\tau - \tau_j}{\tau_{N+1} - \tau_j}, \tag{28}
$$

then

$$
\int_{-1}^{+1} f(\tau)\dot{\ell}_{N+1}(\tau) d\tau = f(+1). \tag{29}
$$

Proof. Integrating the LHS of Eq. (29) by parts yields

$$
f(\tau)\ell_{N+1}(\tau)\Big|_{-1}^{+1} - \int_{-1}^{+1} \dot{f}(\tau)\ell_{N+1}. \tag{30}
$$

Because $\dot{f}(\tau)$ is a polynomial of degree at most $N - 2$ and $\ell_{N+1}(\tau)$ is a polynomial of at most degree N, then the integrand in Eq. (30) is at most degree $2N-2$ and the integral can be evaluated using LGR quadrature as

$$
\int_{-1}^{+1} \dot{f}(\tau)\ell(\tau)_{N+1} d\tau = \sum_{i=1}^{N} w_i \dot{f}(\tau_i)\ell_{N+1}(\tau_i), \tag{31}
$$

where w_i is the i^{th} LGR quadrature weight. Recall that $\ell_i(\tau_i) = 1$ and $\ell_i(\tau_j) = 0$ when $i \neq j$, then Eq. (31) is zero and Eq. (30) reduces to

$$
\begin{aligned}
f(\tau)\ell_{N+1}(\tau)\Big|_{-1}^{+1} &= f(+1)\ell_{N+1}(+1) \\
&\quad - f(-1)\ell_{N+1}(-1) \\
&= f(+1),
\end{aligned} \tag{32}
$$

which completes the proof. □

Equation (29) allows the vector $\mathbf{E}$ of $\tilde{\mathbf{D}}$ to be related to $\mathbf{D}_{N+1}$ of $\tilde{\mathbf{D}}$. The elements of $\mathbf{E}$ are defined as

$$
E_j = \dot{\ell}_j(+1), \quad j = 1, \ldots, N_k. \tag{33}
$$

From Eq. (29) we can write,

$$
E_j = \int_{-1}^{+1} \dot{\ell}_j(\tau)\dot{\ell}_{N+1}(\tau) d\tau \quad j = 1, \ldots, N. \tag{34}
$$

Equation (34) can be evaluated using Gaussian quadrature

$$
E_j = \sum_{i=1}^{N} w_i \dot{\ell}_j(\tau_i)\dot{\ell}_{N+1}(\tau_i), \quad j = 1, \ldots, N \tag{35}
$$

Note that $\dot{\ell}_{N+1}(\tau_i)$ is the i^{th} element of $\mathbf{D}_{N+1}$. Using the definition of the $\tilde{\mathbf{D}}$ matrix and the relationship from Eq. (29) gives

$$
\mathbf{D}_{N+1}^{\mathsf{T}}\mathbf{W}\mathbf{D}_{1:N} = \mathbf{E}, \tag{36}
$$

$$
\mathbf{D}_{N+1}^{\mathsf{T}}\mathbf{W}\mathbf{f}_{1:N} = \mathbf{f}_{N+1}, \tag{37}
$$

where $\mathbf{W} \in \mathbb{R}^{N \times N}$ is defined as

$$
\mathbf{W} = \begin{bmatrix} w_1 & 0 & \ldots & 0 \\ 0 & w_2 & \ldots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \ldots & 0 & w_n \end{bmatrix}, \tag{38}
$$

and w_i, $i = 1, \ldots, N$ are the LGR quadature weights. Equations (36)–(37) allows us to rewrite Eq. (27) as

$$
\begin{aligned}
J_a = &\sum_{i=1}^{N} w_i \mathcal{L}(\mathbf{X}_i, \mathbf{V}_i, \mathbf{U}_i) \\
&- \sum_{i=1}^{N} \langle \mathbf{\Lambda}_i, \mathbf{D}_{i,1:N}\mathbf{X}_{1:N} + D_{i,N+1}\mathbf{X}_{N+1} - \mathbf{V}_i \rangle \\
&- \sum_{i=1}^{N} \left\langle \tilde{\mathbf{\Lambda}}_i, \mathbf{D}_{i,1:N}\mathbf{V}_{1:N} - \mathbf{f}(\mathbf{X}_i, \mathbf{V}_i, \mathbf{U}_i) \right\rangle \\
&- \sum_{i=1}^{N} \left\langle \tilde{\mathbf{\Lambda}}_i, D_{i,N+1}\mathbf{V}_{N+1} \right\rangle \\
&- \left\langle \tilde{\mathbf{\Lambda}}_{N+1}, \mathbf{D}_{N+1}^{\mathsf{T}}\mathbf{W}\mathbf{D}_{1:N}\mathbf{V}_{1:N} + E_0 \mathbf{V}_{N+1} \right\rangle \\
&- \left\langle \tilde{\mathbf{\Lambda}}_{N+1}, \mathbf{D}_{N+1}^{\mathsf{T}}\mathbf{W}\mathbf{f}(\mathbf{X}_{1:N}, \mathbf{V}_{1:N}, \mathbf{U}_{1:N}) \right\rangle \\
&- \mathbf{\Psi}\mathbf{b}^{\mathsf{T}}(\mathbf{X}_1, \mathbf{V}_1, \mathbf{X}_{N+1}, \mathbf{V}_{N+1}) \\
&+ \mathcal{M}(\mathbf{X}_1, \mathbf{V}_1, \mathbf{X}_{N+1}, \mathbf{V}_{N+1})
\end{aligned} \tag{39}
$$

.

To simplify the following derivation, the function arguments will be removed from the equations. For instance, $\mathcal{L}(\mathbf{X}, \mathbf{V}, \mathbf{U})$ will be expressed as $\mathcal{L}$. The KKT conditions are derived by taking the partial derivatives J_a with respect to $\mathbf{X}$, $\mathbf{V}$, $\mathbf{U}$, $\mathbf{\Lambda}$, and $\mathbf{\Psi}$ and setting them equal to zero. Note that $\frac{\partial J_a}{\partial \mathbf{\Lambda}}$ and $\frac{\partial J_a}{\partial \mathbf{\Psi}}$ result in Eqs. (23)–(25), and

$$
\begin{aligned}
\mathbf{D}_i^{\mathsf{T}}\mathbf{\Lambda} = &\nabla_{\mathbf{X}_{1:N}} \left(w_i \mathcal{L}_i + \left\langle \tilde{\mathbf{\Lambda}}_i + \tilde{\mathbf{\Lambda}}_{N+1} D_{i,N+1} w_i, \mathbf{f}_i \right\rangle \right) \\
&- \delta_{1i}(-\nabla_{\mathbf{X}_1}\mathcal{M} + \nabla_{\mathbf{X}_1}\mathbf{\Psi}\mathbf{b}^{\mathsf{T}}),
\end{aligned} \tag{40}
$$

$$
\mathbf{D}_{N+1}^{\mathsf{T}}\mathbf{\Lambda} = \nabla_{X_{N+1}}\mathcal{M} - \nabla_{\mathbf{X}_{N+1}}\mathbf{\Psi}\mathbf{b}^{\mathsf{T}}, \tag{41}
$$

$$
\begin{aligned}
&\mathbf{D}_i^{\mathsf{T}}\left(\tilde{\mathbf{\Lambda}}_{1:N} + \tilde{\mathbf{\Lambda}}_{N+1} D_{i,N+1} w_i \right) \\
&= \nabla_{\mathbf{V}_{1:N}} \left(w_i \mathcal{L}_i + \left\langle \tilde{\mathbf{\Lambda}}_i + \tilde{\mathbf{\Lambda}}_{N+1} D_{i,N+1} w_i, \mathbf{f}_i \right\rangle \right) \\
&+ \mathbf{\Lambda}_i - \delta_{1i}(-\nabla_{\mathbf{V}_1}\mathcal{M} + \nabla_{\mathbf{V}_1}\mathbf{\Psi}\mathbf{b}^{\mathsf{T}}),
\end{aligned} \tag{42}
$$

$$
\mathbf{D}_{N+1}^{\mathsf{T}}\tilde{\mathbf{\Lambda}}_{1:N} + E_0\tilde{\mathbf{\Lambda}}_{N+1} = \nabla_{\mathbf{V}_{N+1}}\mathcal{M} - \nabla_{\mathbf{V}_{N+1}}\mathbf{\Psi}\mathbf{b}^{\mathsf{T}}, \tag{43}
$$

$$
\mathbf{0} = \nabla_{\mathbf{U}_{1:N}} w_i \left(\mathcal{L}_i - \left\langle \tilde{\mathbf{\Lambda}}_{1:N} + \tilde{\mathbf{\Lambda}}_{N+1}\mathbf{D}_{i,N+1} w_i, \mathbf{f} \right\rangle \right), \tag{44}
$$

where $i=1,\ldots,N$ and δ_{ij} is the Kronecker delta function defined as

$$\delta_{ij}=\begin{cases}1, & i=j\\ 0, & i\neq j.\end{cases} \tag{45}$$

Now propose the change of variables

$$\begin{aligned}\boldsymbol{\lambda}_i &= \frac{\boldsymbol{\Lambda}_i}{w_i}, & (46)\\ \boldsymbol{\lambda}_{N+1} &= \mathbf{D}_{N+1}^{\mathsf{T}}\boldsymbol{\Lambda}_{1:N}, & (47)\\ \boldsymbol{\psi}_i &= \boldsymbol{\Psi}_i, & (48)\\ \tilde{\boldsymbol{\lambda}}_i &= \frac{\tilde{\boldsymbol{\Lambda}}_i}{w_i}+\tilde{\boldsymbol{\Lambda}}_{N+1}D_{i,N+1}, & (49)\\ \tilde{\boldsymbol{\lambda}}_{N+1} &= \mathbf{D}_{N+1}^{\mathsf{T}}\tilde{\boldsymbol{\Lambda}}_{1:N}+\tilde{\boldsymbol{\Lambda}}_{N+1}E_0. & (50)\end{aligned}$$

Note that Eqs. (46)–(48) are the same transformations used for the standard LGR method. Finally, define $\mathbf{D}^{\dagger}\in R^{N\times N}$ such that

$$\begin{aligned}D_{11}^{\dagger} &= -D_{11}-\frac{1}{w_1} & (51)\\ D_{ij}^{\dagger} &= -\frac{w_j}{w_i}D_{ji} \quad \text{otherwise}, & (52)\end{aligned}$$

for $i=j=1,2,\ldots,N$. Note that $\mathbf{D}^{\dagger}$ is the same matrix derived by Garg et al. [5] where it was shown that $\mathbf{D}^{\dagger}$ is the differentiation matrix for the space of polynomials of degree $N-1$. Now the KKT conditions can be rewritten as

$$\begin{aligned}\mathbf{D}_i^{\dagger}\boldsymbol{\lambda}_{1:N} &= -\nabla_{\mathbf{X}_{1:N}}\left(\left\langle\tilde{\boldsymbol{\lambda}}_i,\mathbf{f}_i\right\rangle+\mathcal{L}_i\right)\\ &\quad+\frac{\delta_{1i}}{w_1}\left(-\nabla_{X_1}\left(\mathcal{M}-\boldsymbol{\psi}\mathbf{b}^{\mathsf{T}}\right)-\lambda_1\right), & (53)\\ \mathbf{D}_i^{\dagger}\tilde{\boldsymbol{\lambda}}_{1:N} &= -\nabla_{\mathbf{V}_{1:N}}\left(\left\langle\tilde{\boldsymbol{\lambda}}_i,\mathbf{f}_i\right\rangle+\mathcal{L}_i\right)-\boldsymbol{\lambda}_i\\ &\quad+\frac{\delta_{1i}}{w_1}\left(-\nabla_{V_1}\left(\mathcal{M}-\boldsymbol{\psi}\mathbf{b}^{\mathsf{T}}\right)-\tilde{\lambda}_1\right), & (54)\\ \mathbf{0} &= \nabla_{\mathbf{U}_{1:N}}\left(\mathcal{L}_i-\left\langle\tilde{\boldsymbol{\lambda}}_i,\mathbf{f}_i\right\rangle\right). & (55)\\ \boldsymbol{\lambda}_{N+1} &= \nabla_{\mathbf{X}_{N+1}}\left(\mathcal{M}-\boldsymbol{\psi}\mathbf{b}^{\mathsf{T}}\right), & (56)\\ \tilde{\boldsymbol{\lambda}}_{N+1} &= \nabla_{\mathbf{V}_{N+1}}\left(\mathcal{M}-\boldsymbol{\psi}\mathbf{b}^{\mathsf{T}}\right). & (57)\end{aligned}$$

Equations (17)-(18) allow the terms in the second lines of Eqs. (53)–(54) to vanish which results in Eqs (53)–(57) becoming discrete representations of the continuous time first-order optimality conditions from Eqs. (14)–(20).

IV. Example

In this section the costate estimation method for the modified LGR collocation method is demonstrated on an example that contains a nonsmoothcontrol control. Consider the optimal control problem

$$\mathcal{J}(t)=t_f, \tag{58}$$

subject to the dynamic constraints

$$\dot{x}(t) \;=\; \tfrac{t_f}{2}v(t) \quad,\quad \dot{v}(t) \;=\; \tfrac{t_f}{2}u(t), \tag{59}$$

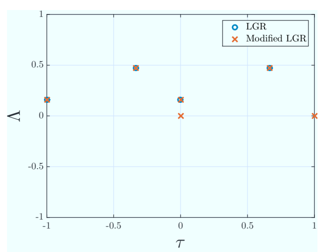
the inequality path constraints

$$(0,-10,-1) \;\leq\; (x(t),v(t),u(t)) \;\leq\; (\infty,10,+1) \tag{60}$$

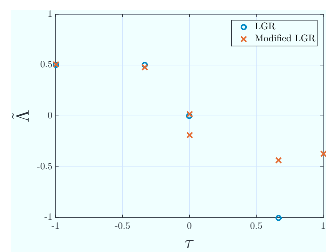
and the boundary conditions

$$(x(-1),x(+1),v(-1),v(+1)) \;=\; (10,0,0,0) \tag{61}$$

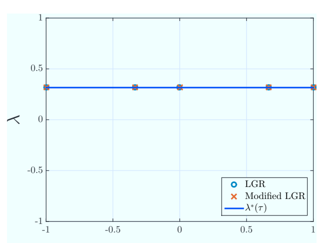
The Lagrange multipliers associated with the collocation constraints associated with the modified Legendre-Gauss-Radau method are shown in Fig. 1a and 1b alongside the Lagrange multipliers of the collocation constraints associated with the standard Legendre-Gauss-Radau method. Using these multipliers together with the costate estimation procedure given in Eqs. (46)–(50), the estimates of the two components of the costate, λ and $\tilde{\lambda}$, obtained from the modified Legendre-Gauss-Radau method are shown in Figs. 1c and 1d. It is seen that the Lagrange multipliers $\boldsymbol{\Lambda}$ and $\tilde{\boldsymbol{\Lambda}}$ differ between the standard and modified Legendre-Gauss-Radau methods, but the costate estimates of the two methods are in excellent agreement. In addition, the estimates of the two components of the costate estimate are in excellent agreement with the optimal costate. This results shows for this example that the modified Legendre-Gauss-Radau method maintains the accuracy of the costate estimate when compared with the standard Legendre-Gauss-Radau collocation method. Finally, Fig. 2 shows control obtained using the modified Legendre-Gauss-Radau method. It is seen that the modified Legendre-Gauss-Radau method finds an accurate approximation of the switch time in the control. Moreover, this switch time uses an extremely sparse mesh (two Legendre-Gauss-Radau points in each of the two mesh intervals). In fact, the analytic solution is known to be a piecewise quadratic. Therefore, an accurate numerical approximation is obtained using the fewest number collocation points possible (two Legendre-Gauss-Radau points in each mesh interval) along with the fewest number of mesh intervals (two).



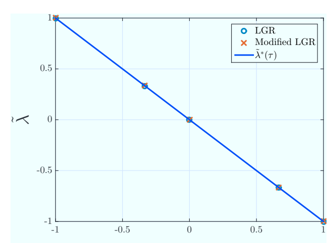
(a) Dual variable, $\boldsymbol{\Lambda}$, for example problem using both the standard and modified LGR collocation methods.



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



(c) Costate Estimate, $\lambda(t)$, for Example using both the standard and modified LGR collocation methods.



(d) Costate Estimate, $\tilde{\lambda}(t)$, for Example using both the standard and modified LGR collocation methods.

Fig. 1: Dual variables and costate estimates for Example using both the standard and modified LGR collocation methods.

V. Conclusions

A modified Legendre-Gauss-Radau collocation method has been developed for solving optimal control problems whose solutions contain a nonsmooth optimal control. The modified

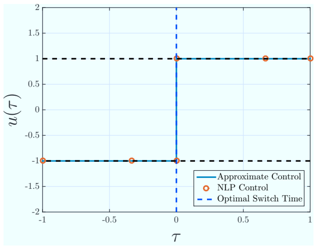


Fig. 2: Optimal Control for the Example Given by Eqs. (58)–(61) Using the Modified LGR Method

method includes the additional constraint where collocation conditions are enforced for each differential equation that is a function of control and a control constraint at the endpoint of the mesh interval that defines the location of nonsmoothness in the solution. The focus of this paper has been on the derivation of the transformed adjoint system associated with the newly developed Legendre-Gauss-Radau method. Using this transformed adjoint system, a transformation has been derived that provides a relationship between the Lagrange multipliers of the nonlinear programming problem and the discrete approximation of the costate of the continuous-time optimal control problem. The effectiveness of the method developed in this paper has been demonstrated on a well known optimal control problem.

Acknowledgments

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

References

[1] P. E. Gill, W. Murray, and M. A. Saunders, "SNOPT: An SQP Algorithm for Large-Scale Constrained Optimization," *SIAM review*, vol. 47, no. 1, pp. 99–131, 2005.

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

[3] D. A. Benson, G. T. Huntington, T. P. Thorvaldsen, and A. V. Rao, "Direct Trajectory Optimization and Costate Estimation via an Orthogonal Collocation Method," *Journal of Guidance, Control, and Dynamics*, vol. 29, no. 6, pp. 1435–1440, November-December 2006.

[4] D. Garg, M. A. Patterson, W. W. Hager, A. V. Rao, D. A. Benson, and G. T. Huntington, "A Unified Framework for the Numerical Solution of Optimal Control Problems Using Pseudospectral Methods," *Automatica*, vol. 46, no. 11, pp. 1843–1851. DOI: 10.1016/j.automatica.2010.06.048, November 2010.

[5] D. Garg, M. A. Patterson, C. L. Darby, C. Francolin, G. T. Huntington, W. W. Hager, and A. V. Rao, "Direct Trajectory Optimization and Costate Estimation of Finite-Horizon and Infinite-Horizon Optimal Control Problems via a Radau Pseudospectral Method," *Computational Optimization and Applications*, vol. 49, no. 2, pp. 335–358. DOI: 10.1007/s10 589–00–09 291–0, June 2011.

[6] D. Garg, W. W. Hager, and A. V. Rao, "Pseudospectral Methods for Solving Infinite-Horizon Optimal Control Problems," *Automatica*, vol. 47, no. 4, pp. 829–837. DOI: 10.1016/j.automatica.2011.01.085, April 2011.

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

[8] G. Elnagar, M. A. Kazemi, and M. Razzaghi, "The Pseudospectral Legendre Method for Discretizing Optimal Control Problems," *IEEE transactions on Automatic Control*, vol. 40, no. 10, pp. 1793–1796, 1995.

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

[10] ——, "Convergence rate for a gauss collocation method applied to unconstrained optimal control," *Journal of Optimization Theory and Applications*, vol. 169, no. 3, pp. 801 – 824, June 2016.

[11] ——, "Lebesgue constants arising in a class of collocation methods," *IMA Journal of Numerical Analysis*, vol. 37, no. 4, pp. 1884–1901, October 2017.

[12] W. W. Hager, H. Hou, S. Mohapatra, and A. V. Rao, "Convergence Rate for an *hp*-Collocation Method Applied to Unconstrained Optimal Control," 2016, arXiv.org/abs/1605.02121.

[13] W. W. Hager, S. Mohapatra, and A. V. Rao, "Convergence Rate for a Gauss Collocation Method applied to Constrained Optimal Control," 2016, arXiv.org/abs/1607.02798.

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

[15] C. L. Darby, W. W. Hager, and A. V. Rao, "An *hp*–Adaptive Pseudospectral Method for Solving Optimal Control Problems," *Optimal Control Applications and Methods*, vol. 32, no. 4, pp. 476–502, 2011.

[16] ——, "Direct Trajectory Optimization Using a Variable Low-Order Adaptive Pseudospectral Method," *Journal of Spacecraft and Rockets*, vol. 48, no. 3, pp. 433–445, May–June 2011.

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

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

[19] ——, "Adaptive Mesh Refinement for Optimal Control Using Nonsmoothness Detection and Mesh Size Reduction," *IEEE Transactions on Control System Technology*, vol. Published Online in Early View, no. 10, pp. 4081–4106, June 2017.

[20] M. Guerra and A. Sarychev, "Measuring Singularity of Generalized Minimizers for Control-Affine Problems," *Journal of Dynamical and Control Systems*, vol. 15, no. 2, pp. 177–221, 2009.

[21] S. Li, R. Zhao, and Q. Zhang, "Optimization method for solving bang-bang and singular control problems," *Journal of Control Theory and Applications*, vol. 10, no. 4, pp. 559–564, 2012.

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

[23] J. E. Cuthrell and L. T. Biegler, "Simultaneous Optimization and Solution Methods for Batch Reactor Control Profiles," *Computers and Chemical Engineering*, vol. 13, no. 1–2, pp. 49–62, January–February 1987.

[24] I. M. Ross and F. Fahroo, "Pseudospectral Knotting Methods for Solving Nonsmooth Optimal Control Problems," *Journal of Guidance Control and Dynamics*, vol. 27, no. 3, pp. 397–405, May–June 2004.

[25] J. Eide and A. V. Rao, "Modified Radau Collocation Method for Solving Optimal Control Problems with Non-Smooth Solutions. Part I: Lavrentiev phenomenon and the search space." Miami Beach, Florida: *2018 IEEE Conference on Decision and Control*, Miami Beach, Florida, 15–17 December 2018. Submitted for Publication, December 2018.