Computational Optimization and Applications (2019) 74:275–314
https://doi.org/10.1007/s10589-019-00100-1

Convergence rate for a Radau hp collocation method applied to constrained optimal control

William W. Hager[1] · **Hongyan Hou**[2] · **Subhashree Mohapatra**[3,4] · **Anil V. Rao**[5] · **Xiang-Sheng Wang**[6]

Received: 25 March 2019 / Published online: 2 May 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019, corrected publication 2019

Abstract
For control problems with control constraints, a local convergence rate is established for an hp-method based on collocation at the Radau quadrature points in each mesh interval of the discretization. If the continuous problem has a sufficiently smooth solution and the Hamiltonian satisfies a strong convexity condition, then the discrete problem possesses a local minimizer in a neighborhood of the continuous solution, and as either the number of collocation points or the number of mesh intervals increase, the discrete solution convergences to the continuous solution in the sup-norm. The convergence is exponentially fast with respect to the degree of the polynomials on each mesh interval, while the error is bounded by a polynomial in the mesh spacing. An advantage of the hp-scheme over global polynomials is that there is a convergence guarantee when the mesh is sufficiently small, while the convergence result for global polynomials requires that a norm of the linearized dynamics is sufficiently small. Numerical examples explore the convergence theory.

Keywords hp Collocation · Radau collocation · Convergence rate · Optimal control · Orthogonal collocation

Mathematics Subject Classification 49M25 · 49M37 · 65K05 · 90C30

October 24, 2017, revised March 25, 2019. The authors gratefully acknowledge support by the Office of Naval Research under Grants N00014-11-1-0068, N00014-15-1-2048 and N00014-18-1-2100, by the National Science Foundation under Grants DMS-1522629, CBET-1404767 and DMS-1819002, and by the U.S. Air Force Research Laboratory under contract FA8651-08-D-0108/0054.

Extended author information available on the last page of the article

1 Introduction

A convergence rate is established for an *hp*-orthogonal collocation method applied to a constrained control problem of the form

$$\left.\begin{array}{lll} \text{minimize} & C(\mathbf{x}(1)) & \\ \text{subject to} & \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), & \mathbf{u}(t) \in \mathcal{U}, \quad t \in \Omega_0, \\ & \mathbf{x}(0) = \mathbf{a}, & (\mathbf{x}, \mathbf{u}) \in \mathcal{C}^1(\Omega_0) \times \mathcal{C}^0(\Omega_0), \end{array}\right\} \tag{1.1}$$

where $\Omega_0 = [0, 1]$, the control constraint set $\mathcal{U} \subset \mathbb{R}^m$ is closed and convex with nonempty interior, the state $\mathbf{x}(t) \in \mathbb{R}^n$, $\dot{\mathbf{x}}$ denotes the derivative of $\mathbf{x}$ with respect to t, $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$, $C : \mathbb{R}^n \to \mathbb{R}$, and $\mathbf{a}$ is the initial condition, which we assume is given; $\mathcal{C}^l(\Omega_0)$ denotes the space of l times continuously differentiable functions mapping Ω_0 to $\mathbb{R}^d$ for some d. The value of d should be clear from context; states and costates always have n components and controls have m components. It is assumed that $\mathbf{f}$ and C are at least continuous.

The development of *hp*-techniques in the context of finite element methods for boundary-value problems began with the work of Gui and Babuška [27–29] and Babuška and Suri [1–3]. In the *hp*-collocation approach that we develop for (1.1), the time domain Ω_0 is initially partitioned into a mesh. To simplify the discussion, we focus on a uniform mesh consisting of K intervals $[t_{k-1}, t_k]$ defined by the mesh points $t_k = k/K$ where $0 \le k \le K$. The dynamics of (1.1) are reformulated using a change of variables. Let $t_{k+1/2} = (t_k + t_{k+1})/2$ be the midpoint of the mesh interval $[t_k, t_{k+1}]$. We make the change of variables $t = t_{k-1/2} + h\tau$, where $h = 1/(2K)$ is half the width of the mesh interval and $\tau \in \Omega := [-1, 1]$; let us define $\mathbf{x}_k : \Omega \to \mathbb{R}^n$ by $\mathbf{x}_k(\tau) = \mathbf{x}(t_{k-1/2} + h\tau)$. Thus $\mathbf{x}_k$ corresponds to the restriction of $\mathbf{x}$ to the mesh interval $[t_{k-1}, t_k]$. Similarly, we define a control $\mathbf{u}_k$ corresponding to the restriction of $\mathbf{u}$ to the mesh interval $[t_{k-1}, t_k]$. In the new variables, the control problem reduces to finding K state-control pairs $(\mathbf{x}_k, \mathbf{u}_k)$, $1 \le k \le K$, each pair defined on the interval $[-1, 1]$, to solve the problem

$$\left.\begin{array}{lll} \text{minimize} & C(\mathbf{x}_K(1)) & \\ \text{subject to} & \dot{\mathbf{x}}_k(\tau) = h\mathbf{f}(\mathbf{x}_k(\tau), \mathbf{u}_k(\tau)), \quad \mathbf{u}_k(\tau) \in \mathcal{U}, & \tau \in \Omega, \\ & \mathbf{x}_k(-1) = \mathbf{x}_{k-1}(1), & 1 \le k \le K, \\ & (\mathbf{x}_k, \mathbf{u}_k) \in \mathcal{C}^1(\Omega) \times \mathcal{C}^0(\Omega). & \end{array}\right\} \tag{1.2}$$

Since the function $\mathbf{x}_0$ does not exist (there is no 0-th mesh interval), we simply define $\mathbf{x}_0(1) = \mathbf{a}$, the initial condition. The condition

$$\mathbf{x}_k(-1) = \mathbf{x}_{k-1}(1) \tag{1.3}$$

in (1.2) corresponds to the initial condition $\mathbf{x}(0) = \mathbf{a}$ when $k = 1$ and to continuity of the state across a mesh interval boundary when $k > 1$. Throughout the paper, (1.3) is referred to as the *continuity condition*.

In the *hp*-scheme developed in this paper, the dynamics for $\mathbf{x}_k$ are approximated by the Radau collocation scheme developed in [11,13,24,25,33]. Let $\mathcal{P}_N$ denote the

space of polynomials of degree at most N defined on the interval Ω, and let $\mathcal{P}_N^n$ denote the n-fold Cartesian product $\mathcal{P}_N \times \cdots \times \mathcal{P}_N$. We analyze a discrete approximation to (1.2) of the form

$$\left.\begin{array}{lll} \text{minimize} & C(\mathbf{x}_K(1)) & \\ \text{subject to} & \dot{\mathbf{x}}_k(\tau_i) = h\mathbf{f}(\mathbf{x}_k(\tau_i), \mathbf{u}_{ki}), \quad 1 \le i \le N, & \mathbf{u}_{ki} \in \mathcal{U}, \\ & \mathbf{x}_k(-1) = \mathbf{x}_{k-1}(1), \quad 1 \le k \le K, & \mathbf{x}_k \in \mathcal{P}_N^n. \end{array}\right\} \tag{1.4}$$

Note that there is no polynomial associated with the control; $\mathbf{u}_{ki}$ corresponds to the value of the control at $t_{k-1/2} + h\tau_i$. In (1.4) the dimension of $\mathcal{P}_N$ is $N+1$ and there are K mesh intervals, so a component of the state variable is chosen from a space of dimension $K(N+1)$. Similarly, there are $KN+K$ equations in (1.4) corresponding to the collocated dynamics at KN points and the K continuity conditions, the initial condition at $t = 0$ and the $K-1$ continuity conditions for the state at the interior mesh points.

For simplicity in the analysis, the same degree polynomials are used in each mesh interval, while in practical implementations of the hp-scheme [11,12,41,43], polynomials of different degrees are often used on different intervals. On intervals where the solution is smooth, high degree polynomials are employed, while on intervals where the solution is nonsmooth, low degree polynomials are used.

We focus on a collocation scheme based on the N Radau quadrature points satisfying

$$-1 < \tau_1 < \tau_2 < \cdots < \tau_N = 1.$$

If P_N denotes the Legendre polynomial of degree N, then the Radau quadrature points are the zeros of $P_N - P_{N-1}$. These quadrature points are sometimes called the flipped Radau points, while the standard Radau points are $-\tau_i$, $1 \le i \le N$. The analysis is the same for either set of points, while the notation is a little cleaner for the flipped points. Besides the N collocation points, our analysis also utilizes the noncollocated point $\tau_0 = -1$ corresponding to the initial condition.

It is pointed out in [34] that for a global collocation scheme where $K = 1$, the discrete dynamics may be infeasible for certain choices of N. In contrast, the analysis in this paper implies that locally, for each choice of the discrete control, there exists a unique discrete state which satisfies the discrete dynamics when K is sufficiently large, or equivalently, when h is sufficiently small, regardless of the choice for N. In this respect, the hp-collocation approach is more robust than a global scheme.

Other global collocation schemes that have been presented in the literature are based on the Lobatto quadrature points [19,22], on the Chebyshev quadrature points [20,23], on the Gauss quadrature points [4,25], and on the extrema of Jacobi polynomials [47]. Kang [39,40] considers control systems in feedback linearizable normal form, and shows that when the Lobatto discretized control problem is augmented with bounds on the states and control, and on certain Legendre polynomial expansion coefficients, then the objectives in the discrete problem converge to the optimal objective of the continuous problem at an exponential rate. Kang's analysis does not involve a coercivity assumption for the continuous problem, but instead imposes bounds in the

discrete problem. Also, in [26] a consistency result is established for a scheme based on Lobatto collocation.

Any of the global schemes could be developed into an hp-collocation scheme. Our rationale for basing our hp-scheme on the Radau collocation points was the following: In numerical experiments such as those in [25], there is often not much difference between the convergence speed of approximations based on either Gauss or Radau collocation, while the Lobatto scheme often converged much slower; and in some cases, the Lobatto costate approximation did not converge due to a null space that arises in the first-order optimality conditions—see [25]. On the other hand, the implementation of an hp-scheme based on the Radau quadrature points was much simpler than the implementation based on the Gauss quadrature points. The Gauss points lie in the interior of each mesh interval, which requires the introduction of the state value at the mesh points. Since one of the Radau points is a mesh point, there is no need to introduce an additional noncollocated point. The implementation ease of Chebyshev quadrature should be similar to that of Gauss and was not pursued. The hp-collocation scheme analyzed in this paper corresponds to the scheme implemented in the popular GPOPS-II software package [45] for solving optimal control problems. This paper, in essence, provides a theoretical justification for the algorithm implemented in the software.

For $\mathbf{x} \in \mathcal{C}^0(\Omega_0)$, we use the sup-norm $\|\cdot\|_\infty$ given by

$$\|\mathbf{x}\|_\infty = \sup\{|\mathbf{x}(t)| : t \in \Omega_0\},$$

where $|\cdot|$ is the Euclidean norm. Given $\mathbf{y} \in \mathbb{R}^n$, the ball with center $\mathbf{y}$ and radius ρ is denoted

$$\mathcal{B}_\rho(\mathbf{y}) = \{\mathbf{x} \in \mathbb{R}^n : |\mathbf{x} - \mathbf{y}| \le \rho\}.$$

The following regularity assumption is assumed to hold throughout the paper.

Smoothness. The problem (1.1) has a local minimizer $(\mathbf{x}^*, \mathbf{u}^*)$ in $\mathcal{C}^1(\Omega_0) \times \mathcal{C}^0(\Omega_0)$. For some $\rho > 0$ and open set $\mathcal{O} \subset \mathbb{R}^{m+n}$ such that

$$\mathcal{B}_\rho(\mathbf{x}^*(t), \mathbf{u}^*(t)) \subset \mathcal{O} \quad \text{for all} \quad t \in \Omega_0,$$

the first two derivative of f and C are Lipschitz continuous on the closure of $\mathcal{O}$ and on $\mathcal{B}_\rho(\mathbf{x}^*(1))$ respectively.

Let $\boldsymbol{\lambda}^*$ denote the solution of the linear costate equation

$$\dot{\boldsymbol{\lambda}}^*(t) = -\nabla_x H(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t)), \quad \boldsymbol{\lambda}^*(1) = \nabla C(\mathbf{x}^*(1)), \tag{1.5}$$

where H is the Hamiltonian defined by $H(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}) = \boldsymbol{\lambda}^{\mathsf{T}} \mathbf{f}(\mathbf{x}, \mathbf{u})$ and ∇ denotes gradient. By the first-order optimality conditions (Pontryagin's minimum principle), we have

$$-\nabla_u H(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t)) \in N_{\mathcal{U}}(\mathbf{u}^*(t)) \quad \text{for all} \quad t \in \Omega_0. \tag{1.6}$$

For any $\mathbf{u} \in \mathcal{U}$,

$$N_{\mathcal{U}}(\mathbf{u}) = \{\mathbf{w} \in \mathbb{R}^m : \mathbf{w}^{\mathsf{T}}(\mathbf{v} - \mathbf{u}) \leq 0 \quad \text{for all} \quad \mathbf{v} \in \mathcal{U}\},$$

while $N_{\mathcal{U}}(\mathbf{u}) = \emptyset$ if $\mathbf{u} \notin \mathcal{U}$.

We will show in Proposition 2.1 that the first-order optimality conditions (Karush–Kuhn–Tucker conditions) for (1.4) are equivalent to the existence of $\boldsymbol{\lambda}_k \in \mathcal{P}^n_{N-1}$, $1 \leq k \leq K$, such that

$$\dot{\boldsymbol{\lambda}}_k(\tau_i) = -h\nabla_x H\left(\mathbf{x}_k(\tau_i), \mathbf{u}_{ki}, \boldsymbol{\lambda}_k(\tau_i)\right), \quad 1 \leq i < N, \tag{1.7}$$

$$\dot{\boldsymbol{\lambda}}_k(1) = -h\nabla_x H\left(\mathbf{x}_k(1), \mathbf{u}_{kN}, \boldsymbol{\lambda}_k(1)\right) + \left[\boldsymbol{\lambda}_k(1) - \boldsymbol{\lambda}_{k+1}(-1)\right] N^2/2,$$

$$\text{where} \quad \boldsymbol{\lambda}_{K+1}(-1) := \nabla C\left(\mathbf{x}_K(1)\right) \tag{1.8}$$

$$N_{\mathcal{U}}(\mathbf{u}_{ki}) \ni -\nabla_u H\left(\mathbf{x}_k(\tau_i), \mathbf{u}_{ki}, \boldsymbol{\lambda}_k(\tau_i)\right), \quad 1 \leq i \leq N. \tag{1.9}$$

Since the $K+1$ mesh interval does not exist, (1.8) includes a definition for $\boldsymbol{\lambda}_{K+1}(-1)$. As we will see in Proposition 2.1, $\boldsymbol{\lambda}_k(-1)$ for $k \leq K$ is the multiplier associated with the continuity condition (1.3). Notice that the system (1.7)–(1.9) for the costate approximation does not contain a continuity condition as in the primal discretization (1.4), so the costate approximation could be discontinuous across the mesh points. Since $\mathcal{P}_{N-1}$ has dimension N and $1 \leq k \leq K$, the approximation to a component of the costate has dimension KN, while (1.7)–(1.8) provides KN equations. Hence, if a continuity condition for the costate were imposed at the mesh points, the system of Eqs. (1.7)–(1.9) along with the continuity condition would be overdetermined.

The following two assumptions are utilized in the convergence analysis.

(A1) The matrix $\nabla^2 C(\mathbf{x}^*(1))$ is positive semidefinite and for some $\alpha > 0$, we have

$$\begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix}^{\mathsf{T}} \nabla^2_{(x,u)} H(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t)) \begin{bmatrix} \mathbf{x} \\ \mathbf{u} \end{bmatrix} \geq \alpha |\mathbf{u}|^2$$

whenever $t \in \Omega_0$, $\mathbf{x} \in \mathbb{R}^n$, and $\mathbf{u} = \mathbf{v} - \mathbf{w}$ for some $\mathbf{v}, \mathbf{w} \in \mathcal{U}$.

(A2) K is large enough, or equivalently h is small enough, that $2hd_1 < 1$ and $2hd_2 < 1$, where

$$d_1 = \sup_{t \in \Omega_0} \|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))\|_\infty \quad \text{and} \quad d_2 = \sup_{t \in \Omega_0} \|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))^{\mathsf{T}}\|_\infty. \tag{1.10}$$

Here $\|\cdot\|_\infty$ is the matrix sup-norm (largest absolute row sum).

The coercivity assumption (A1) ensures that the solution of the discrete problem is a local minimizer. As explained in [37], the condition (A2) is needed to ensure feasibility of the discretized problem (1.4) in a neighborhood of $(\mathbf{x}^*, \mathbf{u}^*)$. In a p scheme where $K = 1$, such as the scheme analyzed in [34], convergence is only guaranteed when $\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))$ is sufficiently small. The convergence theory for the hp-scheme is more robust since (A2) always holds when h is sufficiently small.

Given a local minimizer $(\mathbf{x}^*, \mathbf{u}^*)$ of (1.1), let $\mathbf{x}^*_k$, $\mathbf{u}^*_k$, and $\boldsymbol{\lambda}^*_k$ be the state, control, and costate associated with the mesh interval $[t_{k-1}, t_k]$ and the change of variables $t =$

$t_{k-1/2} + h\tau$, and define $t_{kj} = t_{k-1/2} + h\tau_j$. The domain of $\mathbf{x}_k^*$, $\mathbf{u}_k^*$, or $\boldsymbol{\lambda}_k^*$ is $[-1, +1]$ where -1 corresponds to t_{k-1} and $+1$ corresponds to t_k. We define the following related discrete variables:

$$\left.\begin{aligned} \mathbf{X}_{kj}^* &= \mathbf{x}_k^*(\tau_j) = \mathbf{x}^*(t_{kj}), \quad 0 \le j \le N, \quad 1 \le k \le K, \\ \mathbf{U}_{kj}^* &= \mathbf{u}_k^*(\tau_j) = \mathbf{y}^*(t_{kj}), \quad 1 \le j \le N, \quad 1 \le k \le K, \\ \boldsymbol{\Lambda}_{kj}^* &= \boldsymbol{\lambda}_k^*(\tau_j) = \boldsymbol{\lambda}^*(t_{kj}), \quad 0 \le j \le N, \quad 1 \le k \le K. \end{aligned}\right\} \tag{1.11}$$

Suppose that $\mathbf{x}_k^N \in \mathcal{P}_N^n$, $1 \le k \le K$, is a polynomial which is a stationary point of (1.4) for some discrete controls $\mathbf{u}_k^N$, and suppose that $\boldsymbol{\lambda}_k^N \in \mathcal{P}_{N-1}^n$ satisfy (1.7)–(1.9). We define the following related discrete variables:

$$\begin{aligned} \mathbf{X}_{kj}^N &= \mathbf{x}_k^N(\tau_j), \quad 0 \le j \le N, \quad 1 \le k \le K, \\ \mathbf{U}_{kj}^N &= \mathbf{u}_{kj}^N, \quad\quad 1 \le j \le N, \quad 1 \le k \le K, \\ \boldsymbol{\Lambda}_{kj}^N &= \boldsymbol{\lambda}_k^N(\tau_j), \quad 0 \le j \le N, \quad 1 \le k \le K. \end{aligned}$$

Thus capital letters always refer to discrete variables. As noted earlier, the costate polynomials associated with the discrete problem are typically discontinuous across the mesh points, and $\boldsymbol{\Lambda}_{kN}^N \ne \boldsymbol{\Lambda}_{k+1,0}^N$.

The convergence analysis only involves the smoothness of the optimal state and associated costate on the interior of each mesh interval. Let $\mathcal{H}^p(a, b)$ denote the Sobolev space of functions with square integrable derivatives on (a, b) through order p. Let $\mathcal{PH}^p(\Omega_0)$ denote the space of continuous functions whose restrictions to (t_{k-1}, t_k) are contained in $\mathcal{H}^p(t_{k-1}, t_k)$ for each k between 1 and K (piecewise $\mathcal{H}^p$). The norm on $\mathcal{PH}^p(\Omega_0)$ is the same as the norm on $\mathcal{H}^p(\Omega_0)$ except that the integral is computed over the interior of each mesh interval. In this paper, the error bounds are expressed in terms of a seminorm $|\cdot|_{\mathcal{PH}^p(\Omega_0)}$ which only involves the p-th order derivative:

$$|\mathbf{x}|_{\mathcal{PH}^p(\Omega_0)} = \left(\sum_{k=1}^{K} \int_{t_{k-1}}^{t_k} \left| \frac{d^p \mathbf{x}(t)}{dt^p} \right|^2 dt \right)^{1/2}.$$

The following convergence result relative to the vector sup-norm (largest absolute element) will be established.

Theorem 1.1 *Suppose that* (A1) *and* (A2) *hold. If* $(\mathbf{x}^*, \mathbf{u}^*)$ *is a local minimizer for the continuous problem* (1.1) *with* $\mathbf{x}^*$ *and* $\boldsymbol{\lambda}^* \in \mathcal{PH}^\eta(\Omega_0)$ *for some* $\eta \ge 2$*, then for* N *sufficiently large or for* h *sufficiently small with* $N \ge 2$*, the discrete problem* (1.4) *has a local minimizer and associated multiplier satisfying* (1.7)–(1.9)*, and we have*

$$\begin{aligned} &\max \left\{ \left\| \mathbf{X}^N - \mathbf{X}^* \right\|_\infty, \left\| \mathbf{U}^N - \mathbf{U}^* \right\|_\infty, \left\| \boldsymbol{\Lambda}^N - \boldsymbol{\Lambda}^* \right\|_\infty \right\} \\ &\quad \le h^{p-1} \left(\frac{c}{N} \right)^{p-1} |\mathbf{x}^*|_{\mathcal{PH}^p(\Omega_0)} + h^{q-1} \left(\frac{c}{N} \right)^{q-1.5} |\boldsymbol{\lambda}^*|_{\mathcal{PH}^q(\Omega_0)}, \end{aligned} \tag{1.12}$$

where $p = \min(\eta, N+1)$, $q = \min(\eta, N)$, *and* c *is independent of* h, N, *and* η.

The proof of Theorem 1.1 begins in Section 2 where the discrete first-order optimality conditions are formulated as an inclusion of the form $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$. In Section 4 a bound is obtained for the distance d^* from $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ to $\mathcal{F}(\mathbf{U}^*)$, where $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ denotes the optimal discrete variables defined in (1.11). This bound is based on an estimate given in Sect. 3 for the $\mathcal{H}^1$ approximation error of the polynomial that interpolates $\mathbf{x}^*$ at τ_i, $0 \le i \le N$. The remainder of the paper focuses on showing that the bound for d^* is also a bound for the distance from $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ to a solution of the inclusion $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$. The analysis is based on Proposition 2.2, where it is shown that such a bound can be obtained if a linearized version of the original inclusion is stable under perturbations. More precisely, we need to show that the problem of finding $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ such that

$$\nabla \mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] + \mathbf{Y} \in \mathcal{F}(\mathbf{U})$$

has a unique solution which depends Lipschitz continuously on the perturbation $\mathbf{Y}$. This analysis begins in Section 5 where perturbations in the linearized state and costate discrete dynamics are analyzed. In Section 6 it is shown that solving the linearized inclusion is equivalent to solving a quadratic program, where perturbations in the inclusion appear as linear terms in the quadratic program; assumptions (A1) and (A2) imply the existence of a unique solution to the quadratic program, which in turn implies the existence of a unique solution to the inclusion. Finally, in Section 7 the unique solution of the linearized inclusion is shown to depend Lipschitz continuously on the perturbation. This Lipschitz property and the bound for d^* are combined with Proposition 2.2 to obtain (1.12). The tightness and possible extensions of the error bound (1.12) are explored in Section 8 using some problems with known solutions. In the proof of Theorem 1.1, we need to make the right side of (1.12) sufficiently small to establish the existence of the claimed solution to the discrete problem. The conditions $\eta \ge 2$ and $N \ge 2$ in the statement of the theorem ensure that h^{p-1} and h^{q-1} go to zero as h goes to zero, and $(c/N)^{p-1}$ and $(c/N)^{q-1.5}$ go to zero as N tends to infinity.

Since the discrete costate could be discontinuous across a mesh point, Theorem 1.1 implies convergence of the discrete costate on either side of the mesh point to the continuous costate at the mesh point. The discrete problem provides an estimate for the optimal control at $t = 1$ in the continuous problem, but not at $t = 0$ since this is not a collocation point. Due to the strong convexity assumption (A1), an estimate for the discrete control at $t = 0$ can be obtained from the minimum principle (1.6) since the initial state is given, while we have an estimate for the associated costate at $t = 0$. Alternatively, polynomial interpolation could be used to obtain estimates for the optimal control at $t = 0$.

In a recent paper [36], where we analyze a Gauss collocation scheme on a single interval, $p = q = \min(\eta, N + 1)$. The differences between Radau and Gauss collocation are due to the asymmetry of the Radau points, and the asymmetry in the Radau first-order optimality conditions; that is, for the Radau points, $\boldsymbol{\lambda}_k \in \mathcal{P}^n_{N-1}$ while $\mathbf{x}_k \in \mathcal{P}^n_N$.

Notation. We let Ω denote the interval $[-1, 1]$, while Ω_0 is the interval $[0, 1]$. Let $\mathcal{P}_N$ denote the space of polynomials of degree at most N, while $\mathcal{P}^0_N$ is the subspace consisting of polynomials in $\mathcal{P}_N$ that vanish at $t = -1$ and $t = 1$. The meaning of

the norm $\|\cdot\|_\infty$ is based on context. If $\mathbf{x} \in \mathcal{C}^0(\Omega)$, then $\|\mathbf{x}\|_\infty$ denotes the maximum of $|\mathbf{x}(t)|$ over $t \in [-1, 1]$, where $|\cdot|$ is the Euclidean norm. For a vector $\mathbf{v} \in \mathbb{R}^m$, $\|\mathbf{v}\|_\infty$ is the maximum of $|v_i|$ over $1 \le i \le m$. If $\mathbf{A} \in \mathbb{R}^{m\times n}$, then $\|\mathbf{A}\|_\infty$ is the largest absolute row sum (the matrix norm induced by the ℓ_∞ vector norm). We let $|\mathbf{A}|$ denote the matrix norm induced by the Euclidean vector norm. Throughout the paper, the index k is used for the mesh interval, while the indices i and j are associated with collocation points. If $\mathbf{p} \in \mathbb{R}^{KNn}$, then $\mathbf{p}_k$ for $1 \le k \le K$ refers to vector with components $\mathbf{p}_{kj} \in \mathbb{R}^n$, for $1 \le j \le N$. The dimension of the identity matrix $\mathbf{I}$ is often clear from context; when necessary, the dimension of $\mathbf{I}$ is specified by a subscript. For example, $\mathbf{I}_n$ is the n by n identity matrix. The gradient is denoted ∇, while ∇^2 denotes the Hessian; subscripts indicate the differentiation variables. Throughout the paper, c is a generic constant which has different values in different equations. The value of c is always independent of h, N, and η. The vector $\mathbf{1}$ has all entries equal to one, while the vector $\mathbf{0}$ has all entries equal to zero; again, their dimension should be clear from context. If $\mathbf{D}$ is a matrix, then $\mathbf{D}_j$ is the j-th column of $\mathbf{D}$ and $\mathbf{D}_{i:j}$ is the submatrix formed by columns i through j. We let $\otimes$ denote the Kronecker product. If $\mathbf{U} \in \mathbb{R}^{m\times n}$ and $\mathbf{V} \in \mathbb{R}^{p\times q}$, then $\mathbf{U} \otimes \mathbf{V}$ is the mp by nq block matrix whose (i, j) block is $u_{ij}\mathbf{V}$. We let $\mathcal{L}^2(\Omega)$ denote the usual space of square integrable functions on Ω, while $\mathcal{H}^p(\Omega)$ is the Sobolev space consisting of functions with square integrable derivatives through order p. The seminorm in $\mathcal{H}^p(\Omega)$, corresponding to the $\mathcal{L}^2(\Omega)$ norm of the p-order derivatives, is denoted $|\cdot|_{\mathcal{H}^p(\Omega)}$. The subspace of $\mathcal{H}^1(\Omega)$ corresponding to functions that vanish at $t = -1$ and $t = 1$ is denoted $\mathcal{H}^1_0(\Omega)$.

2 Abstract setting

The Lagrange basis functions associated with the τ_i are

$$L_j(\tau) := \prod_{\substack{i=0\\ i\neq j}}^{N} \frac{\tau - \tau_i}{\tau_j - \tau_i}, \quad 0 \le j \le N.$$

Any $p \in \mathcal{P}_N$ has the expansion

$$p(t) = \sum_{j=0}^{N} p_j L_j(t), \quad p_j = p(\tau_j).$$

Differentiating this identity and evaluating at τ_i gives

$$\dot{p}(\tau_i) = \sum_{j=0}^{N} p_j D_{ij}, \quad D_{ij} = \dot{L}_j(\tau_i), \quad 1 \le i \le N. \tag{2.1}$$

The matrix $\mathbf{D}$ is called a differentiation matrix. Given a feasible point for the discrete problem (1.4), define $\mathbf{X}_{kj} = \mathbf{x}_k(\tau_j)$ and $\mathbf{U}_{ki} = \mathbf{u}_{ki}$. It follows from (2.1) that

$$\sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj} = \dot{\mathbf{x}}_k(\tau_i), \quad 1 \le i \le N.$$

Hence, the discrete problem (1.4) can be reformulated as

$$\left.\begin{array}{lll} \text{minimize} & C(\mathbf{X}_{KN}) & \\ \text{subject to} & \sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj} = h\mathbf{f}(\mathbf{X}_{ki}, \mathbf{U}_{ki}), \quad \mathbf{U}_{ki} \in \mathcal{U}, & 1 \le i \le N, \\ & \mathbf{X}_{k0} = \mathbf{X}_{k-1,N}, & 1 \le k \le K, \end{array}\right\} \tag{2.2}$$

where $\mathbf{X}_{0N} = \mathbf{a}$, the initial condition.

We introduce multipliers $\boldsymbol{\mu}_{ki}$ associated with the constraints in (2.2) and write the Lagrangian as

$$\mathcal{L}(\boldsymbol{\mu}, \mathbf{X}, \mathbf{U}) = C\left(\mathbf{X}_{KN}\right) + \sum_{k=1}^{K}\sum_{i=1}^{N}\left\langle \boldsymbol{\mu}_{ki}, h\mathbf{f}(\mathbf{X}_{ki}, \mathbf{U}_{ki}) - \sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj}\right\rangle + \sum_{k=1}^{K}\left\langle \boldsymbol{\mu}_{k0}, \left(\mathbf{X}_{k-1,N} - \mathbf{X}_{k0}\right)\right\rangle.$$

The first-order optimality conditions for (2.2), often called the Karush–Kuhn–Tucker (KKT) conditions, lead to the following relations (we show the variable with which we differentiate the Lagrangian followed by the associated condition):

$$\mathbf{X}_{k0} \;\Rightarrow \sum_{i=1}^{N} D_{i0}\boldsymbol{\mu}_{ki} = -\boldsymbol{\mu}_{k0}, \tag{2.3}$$

$$\mathbf{X}_{kj} \;\Rightarrow \sum_{i=1}^{N} D_{ij}\boldsymbol{\mu}_{ki} = h\nabla_x H(\mathbf{X}_{kj}, \mathbf{U}_{kj}, \boldsymbol{\mu}_{kj}), \quad 1 \le j < N, \tag{2.4}$$

$$\mathbf{X}_{kN} \Rightarrow \sum_{i=1}^{N} D_{iN}\boldsymbol{\mu}_{ki} = h\nabla_x H(\mathbf{X}_{kN}, \mathbf{U}_{kN}, \boldsymbol{\mu}_{kN}) + \boldsymbol{\mu}_{k+1,0}, \tag{2.5}$$

$$\boldsymbol{\mu}_{K+1,0} := \nabla C(\mathbf{X}_{KN}), \tag{2.6}$$

$$\mathbf{U}_{ki} \;\Rightarrow -\nabla_u H\left(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \boldsymbol{\mu}_{ki}\right) \in N_{\mathcal{U}}(\mathbf{U}_{ki}). \tag{2.7}$$

The KKT multipliers in (2.3)–(2.7) are connected to the polynomials satisfying (1.7)–(1.9) through the Radau quadrature weights ω_i, $1 \le i \le N$. These are positive numbers that sum to 2 and have the property [46, Thm. 3.26] that

$$\int_{-1}^{1} p(\tau)d\tau = \sum_{i=1}^{N} \omega_i\, p(\tau_i)$$

for every $p \in \mathcal{P}_{2N-2}$. Note that $\omega_N = 2/N^2$, which appears in (1.8).

Proposition 2.1 *The multipliers $\boldsymbol{\mu}_k \in \mathbb{R}^{Nn}$ satisfy (2.3)–(2.7) if and only if the polynomial $\boldsymbol{\lambda}_k \in \mathcal{P}^n_{N-1}$ given by $\boldsymbol{\lambda}_k(\tau_i) = \boldsymbol{\mu}_{ki}/\omega_i$, $1 \le i \le N$, satisfies (1.7)–(1.9). Moreover, $\boldsymbol{\mu}_{k0} = \boldsymbol{\lambda}_k(-1)$.*

Proof We start with multipliers $\boldsymbol{\mu}_k$ satisfying (2.3)–(2.7). Define $\boldsymbol{\Lambda}_{ki} = \boldsymbol{\mu}_{ki}/\omega_i$ for $1 \le i \le N$, and let $\boldsymbol{\lambda}_k \in \mathcal{P}^n_{N-1}$ be the polynomial that satisfies $\boldsymbol{\lambda}_k(\tau_i) = \boldsymbol{\Lambda}_{ki}$. Also, set $\boldsymbol{\Lambda}_{k0} = \boldsymbol{\mu}_{k0}$. In terms of $\boldsymbol{\Lambda}_{ki}$ and the matrix $D^{\ddagger}_{ij} = -\omega_j D_{ji}/\omega_i$, the Eqs. (2.4), (2.5) and (2.7) become

$$\sum_{j=1}^{N} D^{\ddagger}_{ij} \boldsymbol{\Lambda}_{kj} = -h \nabla_x H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \boldsymbol{\Lambda}_{ki}), \quad 1 \le i < N, \tag{2.8}$$

$$\sum_{i=1}^{N} D^{\ddagger}_{Ni} \boldsymbol{\Lambda}_{ki} = -[h \nabla_x H(\mathbf{X}_{kN}, \mathbf{U}_{kN}, \boldsymbol{\Lambda}_{kN}) + \boldsymbol{\Lambda}_{k+1,0}/\omega_N], \tag{2.9}$$

$$N_{\mathcal{U}}(\mathbf{U}_{ki}) \ni -\nabla_u H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \boldsymbol{\Lambda}_{ki}), \quad 1 \le i \le N. \tag{2.10}$$

Since the polynomial that is identically equal to $\mathbf{1}$ has derivative $\mathbf{0}$ and since $\mathbf{D}$ is a differentiation matrix, we have $\mathbf{D1} = \mathbf{0}$, which implies that $\mathbf{D}_0 = -\sum_{j=1}^{N} \mathbf{D}_j$, where $\mathbf{D}_j$ is the j-th column of $\mathbf{D}$. Hence, the first definition in (2.3) can be written

$$\boldsymbol{\Lambda}_{k0} = -\sum_{i=1}^{N} \boldsymbol{\mu}_{ki} D_{i0} = \sum_{i=1}^{N}\sum_{j=1}^{N} \boldsymbol{\mu}_{ki} D_{ij} = \sum_{i=1}^{N}\sum_{j=1}^{N} \omega_j \left(\frac{\boldsymbol{\mu}_{ki}}{\omega_i}\right)(\omega_i D_{ij}/\omega_j)$$

$$= -\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i D^{\ddagger}_{ij} \boldsymbol{\Lambda}_{kj} \tag{2.11}$$

$$= \boldsymbol{\Lambda}_{k+1,0} + h \sum_{i=1}^{N} \omega_i \nabla_x H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \boldsymbol{\Lambda}_{ki}), \tag{2.12}$$

where (2.12) is due to (2.8)–(2.9).

In Section 4.2.1 of [25], we introduce a matrix $\mathbf{D}^{\dagger}$ which is a differentiation matrix for the collocation points τ_i, $1 \le i \le N$. That is, if p is a polynomial of degree at most $N-1$ and $\mathbf{p}$ is the vector with components $p(\tau_i)$, $1 \le i \le N$, then $(\mathbf{D}^{\dagger}\mathbf{p})_i = \dot{p}(\tau_i)$. The matrix $\mathbf{D}^{\ddagger}$ only differs from $\mathbf{D}^{\dagger}$ in a single entry: $D^{\ddagger}_{NN} = D^{\dagger}_{NN} - 1/\omega_N$. As a result,

$$(\mathbf{D}^{\ddagger}\mathbf{p})_i = \dot{p}(\tau_i), \quad 1 \le i < N, \quad (\mathbf{D}^{\ddagger}\mathbf{p})_N = \dot{p}(\tau_N) - p(1)/\omega_N. \tag{2.13}$$

It follows that

$$\sum_{j=1}^{N} D^{\ddagger}_{ij} \boldsymbol{\Lambda}_{kj} = \dot{\boldsymbol{\lambda}}_k(\tau_i), \quad 1 \le i < N, \quad \text{and} \tag{2.14}$$

$$\sum_{j=1}^{N} D_{Nj}^{\ddagger} \mathbf{\Lambda}_{kj} = \dot{\boldsymbol{\lambda}}_k(1) - \boldsymbol{\lambda}_k(1)/\omega_N. \tag{2.15}$$

This substitution in (2.11) yields

$$\mathbf{\Lambda}_{k0} = \boldsymbol{\lambda}_k(1) - \sum_{i=1}^{N} \omega_i \dot{\boldsymbol{\lambda}}_k(\tau_i). \tag{2.16}$$

Since $\dot{\boldsymbol{\lambda}}_k \in \mathcal{P}_{N-2}^n$ and N-point Radau quadrature is exact for these polynomial, we have

$$\sum_{i=1}^{N} \omega_i \dot{\boldsymbol{\lambda}}_k(\tau_i) = \int_{-1}^{1} \dot{\boldsymbol{\lambda}}_k(\tau) d\tau = \boldsymbol{\lambda}_k(1) - \boldsymbol{\lambda}_k(-1). \tag{2.17}$$

Combine (2.16) and (2.17) to obtain

$$\mathbf{\Lambda}_{k0} = \boldsymbol{\lambda}_k(-1). \tag{2.18}$$

Let $\mathbf{x}_k \in \mathcal{P}_N^n$ be the polynomial that satisfies $\mathbf{x}_k(\tau_j) = \mathbf{X}_{kj}$ for all $0 \le j \le N$. By (2.14), (1.7) is equivalent to (2.8) which is equivalent to (2.4) after a change of variables. By (2.15) and (2.18), (1.8) is equivalent to (2.9), which is equivalent to (2.5) after a change of variables. Finally, (1.9) is the same as (2.10) which is equivalent to (2.7) after a change of variables. The equivalence between $\mathbf{\Lambda}_{k0}$ and $\boldsymbol{\lambda}_k(-1)$ was derived in (2.18). This shows that the polynomial $\boldsymbol{\lambda}_k(\tau)$ satisfies (1.7)–(1.9). The converse of the proposition follows by reversing all the steps in the derivation. □

The dynamics for (2.2), the first-order optimality conditions (2.8)–(2.10), the formula (2.12) for $\mathbf{\Lambda}_{k0}$, and the terminal costate condition (2.6) can be written as $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$ where

$$\mathcal{T}_{1ki}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = \left(\sum_{j=0}^{N} D_{ij} \mathbf{X}_{kj} \right) - h\mathbf{f}(\mathbf{X}_{ki}, \mathbf{U}_{ki}), \quad 1 \le i \le N, \tag{2.19}$$

$$\mathcal{T}_{2k}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = \mathbf{X}_{k0} - \mathbf{X}_{k-1,N}, \tag{2.20}$$

$$\mathcal{T}_{3ki}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = \left(\sum_{j=1}^{N} D_{ij}^{\ddagger} \mathbf{\Lambda}_{kj} \right) + h\nabla_x H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}), \quad 1 \le i < N, \tag{2.21}$$

$$\mathcal{T}_{3kN}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = \sum_{j=1}^{N} D_{Nj}^{\ddagger} \mathbf{\Lambda}_{kj} + h\nabla_x H(\mathbf{X}_{kN}, \mathbf{U}_{kN}, \mathbf{\Lambda}_{kN}) + \mathbf{\Lambda}_{k+1,0}/\omega_N, \tag{2.22}$$

$$\mathcal{T}_{4k}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = \mathbf{\Lambda}_{k0} - \mathbf{\Lambda}_{k+1,0} - h \sum_{i=1}^{N} \omega_i \nabla_x H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}), \tag{2.23}$$

$$\mathcal{T}_5(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = \nabla C(\mathbf{X}_{KN}) - \mathbf{\Lambda}_{K+1,0}, \tag{2.24}$$

$$\mathcal{T}_{6ki}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = -h\nabla_u H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}), \quad 1 \le i \le N, \tag{2.25}$$

where $1 \le k \le K$. The initial state is $\mathbf{X}_{0N} = \mathbf{X}_{10} = \mathbf{a}$. The components of $\mathcal{F}$ are given by

$$\mathcal{F}_1 = \mathcal{F}_2 = \mathcal{F}_3 = \mathcal{F}_4 = \mathcal{F}_5 = \mathbf{0}, \quad \text{and } \mathcal{F}_{6ki}(\mathbf{U}) = N_{\mathcal{U}}(\mathbf{U}_{ki}).$$

The proof of Theorem 1.1 is based on [18, Proposition 3.1], given below in a slightly simplified form. Other results like this are contained in Theorem 3.1 of [17], in Proposition 5.1 of [31], in Theorem 2.1 of [32], and in Theorem 1 of [30].

Proposition 2.2 *Let $\mathcal{X}$ be a Banach space and let $\mathcal{Y}$ be a linear normed space with the norms in both spaces denoted $\|\cdot\|$. Let $\mathcal{F} : \mathcal{X} \mapsto 2^{\mathcal{Y}}$ and let $\mathcal{T} : \mathcal{X} \mapsto \mathcal{Y}$ with $\mathcal{T}$ continuously Fréchet differentiable in $B_r(\boldsymbol{\theta}^*)$, the ball with center $\boldsymbol{\theta}^*$ and radius r, for some $\boldsymbol{\theta}^* \in \mathcal{X}$ and $r > 0$. Suppose that the following conditions hold for some $\boldsymbol{\delta} \in \mathcal{Y}$ and scalars ϵ and $\gamma > 0$:*

(C1) $\mathcal{T}(\boldsymbol{\theta}^*) + \boldsymbol{\delta} \in \mathcal{F}(\boldsymbol{\theta}^*)$.
(C2) $\|\nabla\mathcal{T}(\boldsymbol{\theta}) - \nabla\mathcal{T}(\boldsymbol{\theta}^*)\| \le \epsilon$ *for all* $\boldsymbol{\theta} \in B_r(\boldsymbol{\theta}^*)$.
(C3) *The map* $(\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*))^{-1}$ *is single-valued and Lipschitz continuous with Lipschitz constant* γ.

If $\epsilon\gamma < 1$ and $\|\boldsymbol{\delta}\| \le (1 - \gamma\epsilon)r/\gamma$, then there exists a unique $\boldsymbol{\theta} \in B_r(\boldsymbol{\theta}^*)$ such that $\mathcal{T}(\boldsymbol{\theta}) \in \mathcal{F}(\boldsymbol{\theta})$. Moreover, we have the estimate

$$\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \le \frac{\gamma}{1 - \gamma\epsilon}\|\boldsymbol{\delta}\|. \tag{2.26}$$

Proof Define $\Phi(\boldsymbol{\theta}) = [\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]^{-1}[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta})$. For all $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2 \in B_r(\boldsymbol{\theta}^*)$, a Taylor expansion with integral remainder term yields

$$[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_2) = [\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_1) + \int_0^1 [\nabla\mathcal{T}(\boldsymbol{\theta}_1 + s(\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1)) - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]\, ds\ (\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1).$$

By (C2), it follows that

$$\|[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)])(\boldsymbol{\theta}_2) - [\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_1)\| \le \epsilon\|\boldsymbol{\theta}_2 - \boldsymbol{\theta}_1\|. \tag{2.27}$$

By (C3) and (2.27), we have

$$\begin{aligned}
&\|\Phi(\boldsymbol{\theta}_1) - \Phi(\boldsymbol{\theta}_2)\| \\
&\quad = \|[\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]^{-1}[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_1) \\
&\qquad -[\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]^{-1}[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_2)\| \\
&\quad \le \gamma\|[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_1) - [\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}_2)\|
\end{aligned}$$

$$\leq \epsilon\gamma \|\boldsymbol{\theta}_1 - \boldsymbol{\theta}_2\|.$$

Since $\epsilon\gamma < 1$, Φ is a contraction on $B_r(\boldsymbol{\theta}^*)$. Subtracting $\nabla\mathcal{T}(\boldsymbol{\theta}^*)(\boldsymbol{\theta}^*)$ from each side of (C1) gives

$$[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]\boldsymbol{\theta}^* + \boldsymbol{\delta} \in [\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}^*),$$

and utilizing the uniqueness in (C3) yields

$$\boldsymbol{\theta}^* = [\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]^{-1}[(\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*))\boldsymbol{\theta}^* + \boldsymbol{\delta}].$$

With this substitution, it follows from (2.27), (C3), and (C2) that

$$\begin{aligned}\|\Phi(\boldsymbol{\theta}) - \boldsymbol{\theta}^*\| &= \|[\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]^{-1}[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}) \\ &\quad -[\mathcal{F} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)]^{-1}[(\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*))(\boldsymbol{\theta}^*) + \boldsymbol{\delta}]\| \\ &\leq \gamma\|[\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}) - [\mathcal{T} - \nabla\mathcal{T}(\boldsymbol{\theta}^*)](\boldsymbol{\theta}^*) - \boldsymbol{\delta}]\| \\ &\leq \gamma(\epsilon\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| + \|\boldsymbol{\delta}\|) \leq \gamma(\epsilon r + \|\boldsymbol{\delta}\|) \end{aligned} \tag{2.28}$$

for all $\boldsymbol{\theta} \in B_r(\boldsymbol{\theta}^*)$. The assumption that $\|\boldsymbol{\delta}\| \leq (1 - \gamma\epsilon)r/\gamma$ can be rearranged to obtain $\gamma(\epsilon r + \|\boldsymbol{\delta}\|) \leq r$, which implies that $\|\Phi(\boldsymbol{\theta}) - \boldsymbol{\theta}^*\| \leq r$ by (2.28). Since Φ maps $B_r(\boldsymbol{\theta}^*)$ into itself and Φ is a contraction on $B_r(\boldsymbol{\theta}^*)$, the contraction mapping principle yields the existence of a unique fixed point $\boldsymbol{\theta} \in B_r(\boldsymbol{\theta}^*)$. Since $\|\Phi(\boldsymbol{\theta}) - \boldsymbol{\theta}^*\| = \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|$ for this fixed point, (2.26) is a consequence of (2.28). □

We use Proposition 2.2 with $\boldsymbol{\theta}^* = (\mathbf{X}^*, \mathbf{U}^*, \boldsymbol{\Lambda}^*)$ defined in (1.11) and $\boldsymbol{\theta} = (\mathbf{X}^N, \mathbf{U}^N, \boldsymbol{\Lambda}^N)$. The norm on $\mathcal{X}$ is given by

$$\|\boldsymbol{\theta}\| = \|(\mathbf{X}, \mathbf{U}, \boldsymbol{\Lambda})\|_\infty = \max\{\|\mathbf{X}\|_\infty, \|\mathbf{U}\|_\infty, \|\boldsymbol{\Lambda}\|_\infty\}. \tag{2.29}$$

The space $\mathcal{Y}$ corresponds to the codomain of $\mathcal{T}$. If $\mathbf{y} \in \mathcal{Y}$, then we let $\mathbf{y}_l$ denote the part of $\mathbf{y}$ associated with $\mathcal{T}_l$, $1 \leq l \leq 6$. The norm of $\mathbf{y} \in \mathcal{Y}$ is given by

$$\|\mathbf{y}\|_{\mathcal{Y}} = \|\mathbf{y}_1\|_\omega + |\mathbf{y}_2| + \|\mathbf{y}_3\|_\omega + |\mathbf{y}_4| + h^{1/2}|\mathbf{y}_5| + h^{-1/2}\|\mathbf{y}_6\|_\infty,$$

where for $\mathbf{z} \in \mathbb{R}^{KNn}$, the ω-norm is defined by

$$\|\mathbf{z}\|_\omega = \left(\sum_{k=1}^{K}\sum_{i=1}^{N} \omega_i |\mathbf{z}_{ki}|^2\right)^{1/2}, \quad \mathbf{z}_{ki} \in \mathbb{R}^n.$$

For $\mathbf{z} \in \mathbb{R}^{Nn}$, the ω-norm is

$$\|\mathbf{z}\|_\omega = \left(\sum_{i=1}^{N} \omega_i |\mathbf{z}_i|^2\right)^{1/2}, \quad \mathbf{z}_i \in \mathbb{R}^n.$$

3 Interpolation error in $\mathcal{H}^1$

Our estimate for the distance from $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ to $\mathcal{F}(\mathbf{U}^*)$ utilizes the following bound for the $\mathcal{H}^1(\Omega)$ error of the interpolant based on the point set τ_i, $0 \le i \le N$.

Lemma 3.1 *If $u \in \mathcal{H}^\eta(\Omega)$ for some $\eta \ge 2$, then there exists a constant c, independent of N and η, such that*

$$|u - u^I|_{\mathcal{H}^1(\Omega)} \le (c/N)^{p-1} |u|_{\mathcal{H}^p(\Omega)}, \quad p = \min\{\eta, N+1\}, \tag{3.1}$$

where $u^I \in \mathcal{P}_N$ is the interpolant of u satisfying $u^I(\tau_i) = u(\tau_i)$, $0 \le i \le N$, and $N > 0$. In the case $\eta = p = 1$, there exists a constant c, independent of N, such that

$$|u - u^I|_{\mathcal{H}^1(\Omega)} \le c|u|_{\mathcal{H}^1(\Omega)}, \tag{3.2}$$

Proof Throughout the analysis, c denotes a generic constant whose value is independent of N and η, and which may have different values in different equations. We first show that if the lemma holds for all $u \in \mathcal{H}_0^1(\Omega) \cap \mathcal{H}^\eta(\Omega)$, then it holds for all $u \in \mathcal{H}^\eta(\Omega)$. Suppose $u \in \mathcal{H}^\eta(\Omega)$ and let ℓ denote the linear function for which $\ell(\pm 1) = u(\pm 1)$. Since $\ell^I = \ell$, it follows that

$$|u - u^I|_{\mathcal{H}^1(\Omega)} = |(u - \ell) - (u - \ell)^I|_{\mathcal{H}^1(\Omega)}.$$

Since $u - \ell \in \mathcal{H}_0^1(\Omega)$, (3.1) gives

$$|u - u^I|_{\mathcal{H}^1(\Omega)} \le (c/N)^{p-1} |u - \ell|_{\mathcal{H}^p(\Omega)}$$

when $\eta \ge 2$. Moreover, when $\eta \ge 2$, $|u - \ell|_{\mathcal{H}^p(\Omega)} = |u|_{\mathcal{H}^p(\Omega)}$ since derivatives of order two or larger applied to the linear function ℓ are zero. This establishes (3.1) for all $u \in \mathcal{H}^\eta(\Omega)$ with $\eta \ge 2$. If $\eta = 1$, then by (3.2), we have

$$|u - u^I|_{\mathcal{H}^1(\Omega)} \le c|u - \ell|_{\mathcal{H}^1(\Omega)} \le c\left(|u|_{\mathcal{H}^1(\Omega)} + (|\ell|_{\mathcal{H}^1(\Omega)}\right). \tag{3.3}$$

Since $\dot{\ell} = (u(1) - u(-1))/2$, the Schwarz inequality gives

$$|\ell|_{\mathcal{H}^1(\Omega)} = \frac{|u(1) - u(-1)|}{\sqrt{2}} = \frac{1}{\sqrt{2}} \left| \int_{-1}^{1} \dot{u}(\tau)\, d\tau \right| \le |u|_{\mathcal{H}^1(\Omega)}. \tag{3.4}$$

Combine (3.3) and (3.4) to obtain (3.2) for all $u \in \mathcal{H}^1(\Omega)$. Henceforth, it is assumed that $u \in \mathcal{H}_0^1(\Omega) \cap \mathcal{H}^\eta(\Omega)$.

Let $\pi_N u$ denote the projection of u into $\mathcal{P}_N^0$ relative to the norm $|\cdot|_{\mathcal{H}^1(\Omega)}$. Define $E_N = u - \pi_N u$ and $e_N = E_N^I = (u - \pi_N u)^I = u^I - \pi_N u$. Since $E_N - e_N = u - u^I$, it follows that

$$|u - u^I|_{\mathcal{H}^1(\Omega)} \le |E_N|_{\mathcal{H}^1(\Omega)} + |e_N|_{\mathcal{H}^1(\Omega)}. \tag{3.5}$$

In [21, Prop. 3.1] it is shown that for $\eta \geq 1$,

$$|E_N|_{\mathcal{H}^1(\Omega)} \leq (c/N)^{p-1}|u|_{\mathcal{H}^p(\Omega)}, \quad \text{where} \quad p = \min\{\eta, N+1\}. \tag{3.6}$$

We will establish the bound

$$|e_N|_{\mathcal{H}^1(\Omega)} \leq c|E_N|_{\mathcal{H}^1(\Omega)}. \tag{3.7}$$

Combine (3.5)–(3.7) to obtain (3.1) and (3.2) for an appropriate choice of c.

By [5, Lem. 4.4] and the fact that $e_N \in \mathcal{P}_N^0$, it follows that

$$|e_N|_{\mathcal{H}^1(\Omega)} \leq cN \left(\int_\Omega \frac{e_N^2(\tau)}{1-\tau^2}\, d\tau \right)^{1/2}. \tag{3.8}$$

Since $e_N \in \mathcal{P}_N^0$ and $e_N^2(\tau)/(1-\tau^2) \in \mathcal{P}_{2N-2}^0$, N-point Radau quadrature is exact, and we have

$$\left(\int_\Omega \frac{e_N^2(\tau)}{1-\tau^2}\, d\tau \right)^{1/2} = \left(\sum_{i=1}^{N-1} \frac{\omega_i e_N^2(\tau_i)}{1-\tau_i^2} \right)^{1/2} = \left(\sum_{i=1}^{N-1} \frac{\omega_i E_N^2(\tau_i)}{1-\tau_i^2} \right)^{1/2}. \tag{3.9}$$

The last equality holds since $e_N = E_N$ at τ_i, $0 \leq i \leq N$. Although Lemma 4.3 in [5] was given for Lobatto quadrature, exactly the same proof can be used for both Gauss and Radau quadrature. Consequently, since $E_N \in \mathcal{H}_0^1(\Omega)$, it follows from [5, Lem. 4.3] that

$$\left(\sum_{i=1}^{N} \frac{\omega_i E_N^2(\tau_i)}{1-\tau_i^2} \right)^{1/2} \leq c \left[\left(\int_\Omega \frac{E_N^2(\tau)}{1-\tau^2}\, d\tau \right)^{1/2} + N^{-1}|E_N|_{\mathcal{H}^1(\Omega)} \right]. \tag{3.10}$$

By [36, Prop. 9.1], we have

$$N \left[\left(\int_\Omega \frac{E_N^2(\tau)}{1-\tau^2}\, d\tau \right)^{1/2} + N^{-1}|E_N|_{\mathcal{H}^1(\Omega)} \right] \leq 2|E_N|_{\mathcal{H}^1(\Omega)}. \tag{3.11}$$

Combine (3.8–3.11) to obtain (3.7). □

Remark 3.1 In the analogue of Lemma 3.1 for the Gauss quadrature points given in [36, Lem. 4.1], the exponent in the error bound is $p - 1.5$ instead of $p - 1$. The difference in the exponent is due to the treatment of endpoints. In the Radau result, the polynomial interpolates at both $\tau = -1$ and $\tau = 1$, while in the Gauss result, the polynomial interpolates only at $\tau = -1$.

4 Analysis of the residual

The distance from $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ to $\mathcal{F}(\mathbf{U}^*)$ is now estimated.

Lemma 4.1 *If $\mathbf{x}^*$ and $\boldsymbol{\lambda}^* \in \mathcal{PH}^\eta(\Omega_0)$ for some $\eta \geq 2$, then there exists a constant c, independent of N, h, and η, such that*

$$\begin{aligned} &\operatorname{dist}[\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*), \mathcal{F}(\mathbf{U}^*)]_{\mathcal{Y}} \\ &\quad \leq h^{p-1/2} \left(\frac{c}{N}\right)^{p-1} |\mathbf{x}^*|_{\mathcal{PH}^p(\Omega_0)} + h^{q-1/2} \left(\frac{c}{N}\right)^{q-1.5} |\boldsymbol{\lambda}^*|_{\mathcal{PH}^q(\Omega_0)}, \end{aligned} \tag{4.1}$$

where $p = \min(\eta, N+1)$ and $q = \min(\eta, N)$.

Proof Since $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ appears throughout the analysis, it is abbreviated $\mathcal{T}^*$. Since the minimum principle (1.6) holds for all $t \in \Omega_0$, it holds at the collocation points, which implies that $\mathcal{T}_6^* \in \mathcal{F}_6(\mathbf{U}^*)$. Also, $\mathcal{T}_2^* = \mathcal{T}_5^* = \mathbf{0}$ since the optimal state is continuous and it satisfies the terminal condition (1.5) in the costate equation. Thus we only need to analyze $\mathcal{T}_1^*$, $\mathcal{T}_3^*$, and $\mathcal{T}_4^*$.

Let us first consider $\mathcal{T}_1^*$. Since $\mathbf{D}$ is a differentiation matrix associated with the collocation points, we have

$$\sum_{j=0}^{N} D_{ij} \mathbf{X}_{kj}^* = \dot{\mathbf{x}}_k^I(\tau_i), \quad 1 \leq i \leq N, \tag{4.2}$$

where $\mathbf{x}_k^I \in \mathcal{P}_N^n$ is the (interpolating) polynomial that passes through $\mathbf{x}_k^*(\tau_j)$ for $0 \leq j \leq N$. Since $\mathbf{x}^*$ satisfies the dynamics of (1.1),

$$h\mathbf{f}(\mathbf{X}_{ki}^*, \mathbf{U}_{ki}^*) = \dot{\mathbf{x}}_k^*(\tau_i). \tag{4.3}$$

Combine (4.2) and (4.3) to obtain

$$\mathcal{T}_{1ki}^* = \dot{\mathbf{x}}_k^I(\tau_i) - \dot{\mathbf{x}}_k^*(\tau_i) = \dot{\mathbf{x}}_k^I(\tau_i) - (\dot{\mathbf{x}}_k^*)^J(\tau_i), \tag{4.4}$$

where $(\dot{\mathbf{x}}_k^*)^J \in \mathcal{P}_{N-1}^n$ is the interpolant that passes through $\dot{\mathbf{x}}_k^*(\tau_i)$ for $1 \leq i \leq N$. Since both $\dot{\mathbf{x}}^I$ and $(\dot{\mathbf{x}}^*)^J$ are polynomials of degree $N-1$ and Radau quadrature is exact for polynomials of degree $2N-2$, it follows that

$$\begin{aligned} \|\mathcal{T}_1^*\|_\omega^2 &= \sum_{k=1}^{K} \sum_{i=1}^{N} \omega_i |\dot{\mathbf{x}}_k^I(\tau_i) - (\dot{\mathbf{x}}_k^*)^J(\tau_i)|^2 \\ &= \sum_{k=1}^{K} \int_{-1}^{1} |\dot{\mathbf{x}}_k^I(\tau) - (\dot{\mathbf{x}}_k^*)^J(\tau)|^2 \, d\tau \\ &\leq 2 \sum_{k=1}^{K} \int_{-1}^{1} \left(|\dot{\mathbf{x}}_k^I(\tau) - \dot{\mathbf{x}}_k^*(\tau)|^2 + |\dot{\mathbf{x}}_k^*(\tau) - (\dot{\mathbf{x}}_k^*)^J(\tau)|^2 \right) d\tau. \end{aligned} \tag{4.5}$$

By Lemma 3.1, we have

$$\|\dot{\mathbf{x}}_k^I - \dot{\mathbf{x}}_k^*\|_{\mathcal{L}^2(\Omega)} \le (c/N)^{p-1} |\mathbf{x}_k^*|_{\mathcal{H}^p(\Omega)}, \quad p = \min\{\eta, N+1\}. \tag{4.6}$$

The second term in (4.5) involves the difference between between $\dot{\mathbf{x}}_k^* \in \mathcal{H}^{(\eta-1)}$ and its interpolant $(\dot{\mathbf{x}}_k^*)^J \in \mathcal{P}_{N-1}^n$ at the N Radau points. By the bound given in [9, (5.4.33)] for the $\mathcal{L}^2$ error in Radau interpolation, this term has exactly the same bound as that on the right side of (4.6). Since $\mathbf{x}_k(\tau) = \mathbf{x}(t_{k-1/2} + h\tau)$, the derivatives contained in the right side of (4.6) satisfy

$$\frac{d^p \mathbf{x}_k^*(\tau)}{d\tau^p} = h^p \left. \frac{d^p \mathbf{x}^*(t)}{dt^p} \right|_{t=t_{k-1/2}+h\tau}.$$

Consequently, after a change of variables, we have

$$\int_{-1}^{1} \left| \frac{d^p \mathbf{x}_k^*(\tau)}{d\tau^p} \right|^2 d\tau = h^{2p-1} \int_{t_{k-1}}^{t_k} \left| \frac{d^p \mathbf{x}^*(t)}{dt^p} \right|^2 dt.$$

Combine this with (4.5) and (4.6) to deduce that $\|\mathcal{T}_1^*\|_\omega$ is bounded by the first term on the right side side of (4.1).

The analysis of $\mathcal{T}_3^*$ is similar to the analysis of $\mathcal{T}_1^*$. Let $\boldsymbol{\lambda}_k^I \in \mathcal{P}_{N-1}^n$ be the polynomial that interpolates $\boldsymbol{\lambda}_k^*(\tau_j)$ for $1 \le j \le N$. By (2.14) and (2.15), we have

$$\sum_{j=1}^{N} D_{ij}^{\ddagger} \boldsymbol{\Lambda}_{kj}^* = \dot{\boldsymbol{\lambda}}_k^I(\tau_i), \quad 1 \le i < N, \tag{4.7}$$

$$\sum_{j=1}^{N} D_{Nj}^{\ddagger} \boldsymbol{\Lambda}_{kj}^* = \dot{\boldsymbol{\lambda}}_k^I(\tau_i) - \boldsymbol{\lambda}_k^*(1)/\omega_N. \tag{4.8}$$

Since $\boldsymbol{\lambda}^*$ satisfies (1.5), it follows that

$$h\nabla_x H(\mathbf{X}_{ki}^*, \mathbf{U}_{ki}^*, \boldsymbol{\Lambda}_{ki}^*) = h\nabla_x H(\mathbf{x}_k^*(\tau_i), \mathbf{u}_k^*(\tau_i), \boldsymbol{\lambda}_k^*(\tau_i)) = -\dot{\boldsymbol{\lambda}}_k^*(\tau_i), \tag{4.9}$$

$1 \le i \le N$. We substitute (4.7)–(4.9) in the definition of $\mathcal{T}_3$ to obtain

$$\mathcal{T}_{3ki}(\mathbf{X}^*, \mathbf{U}^*, \boldsymbol{\Lambda}^*) = \dot{\boldsymbol{\lambda}}_k^I(\tau_i) - \dot{\boldsymbol{\lambda}}_k^*(\tau_i) = \dot{\boldsymbol{\lambda}}_k^I(\tau_i) - (\dot{\boldsymbol{\lambda}}_k^*)^J(\tau_i), \quad 1 \le i \le N,$$

where $(\dot{\boldsymbol{\lambda}}_k^*)^J \in \mathcal{P}_{N-1}^n$ is the polynomial that passes through $\dot{\boldsymbol{\lambda}}_k^*(\tau_i)$, $1 \le i \le N$. Note that the term $-\boldsymbol{\lambda}_k^*(1)/\omega_N$ in (4.8) cancels the corresponding term in $\mathcal{T}_{3k}$ due to the continuity of $\boldsymbol{\lambda}^*$. Since $\dot{\boldsymbol{\lambda}}_k^I \in \mathcal{P}_{N-2}^n$ and $(\dot{\boldsymbol{\lambda}}_k^*)^J \in \mathcal{P}_{N-1}^n$, and since Radau quadrature is exact for polynomials of degree $2N-2$, we obtain, as in (4.5),

$$\|\mathcal{T}_3^*\|_\omega^2 \le 2 \sum_{k=1}^{K} \int_{-1}^{1} \left(|\dot{\boldsymbol{\lambda}}_k^I(\tau) - \dot{\boldsymbol{\lambda}}_k^*(\tau)|^2 + |\dot{\boldsymbol{\lambda}}_k^*(\tau) - (\dot{\boldsymbol{\lambda}}_k^*)^J(\tau)|^2 \right) d\tau. \tag{4.10}$$

The last term in (4.10) has the bound

$$\|(\dot{\boldsymbol{\lambda}}_k^*)^J - \dot{\boldsymbol{\lambda}}_k^*\|_{\mathcal{L}^2(\Omega)} \le h^p (c/N)^{p-1} |\boldsymbol{\lambda}^*|_{\mathcal{H}^p(t_{k-1},t_k)}, \quad p = \min\{\eta, N+1\}, \tag{4.11}$$

corresponding to the $\mathcal{L}^2$ error in interpolation at the Radau points. The other term, however, is different from the state since $\boldsymbol{\lambda}_k^I$ has degree $N-1$ while the state $\mathbf{x}_k^I$ has degree N, and the state interpolates at both the quadrature points and at $\tau = -1$, while $\boldsymbol{\lambda}_k^I$ only interpolates at the quadrature points. The error in the derivative of the interpolant at the Radau points has the bound [9, (5.4.34)]

$$\|\dot{\boldsymbol{\lambda}}_k^I - \dot{\boldsymbol{\lambda}}_k^*\|_{\mathcal{L}^2(\Omega)} \le h^q (c/N)^{q-1.5} |\boldsymbol{\lambda}^*|_{\mathcal{H}^q(\Omega)}, \quad q = \min\{\eta, N\}. \tag{4.12}$$

The exponent changes from $p-1$ in (4.11) to $q-1.5$ due to the fact that $\boldsymbol{\lambda}_k^I$ does not interpolate at $\tau = -1$, and $q \le p$ since the polynomial associated with $\boldsymbol{\lambda}_k^I$ has degree $N-1$. Note that if $\boldsymbol{\lambda}^* \in \mathcal{PH}^\eta(\Omega_0)$, then $\boldsymbol{\lambda}^* \in \mathcal{PH}^{(\eta-1)}(\Omega_0)$, so we can always ensure that the error bound (4.12) dominates the error bound (4.11) by lowering η in (4.11) if necessary. Utilizing the bound (4.12) in (4.10) and changing variables from τ to t, we deduce that $\|\mathcal{T}_3\|_\omega$ is bounded by the second term on the right side of (4.1).

Finally, let us consider $\mathcal{T}_4^*$. Applying (4.9) and utilizing the continuity of $\boldsymbol{\lambda}^*$ and the exactness of Radau quadrature, we have

$$\begin{aligned}
\mathcal{T}_{4k}^* &= \boldsymbol{\lambda}_k^*(-1) - \boldsymbol{\lambda}_{k+1}^*(-1) + \sum_{i=1}^N \omega_i \dot{\boldsymbol{\lambda}}_k^*(\tau_i) \\
&= \boldsymbol{\lambda}_k^*(-1) - \boldsymbol{\lambda}_k^*(1) + \sum_{i=1}^N \omega_i (\dot{\boldsymbol{\lambda}}_k^*)^J(\tau_i) \\
&= \boldsymbol{\lambda}_k^*(-1) - \boldsymbol{\lambda}_k^*(1) + \int_{-1}^1 (\dot{\boldsymbol{\lambda}}_k^*)^J(\tau)\, d\tau = \int_{-1}^1 [(\dot{\boldsymbol{\lambda}}_k^*)^J(\tau) - \dot{\boldsymbol{\lambda}}_k^*(\tau)]\, d\tau.
\end{aligned}$$

By (4.11) and the Schwarz inequality, we have

$$|\mathcal{T}_{4k}^*| \le \sqrt{2}\|(\dot{\boldsymbol{\lambda}}_k^*)^J - \dot{\boldsymbol{\lambda}}_k^*\|_{\mathcal{L}^2(\Omega)} \le h^p (c/N)^{p-1} |\boldsymbol{\lambda}^*|_{\mathcal{H}^p(\Omega)}, \quad p = \min\{\eta, N+1\}.$$

As in the analysis of $\mathcal{T}_3$, we square this, sum over k, change variables from τ to t, and take the square root to obtain a bound that can be dominated by the last term in (4.1). This completes the proof. □

5 Invertibility of linearized dynamics

The inclusion

$$\mathcal{T}(\mathbf{X}, \mathbf{U}, \boldsymbol{\Lambda}) \in \mathcal{F}(\mathbf{U}),$$

corresponding to the first-order optimality conditions for the discrete problem (1.4), will be linearized around $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$. Given $\mathbf{Y} \in \mathcal{Y}$, the linearized problem is to find $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ such that

$$(\nabla \mathcal{T}^*)[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] + \mathbf{Y} \in \mathcal{F}(\mathbf{U}), \tag{5.1}$$

where $\nabla \mathcal{T}^*$ denotes $\nabla \mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$, the derivative of $\mathcal{T}$ evaluated at $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$. Since $\mathbf{\Lambda}$ enters $\mathcal{T}$ in an affine manner, the linearization with respect to $\mathbf{\Lambda}$ is trivial. On the other hand, the discrete state $\mathbf{X}$ and the discrete control $\mathbf{U}$ generally enter $\mathcal{T}$ in a nonlinear fashion. The derivative of $\mathcal{T}$ in (2.19)–(2.25) is built from the following matrices for $1 \le k \le K$:

$$\begin{aligned}
\mathbf{A}_{ki} &= \nabla_x \mathbf{f}(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki})), & \mathbf{B}_{ki} &= \nabla_u \mathbf{f}(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki})), \\
\mathbf{Q}_{ki} &= \nabla^2_{xx} H\left(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki}), \boldsymbol{\lambda}^*(t_{ki})\right), & \mathbf{S}_{ki} &= \nabla^2_{xu} H\left(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki}), \boldsymbol{\lambda}^*(t_{ki})\right), \\
\mathbf{R}_{ki} &= \nabla^2_{uu} H\left(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki}), \boldsymbol{\lambda}^*(t_{ki})\right), & \mathbf{T} &= \nabla^2 C(\mathbf{x}^*(1)).
\end{aligned}$$

As pointed out in (1.11), the optimal variables $(\mathbf{x}^*, \mathbf{u}^*, \boldsymbol{\lambda}^*)$ evaluated at the t_{ki} are equivalent to the transformed optimal variables $(\mathbf{x}^*_k, \mathbf{u}^*_k, \boldsymbol{\lambda}^*_k)$ evaluated at the τ_i. The elements of $\nabla \mathcal{T}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}]$ are the following:

$$\begin{aligned}
\nabla \mathcal{T}^*_{1ki}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \left(\sum_{j=0}^{N} D_{ij} \mathbf{X}_{kj}\right) - h(\mathbf{A}_{ki}\mathbf{X}_{ki} + \mathbf{B}_{ki}\mathbf{U}_{ki}), \quad 1 \le i \le N, \\
\nabla \mathcal{T}^*_{2k}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{X}_{k0} - \mathbf{X}_{k-1,N}, \quad \text{where} \quad \mathbf{X}_{0N} = \mathbf{0}, \\
\nabla \mathcal{T}^*_{3ki}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \left(\sum_{j=1}^{N} D^{\ddagger}_{ij} \mathbf{\Lambda}_{kj}\right) + h(\mathbf{A}^{\mathsf{T}}_{ki}\mathbf{\Lambda}_{ki} + \mathbf{Q}_{ki}\mathbf{X}_{ki} + \mathbf{S}_{ki}\mathbf{U}_{ki}), \quad 1 \le i < N, \\
\nabla \mathcal{T}^*_{3kN}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \left(\sum_{j=1}^{N} D^{\ddagger}_{Nj} \mathbf{\Lambda}_{kj}\right) + h(\mathbf{A}^{\mathsf{T}}_{kN}\mathbf{\Lambda}_{kN} + \mathbf{Q}_{kN}\mathbf{X}_{kN} + \mathbf{S}_{kN}\mathbf{U}_{kN}) \\
&\quad + \mathbf{\Lambda}_{k+1,0}/\omega_N, \\
\nabla \mathcal{T}^*_{4k}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{\Lambda}_{k0} - \mathbf{\Lambda}_{k+1,0} - h \sum_{i=1}^{N} \omega_i (\mathbf{A}^{\mathsf{T}}_{ki}\mathbf{\Lambda}_{ki} + \mathbf{Q}_{ki}\mathbf{X}_{ki} + \mathbf{S}_{ki}\mathbf{U}_{ki}), \\
\nabla \mathcal{T}^*_{5k}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{T}\mathbf{X}_{KN} - \mathbf{\Lambda}_{K+1,0}, \\
\nabla \mathcal{T}^*_{6ki}[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= -h(\mathbf{B}^{\mathsf{T}}_{ki}\mathbf{\Lambda}_{ki} + \mathbf{S}^{\mathsf{T}}_{ki}\mathbf{X}_{ki} + \mathbf{R}_{ki}\mathbf{U}_{ki}), \quad 1 \le i \le N,
\end{aligned} \tag{5.2}$$

where $1 \le k \le K$.

In this section the invertibility of the linearized dynamics for either the state or the costate is analyzed. The following properties of the submatrix $\mathbf{D}_{1:N}$, consisting of the trailing N columns of $\mathbf{D}$, are used in the analysis:

(P1) $\mathbf{D}_{1:N}$ is invertible and $\|\mathbf{D}^{-1}_{1:N}\|_\infty = 2$.

(P2) If $\mathbf{W}$ is the diagonal matrix containing the Radau quadrature weights ω_i on the diagonal, then the rows of the matrix $[\mathbf{W}^{1/2}\mathbf{D}_{1:N}]^{-1}$ have Euclidean norm bounded by $\sqrt{2}$.

These properties are established in Lemma 10.1 of the "Appendix". For further analysis of (P1) and (P2), see the companion paper [10].

Lemma 5.1 *If* (A2) *holds, then for each* $\mathbf{q} \in \mathbb{R}^{Kn}$ *and* $\mathbf{p} \in \mathbb{R}^{KNn}$ *with* $\mathbf{q}_k$ *and* $\mathbf{p}_{ki} \in \mathbb{R}^n$, *the linear system*

$$\sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj} = h\mathbf{A}_{ki}\mathbf{X}_{ki} + \mathbf{p}_{ki}, \quad 1 \le i \le N, \tag{5.3}$$

$$\mathbf{X}_{k0} = \mathbf{X}_{k-1,N} + \mathbf{q}_k, \quad \mathbf{X}_{0N} = \mathbf{0}, \tag{5.4}$$

$1 \le k \le K$, *has a unique solution* $\mathbf{X} \in \mathbb{R}^{K(N+1)n}$. *This solution has the bound*

$$\sup_{\substack{1\le k\le K\\ 1\le j\le N}} \|\mathbf{X}_{kj}\|_\infty \le h^{-1/2}\left(\frac{\sqrt{2}\|\mathbf{p}\|_\omega + |\mathbf{q}|}{(1-2hd_1)^K}\right). \tag{5.5}$$

Remark 5.1 By (5.4), $\|\mathbf{X}_{k0}\|_\infty \le \|\mathbf{X}_{k-1,N}\|_\infty + \|\mathbf{q}_k\|_\infty \le \|\mathbf{X}_{k-1,N}\|_\infty + |\mathbf{q}_k|$. Hence, the entire solution $\mathbf{X}$ of (5.3)–(5.4) has a sup-norm bound of the form (5.5).

Remark 5.2 Recall that d_1 is defined in (1.10). Since the denominator expression $(1 - 2hd_1)^K = (1 - d_1/K)^K$ in the bound (5.5) approaches e^{-d_1} as K tends to infinity, the denominator is bounded away from zero, uniformly in K. Hence, (5.5) also implies a uniform bound, independent of K.

Proof We first show that for given $\mathbf{X}_{k0}$, the linear system (5.3) uniquely determines $\mathbf{X}_{k1}$ through $\mathbf{X}_{kN}$. Since $\mathbf{X}_{0N} = \mathbf{0}$, it follows from (5.4) that $\mathbf{X}_{10} = \mathbf{q}_1$ is known. Consequently, for $k = 1$ up to $k = K$, we can use (5.3) to compute $\mathbf{X}_{k1}$ through $\mathbf{X}_{kN}$, and then (5.4) to evaluate $\mathbf{X}_{k+1,0}$. This shows that (5.3)–(5.4) has a unique solution that can be computed by a recursive process.

Let $\overline{\mathbf{X}}_k$ be the vector obtained by vertically stacking $\mathbf{X}_{k1}$ through $\mathbf{X}_{kN}$, let $\mathbf{A}_k$ be the block diagonal matrix with i-th diagonal block $\mathbf{A}_{ki}$, $1 \le i \le N$, define $\overline{\mathbf{D}} = \mathbf{D}_{1:N} \otimes \mathbf{I}_n$ where $\otimes$ is the Kronecker product, and let $\mathbf{D}_0$ denote the first column of $\mathbf{D}$. With this notation, (5.3)–(5.4) reduce to

$$(\overline{\mathbf{D}} - h\mathbf{A}_k)\overline{\mathbf{X}}_k = \mathbf{p} - (\mathbf{D}_0 \otimes \mathbf{I}_n)\mathbf{X}_{k0} = \mathbf{p} - (\mathbf{D}_0 \otimes \mathbf{I}_n)(\mathbf{X}_{k-1,N} + \mathbf{q}_k). \tag{5.6}$$

By (P1), $\mathbf{D}_{1:N}$ is invertible and $\|\mathbf{D}_{1:N}^{-1}\|_\infty = 2$. Hence, $\|\overline{\mathbf{D}}^{-1}\| = \|\mathbf{D}_{1:N}^{-1} \otimes \mathbf{I}_n\| = 2$, and by (A2), we have $2h\|\mathbf{A}_k\|_\infty \le 2hd_1 < 1$, which implies that

$$h\|\overline{\mathbf{D}}^{-1}\mathbf{A}_k\|_\infty \le h\|\overline{\mathbf{D}}^{-1}\|_\infty\|\mathbf{A}_k\|_\infty \le 2hd_1 < 1.$$

By [38, p. 351], $\mathbf{I} - h\overline{\mathbf{D}}^{-1}\mathbf{A}_k$ is invertible and

$$\|(\mathbf{I} - h\overline{\mathbf{D}}^{-1}\mathbf{A})^{-1}\|_\infty \le 1/(1 - 2hd_1). \tag{5.7}$$

Multiply (5.6) first by $\overline{\mathbf{D}}^{-1}$ and then by $(\mathbf{I} - h\overline{\mathbf{D}}^{-1}\mathbf{A}_k)^{-1}$ to obtain

$$\overline{\mathbf{X}}_k = (\mathbf{I} - h\overline{\mathbf{D}}^{-1}\mathbf{A}_k)^{-1}\left(\overline{\mathbf{D}}^{-1}\mathbf{p}_k + \overline{\mathbf{D}}^{-1}(\mathbf{D}_0 \otimes \mathbf{I}_n)(\mathbf{X}_{k-1,N} + \mathbf{q}_k)\right).$$

It is shown in [36, Lem. 5.1] that $(\overline{\mathbf{D}})^{-1}[\mathbf{D}_0 \otimes \mathbf{I}_n] = -\mathbf{1} \otimes \mathbf{I}_n$. Consequently,

$$\overline{\mathbf{X}}_k = (\mathbf{I} - h\overline{\mathbf{D}}^{-1}\mathbf{A}_k)^{-1}\left(\overline{\mathbf{D}}^{-1}\mathbf{p}_k - \mathbf{1} \otimes (\mathbf{X}_{k-1,N} + \mathbf{q}_k)\right).$$

Take norms and apply (5.7) to get

$$\|\overline{\mathbf{X}}_k\|_\infty \le \left(\frac{1}{1 - 2hd_1}\right)\left(\|\overline{\mathbf{D}}^{-1}\mathbf{p}_k\|_\infty + \|\mathbf{X}_{k-1,N}\|_\infty + \|\mathbf{q}_k\|_\infty\right). \tag{5.8}$$

In [36, Lem. 5.1] it is shown that by (P2), we have

$$\|\overline{\mathbf{D}}^{-1}\mathbf{p}_k\|_\infty \le \sqrt{2}\|\mathbf{p}_k\|_\omega, \quad \|\mathbf{p}_k\|_\omega = \left(\sum_{i=1}^{N} \omega_i |\mathbf{p}_{ki}|^2\right).$$

Insert this bound in (5.8) and utilize the trivial inequality $\|\mathbf{q}_k\|_\infty \le |\mathbf{q}_k|$ to obtain

$$\|\overline{\mathbf{X}}_k\|_\infty \le \left(\frac{1}{1 - 2hd_1}\right)\left(\|\mathbf{X}_{k-1,N}\|_\infty + \sqrt{2}\|\mathbf{p}_k\|_\omega + |\mathbf{q}_k|\right). \tag{5.9}$$

Since $\mathbf{X}_{kN} = \mathbf{0}$ for $k = 0$ and $\|\mathbf{X}_{k,N}\|_\infty \le \|\overline{\mathbf{X}}_k\|_\infty$ for $k > 0$, (5.9) yields

$$\|\overline{\mathbf{X}}_k\|_\infty \le \sum_{j=1}^{k} \frac{\sqrt{2}\|\mathbf{p}_j\|_\omega + |\mathbf{q}_j|}{(1 - 2hd_1)^{k-j+1}} \tag{5.10}$$

for $1 \le k \le K$. The upper bound (5.5) is obtained by replacing $1/(1 - 2hd_1)^{k-j+1}$ by its maximum $1/(1 - 2hd_1)^K$ and by utilizing the Schwarz inequality as in

$$\sum_{j=1}^{k} \|\mathbf{p}_j\|_\omega \le \sqrt{k}\|\mathbf{p}\|_\omega \le h^{-1/2}\|\mathbf{p}\|_\omega \quad \text{and} \quad \sum_{j=1}^{k} |\mathbf{q}_j| \le \sqrt{k}|\mathbf{q}| \le h^{-1/2}|\mathbf{q}|. \tag{5.11}$$

□

The linearized costate dynamics has an analogous bound. The analysis utilizes the following properties of $\mathbf{D}^\ddagger$ established in the companion paper [10]:

(P3) $\mathbf{D}^{\ddagger}$ is invertible and $\|(\mathbf{D}^{\ddagger})^{-1}\|_{\infty} \le 2$.
(P4) The rows of the matrix $[\mathbf{W}^{1/2}\mathbf{D}^{\ddagger}]^{-1}$ have Euclidean norm bounded by $\sqrt{2}$.

Note that the bound $\|(\mathbf{D}^{\ddagger})^{-1}\|_{\infty} \le 2$ in (P3) is implied by (P4) due to the inequality (10.7) contained in the proof of (P1) and (P2).

Lemma 5.2 *If* (A2) *holds, then for each* $\mathbf{q} \in \mathbb{R}^{Kn}$, $\mathbf{p} \in \mathbb{R}^{KNn}$, *and* $\mathbf{\Lambda}_{K+1,0} \in \mathbb{R}^{n}$ *with* $\mathbf{q}_k$ *and* $\mathbf{p}_{ki} \in \mathbb{R}^{n}$, *the linear system*

$$\sum_{j=1}^{N} D^{\ddagger}_{ij}\mathbf{\Lambda}_{kj} = \mathbf{p}_{ki} - h\mathbf{A}_{ki}^{\mathsf{T}}\mathbf{\Lambda}_{ki}, \quad 1 \le i < N, \tag{5.12}$$

$$\sum_{j=1}^{N} D^{\ddagger}_{Nj}\mathbf{\Lambda}_{kj} = \mathbf{p}_{kN} - h\mathbf{A}_{kN}^{\mathsf{T}}\mathbf{\Lambda}_{kN} - \mathbf{\Lambda}_{k+1,0}/\omega_N, \tag{5.13}$$

$$\mathbf{\Lambda}_{k0} = \mathbf{\Lambda}_{k+1,0} + \mathbf{q}_k + h\sum_{i=1}^{N} \omega_i \mathbf{A}_{ki}^{\mathsf{T}}\mathbf{\Lambda}_{ki}, \tag{5.14}$$

$1 \le k \le K$, *has a unique solution* $\mathbf{\Lambda} \in \mathbb{R}^{K(N+1)n}$. *This solution has the bound*

$$\|\mathbf{\Lambda}\|_{\infty} \le \frac{\|\mathbf{\Lambda}_{K+1,0}\|_{\infty} + h^{-1/2}\sqrt{2}\|\mathbf{p}\|_{\omega} + \sum_{k=1}^{K}|\mathbf{q}_k|}{(1-2hd_2)^K}. \tag{5.15}$$

Proof The proof is similar to the proof of Lemma 5.1 except that the recursive solution of (5.12)–(5.14) starts from $k = K$ and descends to $k = 1$. In particular, we first show that for given $\mathbf{\Lambda}_{k+1,0}$, the linear system (5.12)–(5.13) uniquely determines $\mathbf{\Lambda}_{k1}$ through $\mathbf{\Lambda}_{kN}$; then (5.14) can be used to evaluate $\mathbf{\Lambda}_{k0}$.

Define $\overline{\mathbf{D}}^{\ddagger} = \mathbf{D}^{\ddagger} \otimes \mathbf{I}_n$, where $\otimes$ is the Kronecker product. Equations (5.12) and (5.13) can be combined into the single equation

$$\overline{\mathbf{D}}^{\ddagger}\overline{\mathbf{\Lambda}}_k = \mathbf{p}_k - h\mathbf{A}_k^{\mathsf{T}}\overline{\mathbf{\Lambda}}_k - (\mathbf{e}_N \otimes \mathbf{I}_n)\mathbf{\Lambda}_{k+1,0}/\omega_N, \tag{5.16}$$

where $\overline{\mathbf{\Lambda}}_k$ is obtained by vertically stacking $\mathbf{\Lambda}_{k1}$ through $\mathbf{\Lambda}_{kN}$ and $\mathbf{e}_N$ is the vector whose N components are all zero except for the last component which is 1. By (2.14) and (2.15), $\mathbf{D}^{\ddagger}\mathbf{1} = -\mathbf{e}_N/\omega_N$, which implies that

$$\mathbf{D}^{\ddagger\,-1}\mathbf{e}_N = -\omega_N\mathbf{1}. \tag{5.17}$$

Hence, we have

$$\overline{\mathbf{D}}^{\ddagger\,-1}(\mathbf{e}_N \otimes \mathbf{I}_n)/\omega_N = [\mathbf{D}^{\ddagger\,-1} \otimes \mathbf{I}_n](\mathbf{e}_N \otimes \mathbf{I}_n)/\omega_N = -\mathbf{1} \otimes \mathbf{I}_n.$$

Multiply (5.16) by $\overline{\mathbf{D}}^{\ddagger\,-1}$ and rearrange to obtain

$$(\mathbf{I} + h\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{A}_k^{\mathsf{T}})\overline{\mathbf{\Lambda}}_k = \overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k + (\mathbf{1} \otimes \mathbf{I}_n)\mathbf{\Lambda}_{k+1,0}. \tag{5.18}$$

By (A2) and (P3), $h\|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{A}_k^{\mathsf{T}}\|_\infty \le 2hd_2 < 1$. Consequently, the matrix $\mathbf{I}+h\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{A}_k^{\mathsf{T}}$ is invertible with

$$\left\|\left(\mathbf{I}+h\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{A}_k^{\mathsf{T}}\right)^{-1}\right\|_\infty \le \frac{1}{1-2hd_2}.$$

Multiply (5.18) by $(\mathbf{I}+h\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{A}_k^{\mathsf{T}})^{-1}$ and take the norm of each side to obtain

$$\|\overline{\mathbf{\Lambda}}_k\|_\infty \le \left(\frac{1}{1-2hd_2}\right)(\|\mathbf{\Lambda}_{k+1,0}\|_\infty + \|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k\|_\infty) \tag{5.19}$$

$$\le \left(\frac{1}{1-2hd_2}\right)(\|\mathbf{\Lambda}_{k+1,0}\|_\infty + \|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k\|_\infty + \|\mathbf{q}_k\|_\infty) \tag{5.20}$$

The norm of (5.14) gives

$$\begin{aligned}\|\mathbf{\Lambda}_{k0}\|_\infty &\le \|\mathbf{\Lambda}_{k+1,0}\|_\infty + \|\mathbf{q}_k\|_\infty + h\sum_{i=1}^{N}\omega_i\|\mathbf{A}_{ki}^{\mathsf{T}}\|_\infty\|\mathbf{\Lambda}_{ki}\|_\infty\\ &\le \|\mathbf{\Lambda}_{k+1,0}\|_\infty + \|\mathbf{q}_k\|_\infty + 2hd_2\|\overline{\mathbf{\Lambda}}_k\|_\infty\end{aligned}$$

since the ω_i sum to 2. Using the bound for $\|\overline{\mathbf{\Lambda}}_k\|_\infty$ from (5.19) and the fact that $2hd_2 < 1$, we have

$$\begin{aligned}\|\mathbf{\Lambda}_{k0}\|_\infty &\le \|\mathbf{q}_k\|_\infty + \left(\frac{1}{1-2hd_2}\right)(\|\mathbf{\Lambda}_{k+1,0}\|_\infty + 2hd_2\|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k\|_\infty)\\ &\le \left(\frac{1}{1-2hd_2}\right)(\|\mathbf{\Lambda}_{k+1,0}\|_\infty + 2hd_2\|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k\|_\infty + \|\mathbf{q}_k\|_\infty)\\ &\le \left(\frac{1}{1-2hd_2}\right)(\|\mathbf{\Lambda}_{k+1,0}\|_\infty + \|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k\|_\infty + \|\mathbf{q}_k\|_\infty).\end{aligned} \tag{5.21}$$

Since $\mathbf{\Lambda}_{k,0}$ is contained in $\mathbf{\Lambda}_k$, it follows that $\|\mathbf{\Lambda}_{k,0}\|_\infty \le \|\mathbf{\Lambda}_k\|_\infty$. Combine (5.20) and (5.21) to obtain

$$\|\mathbf{\Lambda}_k\|_\infty \le \left(\frac{1}{1-2hd_2}\right)(\|\mathbf{\Lambda}_{k+1}\|_\infty + \|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_k\|_\infty + \|\mathbf{q}_k\|_\infty),$$

where we define $\mathbf{\Lambda}_{K+1,j} = \mathbf{0}$ for $j > 0$ so that $\|\mathbf{\Lambda}_{K+1}\|_\infty = \|\mathbf{\Lambda}_{K+1,0}\|_\infty$. This inequality is applied recursively to obtain

$$\|\mathbf{\Lambda}_k\|_\infty \le \frac{\|\mathbf{\Lambda}_{K+1}\|_\infty}{(1-2hd_2)^{K+1-k}} + \sum_{j=k}^{K}\left(\frac{\|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_j\|_\infty + \|\mathbf{q}_j\|_\infty}{(1-2hd_2)^{j-k+1}}\right).$$

To bound the right side, the factors $1/(1-2hd_2)^{j-k+1}$ are replaced by their maximum $1/(1-2hd_2)^K$ to obtain

$$\|\mathbf{\Lambda}_k\|_\infty \le \frac{\|\mathbf{\Lambda}_{K+1}\|_\infty + \sum_{j=k}^K \left[\|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_j\|_\infty + |\mathbf{q}_j|\right]}{(1-2hd_2)^K}.$$

By the analysis given in [36, Lem. 5.1], (P4) implies that $\|\overline{\mathbf{D}}^{\ddagger\,-1}\mathbf{p}_j\|_\infty \le \sqrt{2}\|\mathbf{p}_j\|_\omega$. Hence, we have

$$\|\mathbf{\Lambda}_k\|_\infty \le \frac{\|\mathbf{\Lambda}_{K+1}\|_\infty + \sum_{j=k}^K \left[\sqrt{2}\|\mathbf{p}_j\|_\omega + |\mathbf{q}_j|\right]}{(1-2hd_2)^K}.$$

The first inequality in (5.11) completes the proof of (5.15). □

6 Invertibility of $\mathcal{F} - \nabla\mathcal{T}^*$

Now let us consider the invertibility of $\mathcal{F} - \nabla\mathcal{T}^*$.

Proposition 6.1 *If* (A1)–(A2) *hold, then for each* $\mathbf{Y} \in \mathcal{Y}$*, there is a unique solution* $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ *to* (5.1).

Proof Similar to the strategy used in [14–16,18,30,33,34,36], a strongly convex quadratic programming problem is formulated; the quadratic program is constructed so that the first-order optimality conditions reduce to (5.1). In particular, we consider the problem

$$\left.\begin{array}{ll} \text{minimize} & \frac{1}{2}\mathcal{Q}(\mathbf{X},\mathbf{U}) + \mathcal{L}(\mathbf{X},\mathbf{U},\mathbf{Y}) \\ \text{subject to} & \sum_{j=0}^N D_{ij}\mathbf{X}_{kj} = h(\mathbf{A}_{ki}\mathbf{X}_{ki} + \mathbf{B}_{ki}\mathbf{U}_{ki}) - \mathbf{y}_{1ki}, \quad \mathbf{U}_{ki} \in \mathcal{U}, \\ & \mathbf{X}_{k0} = \mathbf{X}_{k-1,N} - \mathbf{y}_{2k}, \quad \mathbf{X}_{0N} = \mathbf{0}, \end{array}\right\} \tag{6.1}$$

where $1 \le i \le N$ and $1 \le k \le K$. The quadratic and linear terms in the objective are

$$\mathcal{Q}(\mathbf{X},\mathbf{U}) = \mathbf{X}_{KN}^{\mathsf{T}}\mathbf{T}\mathbf{X}_{KN} + h\sum_{k=1}^K\sum_{i=1}^N \omega_i \begin{bmatrix}\mathbf{X}_{ki}\\ \mathbf{U}_{ki}\end{bmatrix}^{\mathsf{T}}\begin{bmatrix}\mathbf{Q}_{ki} & \mathbf{S}_{ki}\\ \mathbf{S}_{ki}^{\mathsf{T}} & \mathbf{R}_{ki}\end{bmatrix}\begin{bmatrix}\mathbf{X}_{ki}\\ \mathbf{U}_{ki}\end{bmatrix}, \tag{6.2}$$

$$\begin{aligned}\mathcal{L}(\mathbf{X},\mathbf{U},\mathbf{Y}) = \mathbf{y}_5^{\mathsf{T}}\mathbf{X}_{KN} &+ \sum_{k=1}^K\sum_{i=1}^N \omega_i\left(\mathbf{y}_{3ki}^{\mathsf{T}}\mathbf{X}_{ki} - \mathbf{y}_{6ki}^{\mathsf{T}}\mathbf{U}_{ki}\right)\\ &- \sum_{k=1}^K \mathbf{X}_{k0}^{\mathsf{T}}\left(\mathbf{y}_{4k} + \sum_{i=1}^N \omega_i\mathbf{y}_{3ki}\right).\end{aligned} \tag{6.3}$$

In (6.1), the minimization is over $\mathbf{X}$ and $\mathbf{U}$, while $\mathbf{Y}$ is a fixed parameter. By Lemma 5.1, the quadratic programming problem (6.1) is feasible, and the state can be expressed in

terms of the controls. After substituting for $\mathbf{X}$ in terms of $\mathbf{U}$, the quadratic programming problem (6.1) becomes a constrained minimization over $\mathbf{U}$. Since the Radau quadrature weights ω_i are strictly positive, it follows from (A1) that $\mathcal{Q}$ is strongly convex with respect to the control. Hence, there exists a unique optimal solution to (6.1) for any choice of $\mathbf{Y}$. We now show that the first-order optimality conditions for (6.1) reduce to $\nabla\mathcal{T}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] + \mathbf{Y} \in \mathcal{F}(\mathbf{U})$. The first-order optimality conditions hold since $\mathcal{U}$ has nonempty interior. Since the first-order optimality conditions are both necessary and sufficient for optimality in this convex setting, there exists a solution to (5.1). Uniqueness of $\mathbf{U}$ is due to the strong convexity assumption (A1), while the uniqueness of $\mathbf{X}$ and $\mathbf{\Lambda}$ is due to Lemmas 5.1 and 5.2 respectively.

The derivation of the first-order optimality conditions for (6.1) is essentially the same process that we used in Section 2 to write the first-order optimality conditions for the discrete problem (1.4) as $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$. The first two components of $\nabla\mathcal{T}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] + \mathbf{Y} \in \mathcal{F}(\mathbf{U})$ are simply the constraints of (6.1). The Lagrangian L for (6.1) is

$$L(\boldsymbol{\lambda}, \mathbf{X}, \mathbf{U}) = \tfrac{1}{2}\mathcal{Q}(\mathbf{X}, \mathbf{U}) + \mathcal{L}(\mathbf{X}, \mathbf{U}, \mathbf{Y}) + \sum_{k=1}^{K} \left\langle \boldsymbol{\lambda}_{k0}, \left(\mathbf{X}_{k-1,N} - \mathbf{y}_{2k} - \mathbf{X}_{k0}\right)\right\rangle + \sum_{k=1}^{K}\sum_{i=1}^{N} \left\langle \boldsymbol{\lambda}_{ki}, h(\mathbf{A}_{ki}\mathbf{X}_{ki} + \mathbf{B}_{ki}\mathbf{U}_{ki}) - \mathbf{y}_{1ki} - \sum_{j=0}^{N} D_{ij}\mathbf{X}_{kj} \right\rangle$$

The negative derivative of the Lagrangian with respect to $\mathbf{U}_{ki}$ is

$$-h\left[\mathbf{B}_{ki}^{\mathsf{T}}\boldsymbol{\lambda}_{ki} + \omega_i(\mathbf{S}_{ki}^{\mathsf{T}}\mathbf{X}_{ki} + \mathbf{R}_{ki}\mathbf{U}_{ki})\right] + \omega_i\mathbf{y}_{6ki}.$$

After substituting $\boldsymbol{\lambda}_{ki} = \omega_i\mathbf{\Lambda}_{ki}$, the requirement that this vector lies in $N_{\mathcal{U}}(\mathbf{U}_{ki})$ leads the 6th component of (5.1). Equating to zero the derivative of the Lagrangian with respect to $\mathbf{X}_{kj}$, $1 \le j < N$, yields the relation

$$\sum_{i=1}^{N} D_{ij}\boldsymbol{\lambda}_{ki} = h\left[\mathbf{A}_{kj}^{\mathsf{T}}\boldsymbol{\lambda}_{kj} + \omega_j(\mathbf{Q}_{kj}\mathbf{X}_{kj} + \mathbf{S}_{kj}\mathbf{U}_{kj})\right] + \omega_j\mathbf{y}_{3kj}.$$

Equating to zero the derivative of the Lagrangian with respect to $\mathbf{X}_{kN}$ yields the relation

$$\sum_{i=1}^{N} D_{iN}\boldsymbol{\lambda}_{ki} = h\left[\mathbf{A}_{kN}^{\mathsf{T}}\boldsymbol{\lambda}_{kN} + \omega_N(\mathbf{Q}_{kN}\mathbf{X}_{kN} + \mathbf{S}_{kN}\mathbf{U}_{kN})\right] + \omega_N\mathbf{y}_{3kN} + \boldsymbol{\lambda}_{k+1,0},$$

where $\boldsymbol{\lambda}_{K+1,0} = \mathbf{T}\mathbf{X}_{KN} + \mathbf{y}_5$. After substituting $D_{ij} = -D_{ji}^{\ddagger}\omega_j/\omega_i$, $\boldsymbol{\lambda}_{ki} = \omega_i\mathbf{\Lambda}_{ki}$, and $\boldsymbol{\lambda}_{k0} = \mathbf{\Lambda}_{k0}$, we obtain the 3rd and 5th components of (5.1).

Finally, we equate to zero the derivative of the Lagrangian with respect to $\mathbf{X}_{k0}$:

$$\sum_{i=1}^{N} D_{i0}\boldsymbol{\lambda}_{ki} = -\left(\boldsymbol{\lambda}_{k0} + \mathbf{y}_{4k} + \sum_{i=1}^{N} \omega_i \mathbf{y}_{3ki}\right).$$

Utilizing the identity (2.11), it follows that

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i D_{ij}^{\ddagger} \boldsymbol{\Lambda}_{kj} = -\left(\boldsymbol{\Lambda}_{k0} + \mathbf{y}_{4k} + \sum_{i=1}^{N} \omega_i \mathbf{y}_{3ki}\right). \tag{6.4}$$

Multiply the equations in the 3rd component of (5.1) by ω_i and sum over i to obtain

$$\sum_{i=1}^{N}\sum_{j=1}^{N} \omega_i D_{ij}^{\ddagger} \boldsymbol{\Lambda}_{kj} = -\boldsymbol{\Lambda}_{k+1,0} - \sum_{i=1}^{N} \omega_i \left[\mathbf{y}_{3ki} + h\left(\mathbf{A}_{ki}^{\mathsf{T}}\boldsymbol{\Lambda}_{ki} + \mathbf{Q}_{ki}\mathbf{X}_{ki} + \mathbf{S}_{ki}\mathbf{U}_{ki}\right)\right].$$

By (6.4), it follows that

$$\boldsymbol{\Lambda}_{k0} - \boldsymbol{\Lambda}_{k+1,0} - h\sum_{i=1}^{N} \omega_i (\mathbf{A}_{ki}^{\mathsf{T}}\boldsymbol{\Lambda}_{ki} + \mathbf{Q}_{ki}\mathbf{X}_{ki} + \mathbf{S}_{ki}\mathbf{U}_{ki}) + \mathbf{y}_{4k} = \mathbf{0},$$

which is the 4th component of (5.1). This completes the proof. □

7 Lipschitz continuity of $(\mathcal{F} - \nabla\mathcal{T}^*)^{-1}$ and proof of the main theorem

We begin by making the change of variables $\mathbf{X} = \mathbf{Z}(\mathbf{U}) + \boldsymbol{\chi}(\mathbf{Y})$ where $\boldsymbol{\chi}(\mathbf{Y})$ denotes the solution of the state dynamics (5.3) corresponding to $\mathbf{p}_{ki} = -\mathbf{y}_{1ki}$ and $\mathbf{q}_k = -\mathbf{y}_{2k}$, and $\mathbf{Z}(\mathbf{U})$ denotes solution corresponding to $\mathbf{p}_{ki} = h\mathbf{B}_{ki}\mathbf{U}_{ki}$ and $\mathbf{q}_k = 0$. With this change of variables, the quadratic program (6.1) can be rewritten as a purely control constrained problem of the form

$$\text{minimize } \frac{1}{2}\mathcal{Q}(\mathbf{Z}(\mathbf{U}), \mathbf{U}) + \overline{\mathcal{L}}(\mathbf{Z}(\mathbf{U}), \mathbf{U}, \mathbf{Y}) \quad \text{subject to} \quad \mathbf{U}_{ki} \in \mathcal{U}, \tag{7.1}$$

where $1 \le i \le N$, $1 \le k \le K$, and

$$\begin{aligned}\overline{\mathcal{L}}(\mathbf{Z}, \mathbf{U}, \mathbf{Y}) = \mathcal{L}(\mathbf{Z}, \mathbf{U}, \mathbf{Y}) + \boldsymbol{\chi}_{KN}(\mathbf{Y})^{\mathsf{T}}\mathbf{T}\mathbf{Z}_{KN} \\ + h\sum_{k=1}^{K}\sum_{i=1}^{N} \omega_i \left[\boldsymbol{\chi}_{ki}^{\mathsf{T}}(\mathbf{Y})\mathbf{Q}_{ki}\mathbf{Z}_{ki} + \boldsymbol{\chi}_{ki}^{\mathsf{T}}(\mathbf{Y})\mathbf{S}_{ki}\mathbf{U}_{ki}\right].\end{aligned} \tag{7.2}$$

If $[\mathbf{Z}^j, \mathbf{U}^j] = [\mathbf{Z}(\mathbf{U}^j), \mathbf{U}^j]$ denotes the solution of (7.1) corresponding to $\mathbf{Y}^j \in \mathcal{Y}$, $j = 1$ and 2, then by [15, Lem. 4], the solution change $\Delta\mathbf{U} = \mathbf{U}^1 - \mathbf{U}^2$ satisfies the

relation

$$\mathcal{Q}(\Delta\mathbf{Z}, \Delta\mathbf{U}) \leq |\overline{\mathcal{L}}(\Delta\mathbf{Z}, \Delta\mathbf{U}, \Delta\mathbf{Y})|, \tag{7.3}$$

where $\Delta\mathbf{Y} = \mathbf{Y}^1 - \mathbf{Y}^2$ and $\Delta\mathbf{Z} = \mathbf{Z}^1 - \mathbf{Z}^2$. Observe that the quadratic $\mathcal{Q}$ in (6.2) is expressed in terms of the Hessian with respect to $\mathbf{x}$ and $\mathbf{u}$ of the Hamiltonian H evaluated at $(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki}), \boldsymbol{\lambda}^*(t_{ki}))$; by (A1) it follows that

$$\mathcal{Q}(\Delta\mathbf{Z}, \Delta\mathbf{U}) \geq \alpha h \|\Delta\mathbf{U}\|_\omega^2. \tag{7.4}$$

Now consider the terms in $\mathcal{L}$. Let c denote a generic constant which is independent of K and N. By Lemma 5.1 and Remark 5.1, $\|\Delta\mathbf{Z}\|_\infty \leq ch^{1/2}\|\Delta\mathbf{U}\|_\omega$. By the Schwarz inequality, the bound for $\|\Delta\mathbf{Z}\|_\infty$, and the fact that the ω_i are positive and sum to 2, we have

$$\begin{aligned}\left|\sum_{k=1}^{K}\sum_{i=1}^{N}\omega_i\,\Delta\mathbf{y}_{3ki}^{\mathsf{T}}\,\Delta\mathbf{Z}_{ki}\right| &\leq ch^{1/2}\|\Delta\mathbf{U}\|_\omega \sum_{k=1}^{K}\sum_{i=1}^{N}\omega_i|\Delta\mathbf{y}_{3ki}| \\ &\leq ch^{1/2}\|\Delta\mathbf{U}\|_\omega \sum_{k=1}^{K}\|\Delta\mathbf{y}_{3k}\|_\omega \\ &\leq c\|\Delta\mathbf{U}\|_\omega\|\Delta\mathbf{y}_3\|_\omega \leq c\|\Delta\mathbf{U}\|_\omega\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.\end{aligned}$$

Similarly, for the $\mathbf{y}_6$ term in $\mathcal{L}$, the Schwarz inequality gives

$$\begin{aligned}\left|\sum_{k=1}^{K}\sum_{i=1}^{N}\omega_i\,\Delta\mathbf{y}_{6ki}^{\mathsf{T}}\,\Delta\mathbf{U}_{ki}\right| &\leq c\|\Delta\mathbf{y}_6\|_\infty \sum_{k=1}^{K}\sum_{i=1}^{N}\omega_i|\Delta\mathbf{U}_{ki}| \leq c\|\Delta\mathbf{y}_6\|_\infty \sum_{k=1}^{K}\|\Delta\mathbf{U}_k\|_\omega \\ &\leq ch^{-1/2}\|\Delta\mathbf{y}_6\|_\infty\|\Delta\mathbf{U}\|_\omega \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}\|\Delta\mathbf{U}\|_\omega.\end{aligned}$$

The last inequality is due to the $h^{-1/2}\|\Delta\mathbf{y}_6\|_\infty$ term in $\|\Delta\mathbf{Y}\|_{\mathcal{Y}}$.

For the $\mathbf{Z}_{k0}$-term in $\mathcal{L}$, the bound $\|\Delta\mathbf{Z}\|_\infty \leq ch^{1/2}\|\Delta\mathbf{U}\|_\omega$ implies that

$$\begin{aligned}&\sum_{k=1}^{K}\left|\Delta\mathbf{Z}_{k0}^{\mathsf{T}}\left(\Delta\mathbf{y}_{4k} + \sum_{i=1}^{N}\omega_i\,\Delta\mathbf{y}_{3ki}\right)\right| \\ &\quad\leq ch^{1/2}\|\Delta\mathbf{U}\|_\omega \sum_{k=1}^{K}\left(|\Delta\mathbf{y}_{4k}| + \sum_{i=1}^{N}\omega_i|\Delta\mathbf{y}_{3ki}|\right) \\ &\quad\leq c\|\Delta\mathbf{U}\|_\omega\,(|\Delta\mathbf{y}_4| + \|\Delta\mathbf{y}_3\|_\omega) \leq c\|\Delta\mathbf{U}\|_\omega\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.\end{aligned}$$

By Lemma 5.1, we have

$$\|\boldsymbol{\chi}(\Delta\mathbf{Y})\|_\infty \leq ch^{-1/2}(\|\Delta\mathbf{y}_1\|_\omega + |\Delta\mathbf{y}_2|) \leq ch^{-1/2}\|\Delta\mathbf{Y}\|_{\mathcal{Y}}. \tag{7.5}$$

Hence, the $\boldsymbol{\chi}$ terms in (7.2) have a bound such as

$$h\left|\sum_{k=1}^{K}\sum_{i=1}^{N}\omega_i\boldsymbol{\chi}_{ki}^{\mathsf{T}}(\Delta\mathbf{Y})\mathbf{Q}_{ki}\Delta\mathbf{Z}_{ki}\right| \le ch\|\boldsymbol{\chi}(\Delta\mathbf{Y})\|_\infty\|\Delta\mathbf{Z}\|_\infty\sum_{k=1}^{K}\sum_{i=1}^{N}\omega_i$$
$$= c\|\boldsymbol{\chi}(\Delta\mathbf{Y})\|_\infty\|\Delta\mathbf{Z}\|_\infty \le c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}\|\Delta\mathbf{U}\|_\omega.$$

For the terminal term in (7.2), we have the bound

$$|\boldsymbol{\chi}_{KN}(\Delta\mathbf{Y})^{\mathsf{T}}\mathbf{T}\Delta\mathbf{Z}_{KN}| \le c\|\boldsymbol{\chi}(\Delta\mathbf{Y})\|_\infty\|\Delta\mathbf{Z}\|_\infty \le c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}\|\Delta\mathbf{U}\|_\omega.$$

The $\mathbf{y}_5$ term in $\mathcal{L}$ is similar. By the Schwarz inequality,

$$|\Delta\mathbf{y}_5^{\mathsf{T}}\Delta\mathbf{Z}_{KN}| \le c\|\Delta\mathbf{Z}\|_\infty|\Delta\mathbf{y}_5| \le ch^{1/2}\|\Delta\mathbf{U}\|_\omega|\Delta\mathbf{y}_5| \le c\|\Delta\mathbf{U}\|_\omega\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.$$

Note that $h^{1/2}|\Delta\mathbf{y}_5|$ is one of the terms in $\|\Delta\mathbf{Y}\|_{\mathcal{Y}}$. Combine these bounds for the linear term to obtain

$$|\overline{\mathcal{L}}(\Delta\mathbf{Z},\,\Delta\mathbf{U},\,\Delta\mathbf{Y})| \le c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}\|\Delta\mathbf{U}\|_\omega.$$

Hence, (7.4) implies that

$$\|\Delta\mathbf{U}\|_\omega \le ch^{-1}\|\Delta\mathbf{Y}\|_{\mathcal{Y}}. \tag{7.6}$$

Next, the ω-norm on the left side of (7.6) will be converted to an ∞-norm. Due to the identity $\Delta\mathbf{X} = \Delta\mathbf{Z} + \boldsymbol{\chi}(\Delta\mathbf{Y})$ and the bounds (7.5) and (7.6), it follows that

$$\|\Delta\mathbf{X}\|_\infty \le \|\Delta\mathbf{Z}\|_\infty + \|\boldsymbol{\chi}(\Delta\mathbf{Y})\|_\infty$$
$$\le ch^{1/2}\|\Delta\mathbf{U}\|_\omega + ch^{-1/2}\|\Delta\mathbf{Y}\|_{\mathcal{Y}} \le ch^{-1/2}\|\Delta\mathbf{Y}\|_{\mathcal{Y}}. \tag{7.7}$$

Also, using (7.7), we have

$$\|\Delta\mathbf{X}\|_\omega \le h^{-1/2}\|\Delta\mathbf{X}\|_\infty \le ch^{-1}\|\Delta\mathbf{Y}\|_{\mathcal{Y}}. \tag{7.8}$$

Apply Lemma 5.2 with

$$\mathbf{p}_{ki} = -\Delta\mathbf{y}_{3ki} - h(\mathbf{Q}_{ki}\Delta\mathbf{X}_{ki} + \mathbf{S}_{ki}\Delta\mathbf{U}_{ki}), \quad \Delta\boldsymbol{\Lambda}_{K+1,0} = \mathbf{T}\Delta\mathbf{X}_{KN} + \Delta y_5, \quad \text{and}$$
$$\mathbf{q}_k = \sum_{i=1}^{N}\omega_i\left(h\left[\mathbf{Q}_{ki}\Delta\mathbf{X}_{ki} + \mathbf{S}_{ki}\Delta\mathbf{U}_{ki}\right] - \Delta\mathbf{y}_{4ki}\right).$$

By (5.15), we have

$$\|\Delta\boldsymbol{\Lambda}\|_\infty \le c\left(\|\Delta\mathbf{y}_5\|_\infty + \|\Delta\mathbf{X}_{KN}\|_\infty + h^{-1/2}\|\mathbf{p}\|_\omega + \sum_{k=1}^{K}|\mathbf{q}_k|\right). \tag{7.9}$$

By (7.7), $\|\Delta \mathbf{X}_{KN}\|_\infty \le ch^{-1/2}\|\Delta \mathbf{Y}\|_{\mathcal{Y}}$. Taking the ω-norm of $\mathbf{p}$ and using (7.6) and (7.8) gives

$$\|\mathbf{p}\|_\omega \le \|\Delta \mathbf{y}_3\|_\omega + h\|\Delta \mathbf{X}\|_\omega + h\|\Delta \mathbf{U}\|_\omega \le c\|\Delta \mathbf{Y}\|_{\mathcal{Y}}.$$

Finally, (7.6) and (7.8) and the Schwarz inequality yield

$$\begin{aligned}\sum_{k=1}^{K} |\mathbf{q}_k| &\le c\sum_{k=1}^{K} (\|\Delta \mathbf{y}_{4k}\|_\omega + h\|\Delta \mathbf{X}_k\|_\omega + h\|\Delta \mathbf{U}_k\|_\omega)\\ &\le ch^{-1/2}\|\Delta \mathbf{y}_4\|_\omega + h^{1/2}\left(\|\Delta \mathbf{X}\|_\omega + \|\Delta \mathbf{U}\|_\omega\right)] \le ch^{-1/2}\|\Delta \mathbf{Y}\|_\omega.\end{aligned}$$

Inserting these bounds in (7.9), we have

$$\|\Delta \mathbf{\Lambda}\|_\infty \le ch^{-1/2}\|\Delta \mathbf{Y}\|_{\mathcal{Y}}. \tag{7.10}$$

Recall that $\mathbf{R}_{ki} := \nabla^2_{uu} H(\mathbf{x}^*(t_{ki}), \mathbf{u}^*(t_{ki}), \boldsymbol{\lambda}^*(t_{ki}))$. By (A1) $\mathbf{R}_{ki}$ is positive definite with smallest eigenvalue greater than or equal to α. It follows from the 6th component of the inclusion (5.1) that the control solves the quadratic program

$$\min_{\mathbf{U}_{ki}\in\mathcal{U}} \; h\left(\frac{1}{2}\mathbf{U}_{ki}^{\mathsf{T}}\mathbf{R}_{ki} + \mathbf{X}_{ki}^{\mathsf{T}}\mathbf{S}_{ki} + \mathbf{\Lambda}_{ki}^{\mathsf{T}}\mathbf{B}_{ki}\right)\mathbf{U}_{ki} + \mathbf{y}_{6ki}^{\mathsf{T}}\mathbf{U}_{ki}.$$

Again by [15, Lem. 4], the solution change associated with the data change $\Delta \mathbf{Y}$ has the bound

$$h\alpha|\Delta \mathbf{U}_{ki}|^2 \le \left|h\left(\Delta \mathbf{X}_{ki}^{\mathsf{T}}\mathbf{S}_{ki} + \Delta \mathbf{\Lambda}_{ki}^{\mathsf{T}}\mathbf{B}_{ki}\right)\Delta \mathbf{U}_{ki} + \Delta \mathbf{y}_{6ki}\Delta \mathbf{U}_{ki}\right|.$$

Hence, we deduce that

$$\|\Delta \mathbf{U}_{ki}\|_\infty \le |\Delta \mathbf{U}_{ki}| \le c\left(\|\Delta \mathbf{X}_{ki}\|_\infty + \|\Delta \mathbf{\Lambda}_{ki}\|_\infty + h^{-1}\|\Delta \mathbf{y}_{6ki}\|_\infty\right).$$

Utilizing the bounds (7.7) and (7.10), and the $h^{-1/2}$ factor associated with the 6-th component of the $\mathcal{Y}$-norm, yields

$$\|\Delta \mathbf{U}_{ki}\|_\infty \le ch^{-1/2}\|\Delta \mathbf{Y}\|_{\mathcal{Y}}. \tag{7.11}$$

The bounds (7.7), (7.10), and (7.11) combine to establish the following Lipschitz continuity property:

Lemma 7.1 *If* (A1) *and* (A2) *hold, then there exists a unique solution of* (5.1) *for each* $\mathbf{Y} \in \mathcal{Y}$*, and there exists a constant* c*, independent of* K *and* N*, such that the solution change* $\Delta \mathbf{X}$*,* $\Delta \mathbf{U}$*, and* $\Delta \mathbf{\Lambda}$ *relative to the change* $\Delta \mathbf{Y}$ *satisfies*

$$\|(\Delta \mathbf{X}, \Delta \mathbf{U}, \Delta \mathbf{\Lambda})\|_\infty \le ch^{-1/2}\|\Delta \mathbf{Y}\|_{\mathcal{Y}}.$$

Theorem 1.1 is proved using Proposition 2.2. The Lipschitz constant γ of Proposition 2.2 is given by $\gamma = ch^{-1/2}$ where c is the constant of Lemma 7.1. The terms involving $\mathbf{D}$, $\mathbf{D}^{\ddagger}$, $\mathbf{\Lambda}_{k0}$, $\mathbf{\Lambda}_{k+1,0}$, $\mathbf{X}_{k0}$, and $\mathbf{X}_{k-1,N}$ are constants in the derivative $\nabla\mathcal{T}$ and hence these terms cancel when we compute the difference $\nabla\mathcal{T}(\boldsymbol{\theta}) - \nabla\mathcal{T}(\boldsymbol{\theta}^*)$, where $\boldsymbol{\theta} = (\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ and $\boldsymbol{\theta}^* = (\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$. We are left with terms involving the difference of derivatives of $\mathbf{f}$ or C up to second order at points in a neighborhood of $\boldsymbol{\theta}^*$. By the Smoothness assumption, these derivatives are Lipschitz continuous in a neighborhood of $(\mathbf{X}^*, \mathbf{U}^*)$. Hence, there exists constants τ and $r > 0$ such that

$$
\begin{aligned}
&\|\nabla[\mathbf{f}(\mathbf{X}_{ki}, \mathbf{U}_{ki}) - \mathbf{f}(\mathbf{X}^*_{ki}, \mathbf{U}^*_{ki})]\|_\infty \le \tau\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty,\\
&\|\nabla[\nabla_x\mathbf{H}(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}) - \nabla_x\mathbf{H}(\mathbf{X}^*_{ki}, \mathbf{U}^*_{ki}, \mathbf{\Lambda}^*_{ki})]\|_\infty \le \tau\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty,\\
&\|\nabla[\nabla_u\mathbf{H}(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}) - \nabla_u\mathbf{H}(\mathbf{X}^*_{ki}, \mathbf{U}^*_{ki}, \mathbf{\Lambda}^*_{ki})]\|_\infty \le \tau\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty,\\
&\|\nabla[\nabla C(\mathbf{X}_{KN}) - \nabla C(\mathbf{X}^*_{KN})]\|_\infty \le \tau\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty,
\end{aligned}
$$

whenever $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \le r$. In applying Proposition 2.2, we need a bound for the $\mathcal{Y}$-norm of $\nabla\mathcal{T}(\boldsymbol{\theta}) - \nabla\mathcal{T}(\boldsymbol{\theta}^*)$. Taking into account the location of h's in $\mathcal{T}$ and the location of h's in the $\mathcal{Y}$-norm, it follows from the Lipschitz bounds relative to τ that there exists a constant κ such that

$$
\|\nabla\mathcal{T}(\boldsymbol{\theta}) - \nabla\mathcal{T}(\boldsymbol{\theta}^*)\|_{\mathcal{Y}} \le \kappa h^{1/2}\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty,
$$

whenever $\|\boldsymbol{\theta} - \boldsymbol{\theta}^*\|_\infty \le r$. Choose $r > 0$ smaller if necessary to ensure that $c\kappa r < 1$, where c is the constant in Lemma 7.1. In Proposition 2.2, $\epsilon = \kappa h^{1/2} r$ and $\gamma = ch^{-1/2}$. Hence, $\gamma\epsilon = c\kappa r < 1$. Referring to Lemma 4.1, choose N large enough or h small enough so that

$$
\text{dist}[\mathcal{T}(\boldsymbol{\theta}^*), \mathcal{F}(\mathbf{U}^*)] \le \frac{(1 - \gamma\epsilon)r}{\gamma}.
$$

Combine Lemma 4.1 with (2.26) and the formula $\gamma = ch^{-1/2}$ to obtain the bound (1.12) of Theorem 1.1.

The solution to $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$ corresponds to the first-order optimality condition for either (2.2) or (1.4). We use the second-order sufficient optimality conditions to show that this stationary point is a local minimum when it is sufficiently close to $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$. After replacing the KKT multipliers by the transformed quantities given by $\mathbf{\Lambda}_{ki} = \boldsymbol{\lambda}_{ki}/\omega_i$, the Hessian of the Lagrangian is a block diagonal matrix with the following matrices forming the diagonal blocks:

$$
\begin{aligned}
&\omega_i\nabla^2_{(x,u)}H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}), && 1 \le i < N,\\
&\omega_i\nabla^2_{(x,u)}H(\mathbf{X}_{ki}, \mathbf{U}_{ki}, \mathbf{\Lambda}_{ki}) + \nabla^2_{(x,u)}C(\mathbf{X}_{ki}), && i = N,
\end{aligned}
$$

where H is the Hamiltonian and $1 \le k \le K$. The second-order sufficient optimality condition involves showing that the quadratic associated with the Hessian is positive definite when applied to controls of the form $\overline{\mathbf{U}} = \mathbf{V} - \mathbf{W}$, where $\mathbf{V}_{ki}$ and $\mathbf{W}_{ki}$

lie in $\mathcal{U}$ for each k and i, and $\overline{\mathbf{X}}$ is the solution of the linearized discrete dynamics associated with the control $\overline{\mathbf{U}}$ and the initial condition $\overline{\mathbf{X}}_{0N} = \mathbf{0}$. When the Hessian and the linearized dynamics are evaluated at $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) = (\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$, then the positive definiteness is a consequence of (7.4). On the other hand, when $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ is close to $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$, then the matrices in the Hessian and in the linearized dynamics are all close to the matrices corresponding to $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$. Consequently, positives definiteness is preserved for $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ sufficiently close to $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$, and by the second-order sufficient optimality condition [42, Thm. 12.6], $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ is a strict local minimizer. This completes the proof of Theorem 1.1.

8 Numerical illustrations

In this section we analyze the errors associated with the proposed Radau hp-collocation method using numerical examples with known analytic solutions. Consequently, it is possible to precisely determine the error in the hp-approximations. More complex examples, which do not have known analytic solutions, appear in both [45] and at the GPOPS-II examples website: http://www.gpops2.com/Examples/Examples.html. In [45] it is observed that the solutions computed by Radau hp-collocation are in close agreement to the solutions computed by Betts' Sparse Optimization Suite (SOS) [6].

8.1 Example 1

First we consider the unconstrained control problem given by

$$\min \left\{ -x(2) : \dot{x}(t) = \frac{5}{2}(-x(t) + x(t)u(t) - u^2(t)), \ x(0) = 1 \right\}. \tag{8.1}$$

The optimal solution and associated costate are

$$\begin{aligned} x^*(t) &= 4/a(t), \quad a(t) = 1 + 3\exp(2.5t), \\ u^*(t) &= x^*(t)/2, \\ \lambda^*(t) &= -a^2(t)\exp(-2.5t)/[\exp(-5) + 9\exp(5) + 6]. \end{aligned}$$

The time domain [0,2] is divided into equally spaced mesh intervals, and on each mesh interval, we collocate at the Radau points using polynomials of the same degree. We consider polynomials of degree $N = 2$, 3, and 4. Convergence to the true solution is achieved by increasing the number of mesh intervals. Figure 1 plots the base 10 logarithm of the error at the collocation points in the sup-norm versus the base 10 logarithm of mesh size. The results were obtained using the software GPOPS-II [44] and the optimizer IPOPT [7] to solve the discrete nonlinear program. The markers plotted in Figure 1 correspond to the sup-norm error at a given value for h, while the lines have slope $N + 2$ for the state and control, and $N + 1$ for the costate. The vertical placement of each line yields the least squares fit to the markers. Observe that the

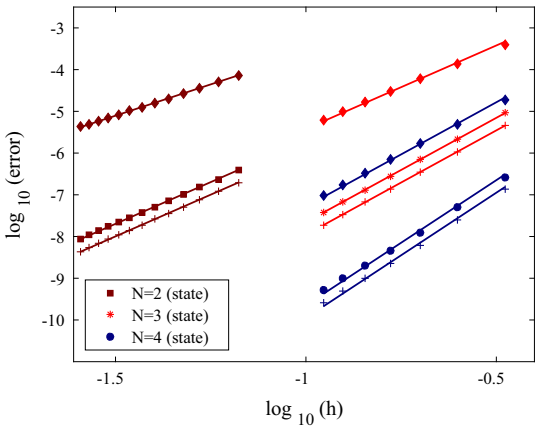

Fig. 1 The logarithm of the sup-norm error in Example 1 as a function of mesh size for polynomials of degree $N = 2$, 3, and 4. The errors in the controls, marked by plus signs, are beneath the state error plots. The errors in the costate, marked by diamonds, are above the state error plots

error decays roughly linearly in this log–log plot, and the pointwise error is roughly $O(h^{N+2})$ in the state and control, and $O(h^{N+1})$ in the costate for fixed N.

The bound given in Theorem 1.1 for fixed N is $O(h^{N-1})$, which is much slower than the observed convergence rate $O(h^{N+1})$. This discrepancy could be due to either the simple nature of the example, or to looseness in the analysis. In our analysis, the exponent of h is reduced by the following effects:

(a) Although the state is approximated by a polynomial of degree N, the costate is approximated by a polynomial of degree $N - 1$. This difference between the state and the costate becomes apparent in Proposition 2.1. We are not free to choose the costate polynomial, its degree comes from the KKT conditions. In the analysis of the residual given in Lemma 4.1, the reduced degree for the costate polynomial implies that the exponent of h in the bound (4.1) is the minimum of N and η rather than the minimum of $N + 1$ and η.

(b) In our analysis at the end of Section 7, we showed that by taking r small enough, the expression $\gamma\epsilon$ in the denominator of (2.26) was strictly bounded from one. The analysis also showed that that the Lipschitz constant satisfied $\gamma \le ch^{-1/2}$. Hence, we lose a half power of h through the Lipschitz constant in the error bound (2.26).

If example 1 indeed represents the typical behavior of the error, then the analysis must be sharpened to address the losses described in (a) and (b).

It is interesting to compare the analysis in this paper to the analysis of Runge–Kutta schemes given in [8,31]. For a fixed N, the Radau scheme in this paper is equivalent to a Runge–Kutta scheme where the $\mathbf{A}$ matrix and $\mathbf{b}$ vector of [31] describing the Runge–Kutta scheme are $\mathbf{D}_{1:N}^{-1}/2$ and the last row of $\mathbf{D}_{1:N}^{-1}/2$ respectively. For $N = 2$ and $N = 3$, the corresponding Runge–Kutta schemes have order 3 and 4 respectively, which means that the error in the Runge–Kutta schemes are $O(h^3)$ and $O(h^4)$ respectively. This exactly matches the costate error for the hp-scheme in this example. A fundamental difference between the results of [31] and the results in this paper is that [31] estimates the error at the mesh points, and there is no information about the error at the intermediate points, while in Theorem 1.1, we estimate the error at both collocation and mesh points. In the hp-framework, it is important to have

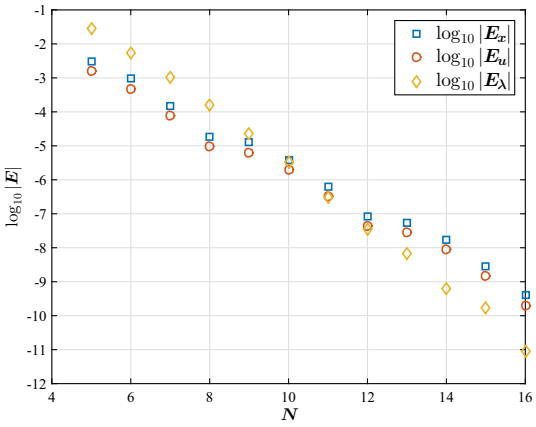

Fig. 2 The base 10 logarithm of the error in the sup-norm as a function of the number of collocation points for Example 1

estimates at the collocation points since K could be fixed, and the convergence is achieved by letting N grow.

Based on the theory developed in the paper [8] of Bonnans and Laurent-Varin, many conditions must be satisfied to achieve high order convergence of a Runge–Kutta scheme for optimal control (4116 conditions for order 7). Potentially, the *hp*-scheme based on Radau collocation could be used to generate high order Runge–Kutta schemes.

Next, we examine in Figure 2 the exponential convergence rate predicted by Theorem 1.1 when there is a single interval and the degree of the polynomials is increased. Since the plot of the base-10 logarithm of the error versus the degree of of the polynomial is nearly linear, the error behaves like $c10^{-\alpha N}$ where $\alpha \approx 0.6$ for either the state or the control and $\alpha \approx 0.8$ for the costate. Since the solution to this problem is infinitely smooth, we can take $\eta = N$ in Theorem 1.1. The error bound in Theorem 1.1 is somewhat complex since it involves the derivatives of the solution. Nonetheless, when we take the base-10 logarithm of the error bound, the asymptotically dominant term appears to be $-N \log_{10} N$ for Example 1. Consequently, the slope of the curve in the error bound varies like $-\log_{10} N$. For N between 4 and 16, $\log_{10} N$ varies from about 0.6 to 1.2. Hence, our observed slopes 0.6 and 0.8 fall in the anticipated range.

8.2 Example 2

Next we consider the problem [35] given by

$$\begin{aligned} \text{minimize} \quad & \frac{1}{2}\int_0^1 [x^2(t) + u^2(t)]\, dt \\ \text{subject to} \quad & \dot{x}(t) = u(t), \quad u(t) \le 1, \quad x(t) \le \frac{2\sqrt{e}}{1-e} \quad \text{for all} \quad t \in [0,1], \\ & x(0) = \frac{5e+3}{4(1-e)}. \end{aligned}$$

The exact solution to this problem is

$$
\begin{aligned}
&0 \le t \le \tfrac{1}{4} : x^*(t) = t - \tfrac{1}{4} + \tfrac{1+e}{1-e}, && u^*(t) = 1,\\
&\tfrac{1}{4} \le t \le \tfrac{3}{4} : x^*(t) = \tfrac{e^{t-\frac{1}{4}}}{1-e}\left(1 + e^{\frac{3}{2}-2t}\right), && u^*(t) = \tfrac{e^{t-\frac{1}{4}}}{1-e}\left(1 - e^{\frac{3}{2}-2t}\right),\\
&\tfrac{3}{4} \le t \le 1 : x^*(t) = \tfrac{2\sqrt{e}}{1-e}, && u^*(t) = 0.
\end{aligned}
$$

The solution of this problem is smooth on the three intervals [0, 0.25], [0.25, 0.75], and [0.75, 1.0], however, at the contact points where one of the constraints changes from active to inactive, there is a discontinuity in the derivative of the optimal control and a discontinuity in the second derivative of the optimal state. The goal with this test problem is to determine whether exponential convergence occurs for the *hp*-scheme with a careful choice of the mesh, and whether a state constrained problem, which is not covered by the error analysis in this paper, possesses similar errors bounds to those for control constrained problems.

First, we solve the problem using $K = 1$, in which case convergence is achieved by increasing the degree N of the polynomials. In Fig. 3a we plot the logarithm of the error at the collocation points in the sup-norm versus the logarithm of the polynomial degree. Convergence occurs, but it is slow due to the discontinuity in the derivatives. The lines in Fig. 3a have slope -2; their vertical placement was chosen to achieve the best least squares fit to the markers (the measured error). Since the logarithm of the error is approximately fit by a line of slope -2, the error decays like c/N^2, which is faster than what might be expected from a bound like that given in Theorem 1.1 with regularity $\mathcal{H}^{2.5-\epsilon}$ for any $\epsilon > 0$.

Next, we divide the time interval [0,1] into three subintervals [0, 0.25], [0.25, 0.75], and [0.75, 1.0], and use different polynomials of the same degree on each subinterval. By this careful choice of the mesh intervals, we obtain an exponential convergence rate in Fig. 3b. Comparing Figs. 3a and b, we see that a huge improvement in the error is possible when we have good estimates for the contact points where the constraints

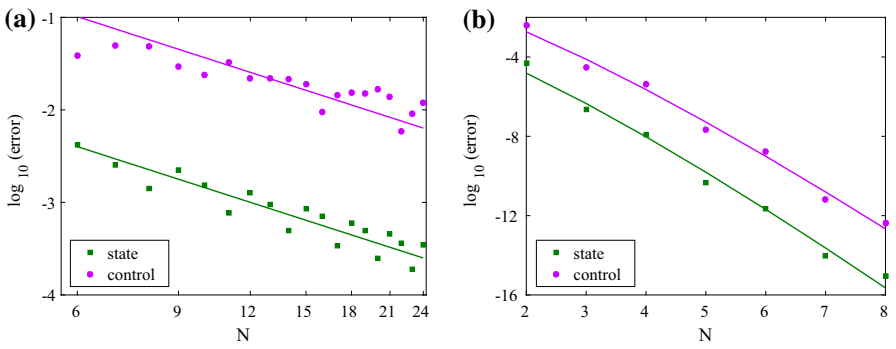

Fig. 3 The error in the solution to Example 2 as a function of the degree of the polynomials used in the *hp*-approximation. In **a** the polynomials are defined on the interval [0, 1]. In **b** there are three mesh intervals [0, 0.25], [0.25, 0.75], and [0.75, 1.0], and different polynomials of the same degree are used on each mesh interval

change between active and inactive. Note that 16-digit accuracy was obtained in Fig. 3b by using MATLAB's QUADPROG to solve the quadratic program associated with the hp-discretization of Example 2.

In a very rough sense, the error bound given by Theorem 1.1 for a smooth problem has the general form $c_1(c_2/N)^N$. The continuous curves plotted in Fig. 3b were obtained by choosing c_1 and c_2 to achieve the least squares best fit to the markers (the measured error). For the state variable, $(c_1, c_2) = (0.0016, 0.1990)$, while for the control $(c_1, c_2) = (0.0950, 0.2801)$. Hence, it seems plausible that a state-constrained control problem may possess an error bound similar to that established in Theorem 1.1 for control constrained problems.

9 Conclusions

A convergence rate is derived for an hp-orthogonal collocation method based on the Radau quadrature points applied to a control problem with convex control constraints. If the problem has a smooth local solution and a Hamiltonian which satisfies a strong convexity assumption, then the discrete approximation has a local minimizer in a neighborhood of the continuous solution. For the hp-scheme, both the number of mesh intervals in the discretization and the degree of the polynomials on each mesh interval can be freely chosen. As the number of mesh intervals increases, convergence occurs at a polynomial rate relative to the mesh width. When there is control over the growth in derivatives, the convergence rate is exponentially fast relative to the polynomial degree. Convergence rates were investigated further using numerical examples. When the polynomial degree is fixed and the mesh width tends to zero, the observed convergence rate was faster than the rate associated with the error bound. For a problem with control and state constraints, exponentially fast convergence was observed when mesh points are located at the contact points where the constraints change between active and inactive. Based on the numerical results, it seems plausible that the convergence result established for control constrained problem could extend to problems with state constraints.

10 Appendix: Proof of (P1) and (P2)

We analyze (P1) and (P2) when τ_i, $1 \le i \le N$, are either the Radau quadrature points analyzed in this paper, or the Gauss quadrature points studied in [36].

Lemma 10.1 *For either the Gauss or Radau quadrature points, the rows of the matrix* $[\mathbf{W}^{1/2}\mathbf{D}_{1:N}]^{-1}$ *have Euclidean length bounded by* $\sqrt{2}$. *For the Gauss quadrature points,* $\|\mathbf{D}_{1:N}^{-1}\|_\infty \le 2$, *and* $\|\mathbf{D}_{1:N}^{-1}\|_\infty$ *approaches 2 as N tends to infinity, while for the Radau quadrature points,* $\|\mathbf{D}_{1:N}^{-1}\|_\infty = 2$.

Proof Given $\mathbf{p} \in \mathbb{R}^N$, let $p \in \mathcal{P}_N$ denote the polynomial that satisfies $p(-1) = 0$ and $p(\tau_i) = p_i$, $1 \le i \le N$. Let $\dot{\mathbf{p}} \in \mathbb{R}^N$ denote the vector with components $\dot{p}_i = \dot{p}(\tau_i)$, and let ℓ_j be the Lagrange polynomial defined by

$$\ell_j(\tau) = \prod_{\substack{i=1 \\ i \neq j}}^{N} \frac{\tau - \tau_i}{\tau_j - \tau_i}, \quad 1 \le j \le N.$$

The identity

$$\dot{p}(\tau) = \sum_{j=1}^{N} \ell_j(\tau) \dot{p}_j \tag{10.1}$$

holds since $\dot{p} \in \mathcal{P}_{N-1}$ and the polynomials on each side of (10.1) are equal at the N quadrature points. Integrate (10.1) to obtain

$$p_i = \int_{-1}^{\tau_i} \dot{p}(\tau)\, d\tau = \sum_{j=1}^{N} \left(\int_{-1}^{\tau_i} \ell_j(\tau)\, d\tau \right) \dot{p}_j. \tag{10.2}$$

Since $\mathbf{D}$ is a differentiation matrix and $p(-1) = 0$, it follows that $\mathbf{D}_{1:N}\mathbf{p} = \dot{\mathbf{p}}$. If the vector $\dot{\mathbf{p}} = \mathbf{0}$, then the polynomial $\dot{p} = 0$ since $\dot{p}$ has degree $N-1$ and vanishes at N points. Since $p(-1) = 0$, it follows that polynomial $p = 0$, which implies that the vector $\mathbf{p} = \mathbf{0}$. Hence, $\mathbf{D}_{1:N}$ is invertible, and $\mathbf{p} = \mathbf{D}^{-1}\dot{\mathbf{p}}$. Comparing the equality $\mathbf{p} = \mathbf{D}^{-1}\dot{\mathbf{p}}$ to (10.2), we deduce that

$$(\mathbf{D}^{-1})_{ij} = \int_{-1}^{\tau_i} \ell_j(\tau)\, d\tau. \tag{10.3}$$

Choose any $s \in [-1, 1]$ and define

$$d_j(s) = \int_{-1}^{s} \ell_j(\tau)\, d\tau \quad \text{and} \quad R(s) = \sum_{j=1}^{N} \frac{d_j(s)^2}{\omega_j}.$$

Observe that $(\mathbf{D}^{-1})_{ij} = d_j(\tau_i)$ and $R(\tau_i)$ is the square of the Euclidean length of row i in $(\mathbf{W}^{1/2}\mathbf{D})^{-1}$. Let $q \in \mathcal{P}_{N-1}$ be the polynomial defined by

$$q(\tau) = \sum_{j=1}^{N} \frac{d_j(s)\ell_j(\tau)}{\omega_j}.$$

Hence, by the triangle and Schwarz inequalities,

$$R(s) = \int_{-1}^{s} q(\tau)\, d\tau \le \int_{-1}^{1} |q(\tau)|\, d\tau \le \sqrt{2} \left(\int_{-1}^{1} q(\tau)^2\, d\tau \right)^{1/2}. \tag{10.4}$$

Since $q^2 \in \mathcal{P}_{2N-2}$, both Radau and Gauss quadrature are exact, and

$$\int_{-1}^{1} q(\tau)^2\, d\tau = \sum_{i=1}^{N} \omega_i q(\tau_i)^2, \tag{10.5}$$

where the τ_j are either the Radau or Gauss quadrature points and the ω_j are the associated weights. Since $\ell_j(\tau_i) = 1$ for $i = j$ and $\ell_j(\tau_i) = 0$ otherwise, it follows from the definition of q that $q(\tau_i) = d_i(s)/\omega_i$. This substitution in (10.5) yields

$$\int_{-1}^{1} q(\tau)^2 \, d\tau = \sum_{i=1}^{N} \frac{d_i(s)^2}{\omega_j} = R(s). \tag{10.6}$$

Equating the expressions (10.4) and (10.6) implies that

$$\left(\int_{-1}^{1} q(\tau)^2 \, d\tau \right)^{1/2} \leq \sqrt{2}.$$

By (10.6), $R(s) \leq 2$ for any $s \in [-1, 1]$. In particular, $R(\tau_i) \leq 2$ for $1 \leq i \leq N$. Since $R(\tau_i)$ is the square of the Euclidean length of row i in $(\mathbf{W}^{1/2}\mathbf{D})^{-1}$, the rows of $(\mathbf{W}^{1/2}\mathbf{D})^{-1}$ have Euclidean length bounded by $\sqrt{2}$. This result holds for both the Radau and Gauss quadrature points since since $q^2 \in \mathcal{P}_{2N-2}$, and both Radau and Gauss quadrature are exact for polynomials of this degree.

If $\mathbf{r}$ is a row of $\mathbf{D}_{1:N}^{-1}$, then by the Schwarz inequality and the fact that the quadrature weights sum to 2 and the rows of the matrix $[\mathbf{W}^{1/2}\mathbf{D}_{1:N}]^{-1}$ have Euclidean length bounded by $\sqrt{2}$, we have

$$\sum_{i=1}^{N} |r_i| = \sum_{i=1}^{N} \sqrt{\omega_i} \left(|r_i|/\sqrt{\omega_i} \right) \leq \left(\sum_{i=1}^{N} \omega_i \right)^{1/2} \left(\sum_{i=1}^{N} r_i^2/\omega_i \right)^{1/2} \leq 2. \tag{10.7}$$

Consequently, the absolute row sums for $\mathbf{D}_{1:N}^{-1}$ are all bounded by 2, or equivalently, $\|\mathbf{D}_{1:N}^{-1}\|_\infty \leq 2$. Given any polynomial $p \in \mathcal{P}_N$ with $p(-1) = 0$ and $|\dot{p}(\tau_i)| \leq 1$ for $1 \leq i \leq N$, it is observed in Section 9 of [36] that $\|\mathbf{D}_{1:N}^{-1}\|_\infty \geq \max\{p(\tau_i) : 1 \leq i \leq N\}$. Take $p(\tau) = 1 + \tau$ to deduce that $\|\mathbf{D}_{1:N}^{-1}\|_\infty \geq 1 + \tau_N$. Hence, $1 + \tau_N \leq \|\mathbf{D}_{1:N}^{-1}\|_\infty \leq 2$. Since $\tau_N = 1$ for the Radau points, it follows that $\|\mathbf{D}_{1:N}^{-1}\|_\infty = 2$. For the Gauss points, τ_N approaches 1 as N tends to infinity; consequently, $\|\mathbf{D}_{1:N}^{-1}\|_\infty$ approaches 2 as N tends to infinity for the Gauss points. □

References

1. Babuška, I., Suri, M.: The h-p version of the finite element method with quasiuniform meshes, RAIRO. Modélisation Mathématique et Analyse Numérique **21**, 199–238 (1987)
2. Babuška, I., Suri, M.: The p- and h-p version of the finite element method, an overview. Comput. Methods Appl. Mech. Eng. **80**, 5–26 (1990)
3. Babuška, I., Suri, M.: The p and h-p version of the finite element method, basic principles and properties. SIAM Rev. **36**, 578–632 (1994)
4. 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**, 1435–1440 (2006)
5. Bernardi, C., Maday, Y.: Polynomial interpolation results in Sobolev spaces. J. Comput. Appl. Math. **43**, 53–82 (1992)

6. Betts, J.T.: Sparse optimization suite. In: Applied Mathematical Analysis. LLC, Issaquah (2013)
7. Biegler, L.T., Zavala, V.M.: Large-scale nonlinear programming using IPOPT: an integrating framework for enterprise-wide optimization. Comput. Chem. Eng. **33**, 575–582 (2008)
8. Bonnans, J.F., Laurent-Varin, J.: Computation of order conditions for symplectic partitioned Runge–Kutta schemes with application to optimal control. Numer. Math. **103**, 1–10 (2006)
9. Canuto, C., Hussaini, M., Quarteroni, A., Zang, T.: Spectral Methods, Fundamentals in Single Domains. Springer, Berlin (2006)
10. Chen, W., Du, W., Hager, W.W., Yang, L.: Bounds for integration matrices that arise in Gauss and Radau collocation. Comput. Optim. Appl. (2019). https://doi.org/10.1007/s10589-019-00099-5
11. Darby, C.L., Hager, W.W., Rao, A.V.: Direct trajectory optimization using a variable low-order adaptive pseudospectral method. J. Spacecr. Rockets **48**, 433–445 (2011)
12. Darby, C.L., Hager, W.W., Rao, A.V.: An hp-adaptive pseudospectral method for solving optimal control problems. Optim. Control Appl. Methods **32**, 476–502 (2011)
13. Dennis, M.E., Hager, W.W., Rao, A.V.: Computational method for optimal guidance and control using adaptive Gaussian quadrature collocation. J. Guid. Control Dyn. (2019). https://doi.org/10.2514/1.G003943
14. Dontchev, A., Hager, W.W., Poore, A., Yang, B.: Optimality, stability, and convergence in nonlinear control. Appl. Math. Optim. **31**, 297–326 (1995)
15. Dontchev, A.L., Hager, W.W.: Lipschitzian stability in nonlinear control and optimization. SIAM J. Control Optim. **31**, 569–603 (1993)
16. Dontchev, A.L., Hager, W.W.: A new approach to Lipschitz continuity in state constrained optimal control. Syst. Control Lett. **35**, 137–143 (1998)
17. Dontchev, A.L., Hager, W.W.: The Euler approximation in state constrained optimal control. Math. Comput. **70**, 173–203 (2000)
18. Dontchev, A.L., Hager, W.W., Veliov, V.M.: Second-order Runge–Kutta approximations in constrained optimal control. SIAM J. Numer. Anal. **38**, 202–226 (2000)
19. Elnagar, G., Kazemi, M., Razzaghi, M.: The pseudospectral Legendre method for discretizing optimal control problems. IEEE Trans. Automat. Control **40**, 1793–1796 (1995)
20. Elnagar, G.N., Kazemi, M.A.: Pseudospectral Chebyshev optimal control of constrained nonlinear dynamical systems. Comput. Optim. Appl. **11**, 195–217 (1998)
21. Elschner, J.: The h-p-version of spline approximation methods for Melin convolution equations. J. Integral Equ. Appl. **5**, 47–73 (1993)
22. Fahroo, F., Ross, I.M.: Costate estimation by a Legendre pseudospectral method. J. Guid. Control Dyn. **24**, 270–277 (2001)
23. Fahroo, F., Ross, I.M.: Direct trajectory optimization by a Chebyshev pseudospectral method. J. Guid. Control Dyn. **25**, 160–166 (2002)
24. Garg, D., Patterson, M.A., Darby, C.L., Françolin, 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 using a Radau pseudospectral method. Comput. Optim. Appl. **49**, 335–358 (2011)
25. 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**, 1843–1851 (2010)
26. Gong, Q., Ross, I.M., Kang, W., Fahroo, F.: Connections between the covector mapping theorem and convergence of pseudospectral methods for optimal control. Comput. Optim. Appl. **41**, 307–335 (2008)
27. Gui, W., Babuška, I.: The h, p and h-p versions of the finite element method in 1 dimension. Part I. The error analysis of the p-version. Numer. Math. **49**, 577–612 (1986)
28. Gui, W., Babuška, I.: The h, p and h-p versions of the finite element method in 1 dimension. Part II. The error analysis of the h-and h-p versions. Numer. Math. **49**, 613–657 (1986)
29. Gui, W., Babuška, I.: The h, p and h-p versions of the finite element method in 1 dimension. Part III. The adaptive h-p version. Numer. Math. **49**, 659–683 (1986)
30. Hager, W.W.: Multiplier methods for nonlinear optimal control. SIAM J. Numer. Anal. **27**, 1061–1080 (1990)
31. Hager, W.W.: Runge–Kutta methods in optimal control and the transformed adjoint system. Numer. Math. **87**, 247–282 (2000)
32. Hager, W.W.: Numerical analysis in optimal control. In: Hoffmann, K.-H., Lasiecka, I., Leugering, G., Sprekels, J., Tröltzsch, F. (eds.) International Series of Numerical Mathematics, vol. 139, pp. 83–93. Birkhauser Verlag, Basel (2001)

33. Hager, W.W., Hou, H., Rao, A.V.: Convergence rate for a Radau collocation method applied to unconstrained optimal control (2015). arXiv:1508.03783
34. 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**, 801–824 (2016)
35. Hager, W.W., Ianculescu, G.: Dual approximations in optimal control. SIAM J. Control Optim. **22**, 423–465 (1984)
36. 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)
37. Hager, W.W., Liu, J., Mohapatra, S., Rao, A.V., Wang, X.-S.: A pseudospectral method for optimal control based on collocation at the gauss points. In: 2018 IEEE Conference on Decision and Control (CDC), pp. 2490–2495 (Dec 2018)
38. Horn, R.A., Johnson, C.R.: Matrix Analysis. Cambridge University Press, Cambridge (2013)
39. Kang, W.: The rate of convergence for a pseudospectral optimal control method. In: Proceeding of the 47th IEEE Conference on Decision and Control, pp. 521–527. IEEE (2008)
40. Kang, W.: Rate of convergence for the Legendre pseudospectral optimal control of feedback linearizable systems. J. Control Theory Appl. **8**, 391–405 (2010)
41. Liu, F., Hager, W.W., Rao, A.V.: Adaptive mesh refinement method for optimal control using nonsmoothness detection and mesh size reduction. J. Frankl. Inst. **352**, 4081–4106 (2015)
42. Nocedal, J., Wright, S.J.: Numerical Optimization, 2nd edn. Springer, New York (2006)
43. Patterson, M.A., Hager, W.W., Rao, A.V.: A *ph* mesh refinement method for optimal control. Optim. Control Appl. Meth. **36**, 398–421 (2015)
44. Patterson, M.A., Rao, A.V.: GPOPS-II: A MATLAB software for solving multi-phase optimal control problems using hp-adaptive Gaussian quadrature collocation methods and sparse non-linear programming. ACM Trans. Math. Softw. **41**, 1–37 (2014)
45. Patterson, M.A., Rao, A.V.: $\mathbb{GPOPS} - \mathbb{II}$, a MATLAB software for solving multiple-phase optimal control problems using *hp*-adaptive Gaussian quadrature collocation methods and sparse nonlinear programming. ACM Trans. Math. Softw. **41**, 1–37 (2015)
46. Shen, J., Tang, T., Wang, L.-L.: Spectral Methods. Springer, Berlin (2011)
47. Williams, P.: Jacobi pseudospectral method for solving optimal control problems. J. Guid. Control Dyn. **27**, 293–297 (2004)

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

Affiliations

William W. Hager[1] · Hongyan Hou[2] · Subhashree Mohapatra[3,4] · Anil V. Rao[5] · Xiang-Sheng Wang[6]

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

Hongyan Hou
hongyan.hou@mnstate.edu

Subhashree Mohapatra
subhashree3mohapatra@gmail.com

Anil V. Rao
anilvrao@ufl.edu
http://www.mae.ufl.edu/rao

Xiang-Sheng Wang
xswang@louisiana.edu
http://www.ucs.louisiana.edu/~xxw6637/

[1] Department of Mathematics, University of Florida, P.O. Box 118105, Gainesville, FL 32611-8105, USA

[2] Mathematics Department, Minnesota State University Moorhead, 1104 7th Avenue South, P.O. Box 104, Moorhead, MN 56563, USA

[3] Department of Mathematics, University of Florida, Gainesville, FL 32611, USA

[4] Present Address: Department of Mathematics, SRM Institute of Science and Technology, Kattankulathur 603203, India

[5] Department of Mechanical and Aerospace Engineering, University of Florida, P.O. Box 116250, Gainesville, FL 32611-6250, USA

[6] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70503, USA