

CONVERGENCE RATE FOR A GAUSS COLLOCATION METHOD APPLIED TO CONSTRAINED OPTIMAL CONTROL*

WILLIAM W. HAGER[†], JUN LIU[‡], SUBHASHREE MOHAPATRA[†], ANIL V. RAO[§], AND
XIANG-SHENG WANG[¶]

Abstract. A local convergence rate is established for a Gauss orthogonal collocation method applied to optimal control problems with control constraints. If the Hamiltonian possesses a strong convexity property, then the theory yields convergence for problems whose optimal state and costate possess two square integrable derivatives. The convergence theory is based on a stability result for the sup-norm change in the solution of a variational inequality relative to a 2-norm perturbation, and on a Sobolev space bound for the error in interpolation at the Gauss quadrature points and the additional point -1 . The tightness of the convergence theory is examined using a numerical example.

Key words. Gauss collocation method, convergence rate, optimal control, orthogonal collocation

AMS subject classifications. 49M25, 49M37, 65K05, 90C30

DOI. 10.1137/16M1096761

1. Introduction. In earlier work [24, 25, 26], we analyze the convergence rate for orthogonal collocation methods applied to unconstrained control problems. In this analysis, it is assumed that the problem solution is smooth, in which case the theory implies that the discrete approximations converge to the solution of the continuous problem at potentially an exponential rate. But when control constraints are present, the solution often possesses limited regularity. The convergence theory developed in the earlier work for unconstrained problems required that the optimal state had at least four derivatives, while for constrained problems, the optimal state may have only two derivatives, at best [4, 7, 20, 28]. The earlier convergence theory was based on a stability analysis for a linearization of the unconstrained control problem; the theory showed that the sup-norm change in the solution was bounded relative to the sup-norm perturbation in the linear system. Here we introduce a convex control constraint, in which case the linearized problem is a variational inequality, or equivalently a differential inclusion, not a linear system. We obtain a bound for the sup-norm change in the solution relative to a 2-norm perturbation in the variational inequality. By using the 2-norm for the perturbation rather than the sup-norm, we are able to avoid both Lebesgue constants and the Markov bound [34] for the sup-norm of the derivative of a polynomial relative to the sup-norm of the original polynomial. Using best approximation results in Sobolev spaces [3, 13], we obtain convergence when the

*Received by the editors October 3, 2016; accepted for publication (in revised form) January 8, 2018; published electronically April 10, 2018.

<http://www.siam.org/journals/sicon/56-2/M109676.html>

Funding: The authors gratefully acknowledge support by the Office of Naval Research under grant N00014-15-1-2048, by the National Science Foundation under grant DMS-1522629, and by the U.S. Air Force Research Laboratory under contract FA8651-08-D-0108/0054.

[†]Department of Mathematics, University of Florida, Gainesville, FL 32611-8105 (hager@ufl.edu, <http://people.clas.ufl.edu/hager/>, subhashree3mohapatra@gmail.com).

[‡]Department of Mathematics and Statistics, Southern Illinois University Edwardsville, Edwardsville, IL 62026 (juliu@siue.edu, <http://www.siue.edu/~juliu/>).

[§]Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611-6250 (anilvrao@ufl.edu, <http://www.mae.ufl.edu/people/rao>).

[¶]Department of Mathematics, University of Louisiana at Lafayette, Lafayette, LA 70503 (xswang@louisiana.edu, <http://www.ucslouisiana.edu/~xxw6637/>).

optimal state and costate have only two square integrable derivatives, which implies that the theory is applicable to a class of control constrained problems for which the optimal control is Lipschitz continuous.

The specific collocation scheme analyzed in this paper, presented in [2, 18], is based on collocation at the Gauss quadrature points, or equivalently, at the roots of a Legendre polynomial. Other sets of collocation points that have been studied in the literature include the Lobatto quadrature points [11, 14, 19], the Chebyshev quadrature points [12, 15], the Radau quadrature points [16, 17, 33, 36], and extrema of Jacobi polynomials [39]. Kang [31, 32] obtains a convergence rate for the Lobatto scheme applied to control systems in feedback linearizable normal form by inserting bounds in the discrete problem for the states, the controls, and certain Legendre polynomial expansion coefficients. In our approach, the discretized problem is obtained by simply collocating at the Gauss quadrature points.

Our approximation to the control problem uses a global polynomial defined on the problem domain. Earlier work, including [6, 8, 9, 10, 22, 30, 37], utilizes a piecewise polynomial approximation, in which case convergence is achieved by letting the mesh spacing approach zero, while keeping the polynomial degree fixed. For an orthogonal collocation scheme based on global polynomials, convergence is achieved by letting the degree of the polynomials tend to infinity. Our results show that even when control constraints are present, and a solution possesses limited regularity, convergence can still be achieved with global polynomials.

We consider control problems of the form

$$(1.1) \quad \begin{aligned} & \text{minimize} && C(\mathbf{x}(1)) \\ & \text{subject to} && \dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{u}(t)), \quad \mathbf{u}(t) \in \mathcal{U}, \quad t \in \Omega, \\ & && \mathbf{x}(-1) = \mathbf{x}_0, \quad (\mathbf{x}, \mathbf{u}) \in \mathcal{C}^1(\Omega; \mathbb{R}^n) \times \mathcal{C}^0(\Omega; \mathbb{R}^m), \end{aligned}$$

where $\Omega = [-1, 1]$, 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{x}_0 is the initial condition which we assume is given, $\mathbf{f} : \mathbb{R}^n \times \mathbb{R}^m \rightarrow \mathbb{R}^n$, and $C : \mathbb{R}^n \rightarrow \mathbb{R}$, $\mathcal{C}^l(\Omega; \mathbb{R}^n)$ denotes the space of l times continuously differentiable functions mapping Ω to \mathbb{R}^n . It is assumed that \mathbf{f} and C are at least continuous.

Let \mathcal{P}_N denote the space of polynomials of degree at most N , and let \mathcal{P}_N^n denote the n -fold Cartesian product $\mathcal{P}_N \times \cdots \times \mathcal{P}_N$. We analyze the discretization of (1.1) given by

$$(1.2) \quad \begin{aligned} & \text{minimize} && C(\mathbf{x}(1)) \\ & \text{subject to} && \dot{\mathbf{x}}(\tau_i) = \mathbf{f}(\mathbf{x}(\tau_i), \mathbf{u}_i), \quad \mathbf{u}_i \in \mathcal{U}, \quad 1 \leq i \leq N, \\ & && \mathbf{x}(-1) = \mathbf{x}_0, \quad \mathbf{x} \in \mathcal{P}_N^n. \end{aligned}$$

The polynomials used to approximate the state should satisfy the dynamics exactly at the collocation points τ_i , $1 \leq i \leq N$. The parameter \mathbf{u}_i represents an approximation to the control at time τ_i . The dimension of \mathcal{P}_N is $N + 1$, while there are $N + 1$ equations in (1.2) corresponding to the collocated dynamics at N points and the initial condition. We collocate at the Gauss quadrature points, which are symmetric about $t = 0$ and which satisfy

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

The analysis also makes use of the two noncollocated points

$$\tau_0 = -1 \quad \text{and} \quad \tau_{N+1} = +1.$$

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

$$\|\mathbf{x}\|_\infty = \sup\{|\mathbf{x}(t)| : t \in [0, 1]\},$$

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

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

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 $\mathcal{C}^1(\Omega; \mathbb{R}^n) \times \mathcal{C}^0(\Omega; \mathbb{R}^m)$. There exists an open set $\mathcal{O} \subset \mathbb{R}^{m+n}$ and $\rho > 0$ such that

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

Moreover, the first two derivatives 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

$$(1.3) \quad \dot{\boldsymbol{\lambda}}^*(t) = -\nabla_x H(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t)), \quad \boldsymbol{\lambda}^*(1) = \nabla C(\mathbf{x}^*(1)),$$

where H is the Hamiltonian defined by $H(\mathbf{x}, \mathbf{u}, \boldsymbol{\lambda}) = \boldsymbol{\lambda}^\top \mathbf{f}(\mathbf{x}, \mathbf{u})$ and ∇ denotes gradient. From the first-order optimality conditions (Pontryagin’s minimum principle), it follows that

$$(1.4) \quad -\nabla_u H(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t)) \in N_{\mathcal{U}}(\mathbf{u}^*(t)) \quad \text{for all } t \in \Omega,$$

where $N_{\mathcal{U}}$ is the normal cone. For any $\mathbf{u} \in \mathcal{U}$,

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

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

Since the collocation problem (1.2) is finite dimensional, the first-order optimality conditions, or Karush–Kuhn–Tucker conditions, hold when a constraint qualification [35] is satisfied. We show in Lemma 2.1 that the first-order optimality conditions are equivalent to the existence of $\boldsymbol{\lambda} \in \mathcal{P}_N^n$ such that

$$(1.5) \quad \dot{\boldsymbol{\lambda}}(\tau_i) = -\nabla_x H(\mathbf{x}(\tau_i), \mathbf{u}_i, \boldsymbol{\lambda}(\tau_i)), \quad 1 \leq i \leq N,$$

$$(1.6) \quad \boldsymbol{\lambda}(1) = \nabla C(\mathbf{x}(1)),$$

$$(1.7) \quad N_{\mathcal{U}}(\mathbf{u}_i) \ni -\nabla_u H(\mathbf{x}(\tau_i), \mathbf{u}_i, \boldsymbol{\lambda}(\tau_i)), \quad 1 \leq i \leq N.$$

The following assumptions are utilized in the convergence analysis.

- (A1) For some $\alpha > 0$, the smallest eigenvalue of the Hessian matrices $\nabla^2 C(\mathbf{x}^*(1))$ and $\nabla_{(x,u)}^2 H(\mathbf{x}^*(t), \mathbf{u}^*(t), \boldsymbol{\lambda}^*(t))$ are greater than α , uniformly for $t \in [0, 1]$.
- (A2) For some $\beta < 1/2$, the Jacobian of the dynamics satisfies

$$\|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))\|_\infty \leq \beta \quad \text{and} \quad \|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))^\top\|_\infty \leq \beta$$

for all $t \in \Omega$, where $\|\cdot\|_\infty$ is the matrix sup-norm (largest absolute row sum), and the Jacobian $\nabla_x \mathbf{f}$ is an n by n matrix whose i th row is $(\nabla_x f_i)^\top$.

The condition (A2) ensures (see Lemma 5.1) that in the discrete linearized problem, it is possible to solve for the discrete state in terms of the discrete control. As

shown in [24], this property holds in an hp -collocation framework when the domain Ω is partitioned into K mesh intervals with K large enough that

$$\|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))\|_\infty / K \leq \beta \quad \text{and} \quad \|\nabla_x \mathbf{f}(\mathbf{x}^*(t), \mathbf{u}^*(t))^\top\|_\infty / K \leq \beta$$

for all $t \in \Omega$.

The coercivity assumption (A1) is not only a sufficient condition for the local optimality of a feasible point $(\mathbf{x}^*, \mathbf{u}^*)$ of (1.1), but it yields the stability of the discrete linearized problem (see Lemma 6.2). One would hope that (A1) could be weakened to only require coercivity relative to a subspace associated with the linearized dynamics similar to what is done in [6]. To formulate this weakened condition, we introduce the following 6 matrices:

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

With this notation and with $\langle \cdot, \cdot \rangle$ denoting the L^2 inner product, the weaker version of (A1) is that

$$\mathbf{x}(1)^\top \mathbf{T} \mathbf{x}(1) + \langle \mathbf{x}, \mathbf{Q} \mathbf{x} \rangle + \langle \mathbf{u}, \mathbf{R} \mathbf{u} \rangle + 2 \langle \mathbf{x}, \mathbf{S} \mathbf{u} \rangle \geq \alpha \langle \mathbf{u}, \mathbf{u} \rangle,$$

whenever (\mathbf{x}, \mathbf{u}) satisfies $\dot{\mathbf{x}} = \mathbf{A} \mathbf{x} + \mathbf{B} \mathbf{u}$ with $\mathbf{x}(-1) = \mathbf{0}$ and $\mathbf{u} = \mathbf{v} - \mathbf{w}$ for some \mathbf{v} and $\mathbf{w} \in L^2$ satisfying $\mathbf{v}(t)$ and $\mathbf{w}(t) \in \mathcal{U}$ for almost every $t \in [-1, 1]$. For the Euler integration scheme, we show in [6, Lem. 11] that this weaker condition implies an analogous coercivity property for the discrete problem. The extension of this result from the Euler scheme to orthogonal collocation schemes remains an open problem.

Let \mathbf{D} be the N by $N + 1$ matrix defined by

$$(1.8) \quad D_{ij} = \dot{L}_j(\tau_i), \text{ where } L_j(\tau) := \prod_{\substack{l=0 \\ l \neq j}}^N \frac{\tau - \tau_l}{\tau_j - \tau_l}, \quad 1 \leq i \leq N \text{ and } 0 \leq j \leq N.$$

The matrix \mathbf{D} is a differentiation matrix in the sense that $(\mathbf{D} \mathbf{p})_i = \dot{p}(\tau_i)$, $1 \leq i \leq N$, whenever $p \in \mathcal{P}_N$ is the polynomial that satisfies $p(\tau_j) = p_j$ for $0 \leq j \leq N$. The submatrix $\mathbf{D}_{1:N}$, consisting of the trailing N columns of \mathbf{D} , has the following properties which are utilized in the analysis:

- (P1) $\mathbf{D}_{1:N}$ is invertible and $\|\mathbf{D}_{1:N}^{-1}\|_\infty \leq 2$.
- (P2) If \mathbf{W} is the diagonal matrix containing the Gauss quadrature weights ω_i , $1 \leq i \leq N$, on the diagonal, then the rows of the matrix $[\mathbf{W}^{1/2} \mathbf{D}_{1:N}]^{-1}$ have Euclidean norm bounded by $\sqrt{2}$.

The invertibility of $\mathbf{D}_{1:N}$ is proved in [18, Prop. 1]. The bound for the inverse appearing in (P1) is established in Appendix 1. (P2) has been checked numerically for N up to 300 in [26]. Some intuition concerning the general validity of (P2) is as follows: It is observed numerically that the last row of the matrix $[\mathbf{W}^{1/2} \mathbf{D}_{1:N}]^{-1}$ has the largest Euclidean norm among all the rows. Based on the formula for $\mathbf{D}_{1:N}^{-1}$ given in [18, sect. 4.1.2], the i th element in the last row approaches ω_i as N tends to infinity. Hence, the i th element in the last row of $[\mathbf{W}^{1/2} \mathbf{D}_{1:N}]^{-1}$ approaches $\sqrt{\omega_i}$ as N tends to infinity. Since the quadrature weights sum to 2, the Euclidean norm of the last row of $[\mathbf{W}^{1/2} \mathbf{D}_{1:N}]^{-1}$ should be close to $\sqrt{2}$. Despite the strong numerical evidence for (P2), a proof of (P2) for general N is still missing.

The properties (P1) and (P2) are stated separately since they are used in different ways in the analysis. However, (P2) implies (P1) by the Schwarz inequality. That is, if \mathbf{r} is a row from $\mathbf{D}_{1:N}^{-1}$, then we have

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

since the quadrature weights sum to 2 and when (P2) holds, the Euclidean norm of a row from $[\mathbf{W}^{1/2}\mathbf{D}_{1:N}]^{-1}$ is at most $\sqrt{2}$.

If $\mathbf{x}^N \in \mathcal{P}_N^n$ is a solution of (1.2) associated with the discrete controls $\mathbf{u}_i, 1 \leq i \leq N$, and if $\boldsymbol{\lambda}^N \in \mathcal{P}_N^n$ satisfies (1.5)–(1.7), then we define

$$\begin{aligned} \mathbf{X}^N &= [\mathbf{x}^N(-1), \mathbf{x}^N(\tau_1), \dots, \mathbf{x}^N(\tau_N), \mathbf{x}^N(+1)], \\ \mathbf{X}^* &= [\mathbf{x}^*(-1), \mathbf{x}^*(\tau_1), \dots, \mathbf{x}^*(\tau_N), \mathbf{x}^*(+1)], \\ \mathbf{U}^N &= [\mathbf{u}_1, \dots, \mathbf{u}_N], \\ \mathbf{U}^* &= [\mathbf{u}^*(\tau_1), \dots, \mathbf{u}^*(\tau_N)], \\ \boldsymbol{\Lambda}^N &= [\boldsymbol{\lambda}^N(-1), \boldsymbol{\lambda}^N(\tau_1), \dots, \boldsymbol{\lambda}^N(\tau_N), \boldsymbol{\lambda}^N(+1)], \\ \boldsymbol{\Lambda}^* &= [\boldsymbol{\lambda}^*(-1), \boldsymbol{\lambda}^*(\tau_1), \dots, \boldsymbol{\lambda}^*(\tau_N), \boldsymbol{\lambda}^*(+1)]. \end{aligned}$$

The following convergence result relative to the vector ∞ -norm (largest absolute element) is established. Here $\mathcal{H}^p(\Omega; \mathbb{R}^n)$ denotes the Sobolev space of functions with square integrable derivatives through order p and norm denoted $\|\cdot\|_{\mathcal{H}^p(\Omega; \mathbb{R}^n)}$.

THEOREM 1.1. *Suppose $(\mathbf{x}^*, \mathbf{u}^*)$ is a local minimizer for the continuous problem (1.1) with $(\mathbf{x}^*, \boldsymbol{\lambda}^*) \in \mathcal{H}^\eta(\Omega; \mathbb{R}^n)$ for some $\eta \geq 2$. If both (A1)–(A2) and (P1)–(P2) hold, then for N sufficiently large, the discrete problem (1.2) has a local minimizer $\mathbf{x}^N \in \mathcal{P}_N^n$ and $\mathbf{u} \in \mathbb{R}^{mN}$, and an associated multiplier $\boldsymbol{\lambda}^N \in \mathcal{P}_N^n$ satisfying (1.5)–(1.7); moreover, there exists a constant c independent of N and η such that*

$$(1.9) \quad \begin{aligned} &\max \{ \|\mathbf{X}^N - \mathbf{X}^*\|_\infty, \|\mathbf{U}^N - \mathbf{U}^*\|_\infty, \|\boldsymbol{\Lambda}^N - \boldsymbol{\Lambda}^*\|_\infty \} \\ &\leq \left(\frac{c}{N} \right)^{p-3/2} (\|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega; \mathbb{R}^n)} + \|\boldsymbol{\lambda}^*\|_{\mathcal{H}^p(\Omega; \mathbb{R}^n)}), \quad p := \min\{\eta, N + 1\}. \end{aligned}$$

This result was established in [26] for an unconstrained control problem, but with the exponent 3/2 replaced by 3 and with $\eta \geq 4$. Hence, the analysis is extended to control constrained problems and the exponent of N in the convergence estimate is improved by 1.5. Since typical control constrained problems have regularity at most $\eta = 2$ when (A1) holds, there is no guarantee of convergence with the previous estimate.

The paper is organized as follows. In section 2 the discrete optimization problem (1.2) is reformulated as a differential inclusion obtained from the first-order optimality conditions, and a general approach to convergence analysis is presented. We also establish the connection between the Karush–Kuhn–Tucker conditions and the polynomial conditions (1.5)–(1.7). In section 3 we use results from [3] to bound the derivative of the interpolation error in \mathcal{L}^2 . Section 4 estimates how closely the solution to the continuous problem satisfies the first-order optimality conditions for the discrete problem, while section 5 establishes the invertibility of the linearized dynamics for the discrete problem. Section 6 proves a Lipschitz property for the linearized optimality conditions, which yields a proof of Theorem 1.1. A numerical example given in section 7 indicates the potential for further improvements to the convergence rate exponent. Section 10 contains a result of Maday concerning the error in best \mathcal{H}^1 approximation relative to an \mathcal{L}^2 norm with a singular weight function.

Notation. We let \mathcal{P}_N denote the space of polynomials of degree at most N , while \mathcal{P}_N^0 is the subspace consisting of polynomials in \mathcal{P}_N that vanish at $t = -1$ and $t = 1$. The Gauss collocation points $\tau_i, 1 \leq i \leq N$, are the roots of the Legendre polynomial P_N of degree N . The associated Gauss quadrature weights $\omega_i, 1 \leq i \leq N$, are given by

$$(1.10) \quad \omega_i = \frac{2}{(1 - \tau_i^2)P'_N(\tau_i)^2}.$$

For any $p \in \mathcal{P}_{2N-1}$, we have [38, Thm. 3.6.24]

$$(1.11) \quad \int_{\Omega} p(t) dt = \sum_{i=1}^N \omega_i p(\tau_i).$$

Derivatives with respect to t are denoted with either a dot above the function as in $\dot{\mathbf{x}}$, which is common in the optimal control literature, or with an accent as in p' , which is common in the numerical analysis literature. The meaning of the norm $\|\cdot\|_{\infty}$ is based on context. If $\mathbf{x} \in \mathcal{C}^0(\mathbb{R}^n)$, 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 $1 \leq i \leq m$. 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 vector sup-norm). We often partition a vector $\mathbf{p} \in \mathbb{R}^{nN}$ into subvectors $\mathbf{p}_i \in \mathbb{R}^n, 1 \leq i \leq N$. Similarly, if $\mathbf{p} \in \mathbb{R}^{mN}$, then $\mathbf{p}_i \in \mathbb{R}^m$. 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 is independent of the polynomial degree N and the smoothness η , and which may have different values in different equations. 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 the differentiation matrix introduced in (1.8), 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 matrix composed of $p \times q$ blocks; the (i, j) block is $u_{ij}\mathbf{V}$. We let $\mathcal{L}^2(\Omega)$ denote the usual space of functions square integrable on Ω , while $\mathcal{H}^p(\Omega)$ is the Sobolev space consisting of functions with square integrable derivatives through order p . The norm in $\mathcal{H}^p(\Omega)$ is denoted $\|\cdot\|_{\mathcal{H}^p(\Omega)}$. The seminorm in $\mathcal{H}^1(\Omega)$ corresponding to the $\mathcal{L}^2(\Omega)$ norm of the derivative is denoted $|\cdot|_{\mathcal{H}^1(\Omega)}$. The subspace of $\mathcal{H}^1(\Omega)$ corresponding to functions that vanish at $t = -1$ and $t = 1$ is denoted $\mathcal{H}_0^1(\Omega)$. We let $\mathcal{H}^p(\Omega; \mathbb{R}^n)$ denote the n -fold Cartesian product $\mathcal{H}^p(\Omega) \times \dots \times \mathcal{H}^p(\Omega)$.

2. Abstract setting. In the introduction, we formulated the discrete optimization problem (1.2) and the necessary conditions (1.5)–(1.7) in polynomial spaces. However, to prove Theorem 1.1, we reformulate the first-order optimality conditions in Cartesian space. Given a feasible point $\mathbf{x} \in \mathcal{P}_N^n$ and $\mathbf{u} \in \mathbb{R}^{mN}$ for the discrete problem (1.2), define $\mathbf{X}_j = \mathbf{x}(\tau_j), 0 \leq j \leq N + 1$, and $\mathbf{U}_i = \mathbf{u}_i, 1 \leq i \leq N$. As noted earlier, \mathbf{D} is a differentiation matrix in the sense that

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

Since $\dot{\mathbf{x}} \in \mathcal{P}_{N-1}^n$, it follows from the exactness result (1.11) for Gaussian quadrature that when \mathbf{x} satisfies the dynamics of (1.2), we have

$$\mathbf{X}_{N+1} = \mathbf{x}(1) = \mathbf{x}(-1) + \int_{\Omega} \dot{\mathbf{x}}(t) dt = \mathbf{X}_0 + \sum_{j=1}^N \omega_j \mathbf{f}(\mathbf{X}_j, \mathbf{U}_j).$$

Hence, the discrete problem (1.2) can be reformulated as the nonlinear programming problem

$$\begin{aligned} & \text{minimize} && C(\mathbf{X}_{N+1}) \\ (2.1) \quad & \text{subject to} && \sum_{j=0}^N D_{ij} \mathbf{X}_j = \mathbf{f}(\mathbf{X}_i, \mathbf{U}_i), \quad \mathbf{U}_i \in \mathcal{U}, \quad 1 \leq i \leq N, \\ & && \mathbf{X}_0 = \mathbf{x}_0, \quad \mathbf{X}_{N+1} = \mathbf{X}_0 + \sum_{j=1}^N \omega_j \mathbf{f}(\mathbf{X}_j, \mathbf{U}_j). \end{aligned}$$

To prove Theorem 1.1, we analyze the existence and stability of solutions to the first-order optimality conditions associated with the nonlinear programming problem.

We introduce multipliers $\boldsymbol{\mu}_j \in \mathbb{R}^n$, $0 \leq j \leq N + 1$, corresponding to each of the constraints in the nonlinear program. The first-order optimality conditions correspond to stationary points of the Lagrangian

$$\begin{aligned} & C(\mathbf{X}_{N+1}) + \sum_{i=1}^N \left\langle \boldsymbol{\mu}_i, \mathbf{f}(\mathbf{X}_i, \mathbf{U}_i) - \sum_{j=0}^N D_{ij} \mathbf{X}_j \right\rangle + \langle \boldsymbol{\mu}_0, \mathbf{x}_0 - \mathbf{X}_0 \rangle \\ & + \left\langle \boldsymbol{\mu}_{N+1}, \mathbf{X}_0 - \mathbf{X}_{N+1} + \sum_{i=1}^N \omega_i \mathbf{f}(\mathbf{X}_i, \mathbf{U}_i) \right\rangle. \end{aligned}$$

The stationarity conditions for the Lagrangian appear below.

$$(2.2) \quad \mathbf{X}_0 \Rightarrow \boldsymbol{\mu}_{N+1} = \boldsymbol{\mu}_0 + \sum_{i=1}^N D_{i0} \boldsymbol{\mu}_i,$$

$$(2.3) \quad \mathbf{X}_j \Rightarrow \sum_{i=1}^N D_{ij} \boldsymbol{\mu}_i = \nabla_x H(\mathbf{X}_j, \mathbf{U}_j, \boldsymbol{\mu}_j + \omega_j \boldsymbol{\mu}_{N+1}), \quad 1 \leq j \leq N,$$

$$(2.4) \quad \mathbf{X}_{N+1} \Rightarrow \boldsymbol{\mu}_{N+1} = \nabla C(\mathbf{X}_{N+1}),$$

$$(2.5) \quad \mathbf{U}_i \Rightarrow -\nabla_u H(\mathbf{X}_i, \mathbf{U}_i, \boldsymbol{\mu}_i + \omega_i \boldsymbol{\mu}_{N+1}) \in N_{\mathcal{U}}(\mathbf{U}_i), \quad 1 \leq i \leq N.$$

Since there are no state constraints, the conditions (2.2)–(2.4) are obtained by setting to zero the derivative of the Lagrangian with respect to the indicated variables. The condition (2.5) corresponds to stationarity of the Lagrangian respect to the control. The relation between multipliers satisfying (2.2)–(2.5) and $\boldsymbol{\lambda} \in \mathcal{P}_N^n$ satisfying (1.5)–(1.7) is as follows.

PROPOSITION 2.1. *The multipliers $\boldsymbol{\mu} \in \mathbb{R}^{n(N+2)}$ satisfy (2.2)–(2.5) if and only if the polynomial $\boldsymbol{\lambda} \in \mathcal{P}_N^n$ satisfying the $N + 1$ interpolation conditions $\boldsymbol{\lambda}(1) = \boldsymbol{\mu}_{N+1}$ and $\boldsymbol{\lambda}(\tau_i) = \boldsymbol{\mu}_{N+1} + \boldsymbol{\mu}_i/\omega_i$, $1 \leq i \leq N$, is a solution of (1.5)–(1.7) and $\boldsymbol{\lambda}(-1) = \boldsymbol{\mu}_0$.*

Proof. We start with multipliers $\boldsymbol{\mu}$ satisfying (2.2)–(2.5) and show that $\boldsymbol{\lambda} \in \mathcal{P}_N^n$ satisfying the interpolation conditions $\boldsymbol{\lambda}(1) = \boldsymbol{\mu}_{N+1}$ and $\boldsymbol{\lambda}(\tau_i) = \boldsymbol{\mu}_{N+1} + \boldsymbol{\mu}_i/\omega_i$,

$1 \leq i \leq N$, is a solution of (1.5)–(1.7) with $\lambda(-1) = \mu_0$. The converse follows by reversing all the steps in the derivation. Define $\Lambda_i = \mu_{N+1} + \mu_i/\omega_i$ for $1 \leq i \leq N$, $\Lambda_{N+1} = \mu_{N+1}$, and $\Lambda_0 = \mu_0$. Hence, we have $\mu_i = \omega_i(\Lambda_i - \Lambda_{N+1})$ for $1 \leq i \leq N$. In (2.5) we divide by ω_i and substitute $\Lambda_i = \mu_{N+1} + \mu_i/\omega_i$. In (2.3) we divide by ω_j , and substitute $\Lambda_j = \mu_{N+1} + \mu_j/\omega_j$ and

$$(2.6) \quad D_{ij} = - \left(\frac{\omega_j}{\omega_i} \right) D_{ji}^\dagger, \quad D_{i,N+1}^\dagger = - \sum_{j=1}^N D_{ij}^\dagger, \quad 1 \leq i \leq N.$$

With these modifications, (2.3)–(2.5) become

$$(2.7) \quad \sum_{j=1}^{N+1} D_{ij}^\dagger \Lambda_j = -\nabla_x H(\mathbf{X}_i, \mathbf{U}_i, \Lambda_i),$$

$$(2.8) \quad \Lambda_{N+1} = \nabla C(\mathbf{X}_{N+1}),$$

$$(2.9) \quad N_{\mathcal{U}}(\mathbf{U}_i) \ni -\nabla_u H(\mathbf{X}_i, \mathbf{U}_i, \Lambda_i),$$

$1 \leq i \leq N$. In [18, Thm. 1] it is shown that if $\lambda \in \mathcal{P}_N^n$ is a polynomial that satisfies the conditions $\lambda(\tau_i) = \Lambda_i$ for $1 \leq i \leq N + 1$, then

$$(2.10) \quad \sum_{j=1}^{N+1} D_{ij}^\dagger \Lambda_j = \dot{\lambda}(\tau_i), \quad 1 \leq i \leq N.$$

This identity coupled with (2.7)–(2.9) implies that (1.5)–(1.7) hold.

Now let us consider the final term in (2.2). Since the polynomial that is identically equal to $\mathbf{1}$ has derivative $\mathbf{0}$ and since \mathbf{D} is a differentiation matrix, we have $\mathbf{D}\mathbf{1} = \mathbf{0}$, 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 final term in (2.2) can be written

$$(2.11) \quad \begin{aligned} \sum_{i=1}^N \mu_i D_{i0} &= - \sum_{i=1}^N \sum_{j=1}^N \mu_i D_{ij} = - \sum_{i=1}^N \sum_{j=1}^N \omega_j \left(\frac{\mu_i}{\omega_i} \right) \left(\frac{\omega_i D_{ij}}{\omega_j} \right) \\ &= \sum_{i=1}^N \sum_{j=1}^N \omega_i D_{ij}^\dagger (\Lambda_j - \Lambda_{N+1}) = \sum_{i=1}^N \sum_{j=1}^{N+1} \omega_i D_{ij}^\dagger \Lambda_j. \end{aligned}$$

Again, if $\lambda \in \mathcal{P}_N^n$ is the interpolating polynomial that satisfies $\lambda(\tau_i) = \Lambda_i$ for $1 \leq i \leq N + 1$, then by (2.10), (2.11), and the exactness of Gaussian quadrature for polynomials in \mathcal{P}_{N-1}^n , we have

$$(2.12) \quad \sum_{i=1}^N \mu_i D_{i0} = \sum_{i=1}^N \omega_i \dot{\lambda}(\tau_i) = \int_{\Omega} \dot{\lambda}(\tau) d\tau = \lambda(1) - \lambda(-1).$$

Since $\lambda(1) = \lambda_{N+1} = \mu_{N+1}$, we deduce from (2.2) and (2.12) that $\lambda(-1) = \mu_0$. \square

In the proof of Proposition 2.1, $\Lambda_0 = \mu_0$ and $\Lambda_{N+1} = \mu_{N+1}$. We combine (2.2), (2.7), and (2.11) to obtain

$$(2.13) \quad \Lambda_{N+1} = \Lambda_0 - \sum_{i=1}^N \omega_i \nabla_x H(\mathbf{X}_i, \mathbf{U}_i, \Lambda_i).$$

Based on Proposition 2.1, the optimality conditions (2.2)–(2.5) are equivalent to (1.5)–(1.7), which are equivalent to (2.7)–(2.9) and (2.13). This latter formulation, which we refer to as the transformed adjoint system in our earlier work [22], is most convenient for the subsequent analysis. This leads us to write the first-order optimality conditions for (1.2) as an inclusion $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$, where

$$(\mathcal{T}_0, \mathcal{T}_1, \dots, \mathcal{T}_6)(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathbb{R}^n \times \mathbb{R}^{nN} \times \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^{nN} \times \mathbb{R}^n \times \mathbb{R}^{mN}.$$

The 7 components of \mathcal{T} are defined as

$$\begin{aligned} \mathcal{T}_0(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= \mathbf{X}_0 - \mathbf{x}_0, \\ \mathcal{T}_{1i}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= \left(\sum_{j=0}^N D_{ij} \mathbf{X}_j \right) - \mathbf{f}(\mathbf{X}_i, \mathbf{U}_i), \quad 1 \leq i \leq N, \\ \mathcal{T}_2(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= \mathbf{X}_{N+1} - \mathbf{X}_0 - \sum_{j=1}^N \omega_j \mathbf{f}(\mathbf{X}_j, \mathbf{U}_j), \\ \mathcal{T}_3(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= \mathbf{\Lambda}_{N+1} - \mathbf{\Lambda}_0 + \sum_{i=1}^N \omega_i \nabla_x H(\mathbf{X}_i, \mathbf{U}_i, \mathbf{\Lambda}_i), \\ \mathcal{T}_{4i}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= \left(\sum_{j=1}^{N+1} D_{ij}^\dagger \mathbf{\Lambda}_j \right) + \nabla_x H(\mathbf{X}_i, \mathbf{U}_i, \mathbf{\Lambda}_i), \quad 1 \leq i \leq N, \\ \mathcal{T}_5(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= \mathbf{\Lambda}_{N+1} - \nabla C(\mathbf{X}_{N+1}), \\ \mathcal{T}_{6i}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) &= -\nabla_u H(\mathbf{X}_i, \mathbf{U}_i, \mathbf{\Lambda}_i), \quad 1 \leq i \leq N. \end{aligned}$$

The components of \mathcal{F} are given by

$$\mathcal{F}_0 = \mathcal{F}_1 = \dots = \mathcal{F}_5 = \mathbf{0}, \quad \text{while } \mathcal{F}_{6i}(\mathbf{U}) = N_{\mathcal{U}}(\mathbf{U}_i).$$

The first three components of the inclusion $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$ are the constraints of (2.1), the next three components describe the discrete costate dynamics, and the last component is the discrete version of the Pontryagin minimum principle. The proof of Theorem 1.1 is based on an existence and stability result for local solutions of the inclusion $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$. We will apply [10, Prop. 3.1], which is repeated below for convenience. Other results like this are contained in [8, Thm. 3.1], in [21, Thm. 1], in [22, Prop. 5.1], and in [23, Thm. 2.1].

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 $B_r(\boldsymbol{\theta}^*)$ 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) $\|\nabla \mathcal{T}(\boldsymbol{\theta}) - \nabla \mathcal{T}(\boldsymbol{\theta}^*)\| \leq \epsilon$ 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 $\|\boldsymbol{\delta}\| \leq (1 - \gamma\epsilon)r/\gamma$, 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*

$$(2.14) \quad \|\boldsymbol{\theta} - \boldsymbol{\theta}^*\| \leq \frac{\gamma}{1 - \gamma\epsilon} \|\boldsymbol{\delta}\|.$$

We apply Proposition 2.2 with $\theta^* = (\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ and $\theta = (\mathbf{X}^N, \mathbf{U}^N, \mathbf{\Lambda}^N)$, where the discrete variables were defined before Theorem 1.1. The key steps in the analysis are the estimation of the residual $\|\mathcal{T}(\theta^*)\|$, the proof that $\mathcal{F} - \nabla\mathcal{T}(\theta^*)$ is invertible, and the proof that $(\mathcal{F} - \nabla\mathcal{T}(\theta^*))^{-1}$ is Lipschitz continuous with respect to the norms in \mathcal{X} and \mathcal{Y} . In our context, we use the sup-norm for \mathcal{X} . In particular,

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

For this norm, the left side of (1.9) and the left side of (2.14) are the same. The norm on \mathcal{Y} enters into the estimation of both the distance from $\|\mathcal{T}(\theta^*)\|$ to $\mathcal{F}(\theta^*)$ ($\|\delta\|$ in (2.14)) and the Lipschitz constant γ for $(\mathcal{F} - \nabla\mathcal{T}(\theta^*))^{-1}$. In our context, we think of an element of \mathcal{Y} as a large vector with components $\mathbf{y}_l \in \mathbb{R}^n$ or \mathbb{R}^m . There are N components in \mathbb{R}^m associated with \mathcal{T}_6 , one component in \mathbb{R}^n associated with each of $\mathcal{T}_0, \mathcal{T}_2, \mathcal{T}_3$, and \mathcal{T}_5 , and N components in \mathbb{R}^n associated with \mathcal{T}_1 and \mathcal{T}_4 . Hence, \mathcal{Y} has dimension $mN + 4n + 2nN$ which matches the dimension of \mathcal{X} since $\dim(\mathbf{U}) = mN$, $\dim(\mathbf{X}) = (n + 2)N$, and $\dim(\mathbf{\Lambda}) = (n + 2)N$. For the norm of $\mathbf{y} \in \mathcal{Y}$, we take

$$\|\mathbf{y}\|_{\mathcal{Y}} = |\mathbf{y}_0| + |\mathbf{y}_2| + |\mathbf{y}_3| + |\mathbf{y}_5| + \|\mathbf{y}_6\|_\infty + \|\mathbf{y}_1\|_\omega + \|\mathbf{y}_4\|_\omega.$$

Here ω -norm used for \mathbf{y}_1 (state dynamics) and \mathbf{y}_4 (costate dynamics) is defined by

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

Note that the ω -norm has the upper bound

$$(2.15) \quad \|\mathbf{z}\|_\omega \leq \sqrt{2n} \|\mathbf{z}\|_\infty$$

since the ω_i are positive and sum to 2.

3. Interpolation error in \mathcal{H}^1 . Our error analysis is based on a result concerning the error in interpolation at the point set $\tau_i, 0 \leq i \leq N$, where τ_i for $i > 0$ are the N Gauss quadrature points on Ω , and $\tau_0 = -1$. In [3, Thm. 4.8], Bernardi and Maday give an overview of the analysis of error in \mathcal{H}^1 for interpolation at Gauss quadrature points. Here we take into account the additional interpolation point $\tau_0 = -1$ and provide a complete derivation of the interpolation error estimate.

LEMMA 3.1. *If $u \in \mathcal{H}^\eta(\Omega)$ for some $\eta \geq 1$, then there exists a constant c , independent of N and η , such that*

$$(3.1) \quad \|u - u^I\|_{\mathcal{H}^1(\Omega)} \leq (c/N)^{p-3/2} \|u\|_{\mathcal{H}^p(\Omega)}, \quad p = \min\{\eta, N + 1\},$$

where $u^I \in \mathcal{P}_N$ is the interpolant of u satisfying $u^I(\tau_i) = u(\tau_i), 0 \leq i \leq N$, and $N > 0$.

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. Let ℓ denote the linear function for which $\ell(\pm 1) = u(\pm 1)$. If the lemma holds for all $u \in \mathcal{H}_0^1(\Omega) \cap \mathcal{H}^\eta(\Omega)$, then it holds for all $u \in \mathcal{H}^\eta(\Omega)$ since $|u - u^I|_{\mathcal{H}^1(\Omega)} = |(u - \ell) - (u - \ell)^I|_{\mathcal{H}^1(\Omega)}$ and $\|u - \ell\|_{\mathcal{H}^p(\Omega)} \leq c\|u\|_{\mathcal{H}^p(\Omega)}$. Hence, without loss of generality, 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

$$(3.2) \quad |u - u^I|_{\mathcal{H}^1(\Omega)} \leq |E_N|_{\mathcal{H}^1(\Omega)} + |e_N|_{\mathcal{H}^1(\Omega)}.$$

In [13, Prop. 3.1] it is shown that

$$(3.3) \quad |E_N|_{\mathcal{H}^1(\Omega)} \leq (c/N)^{p-1} \|u\|_{\mathcal{H}^p(\Omega)}, \quad \text{where } p = \min\{\eta, N + 1\}.$$

We establish below the bound

$$(3.4) \quad |e_N|_{\mathcal{H}^1(\Omega)} \leq c\sqrt{N}|E_N|_{\mathcal{H}^1(\Omega)}.$$

Estimate (3.1) follows, for an appropriate choice of c , by combining (3.2)–(3.4).

The proof of (3.4) proceeds as follows: Let ϕ_N be defined by

$$(3.5) \quad \phi_N(\tau) = e_N(\tau) - e_N(1)w_N(\tau), \quad \text{where } w_N(\tau) = \frac{(1 + \tau)P'_N(\tau)}{N(N + 1)}.$$

Since P_N , the Legendre polynomial of degree N , satisfies $P'_N(1) = N(N + 1)/2$, it follows that $w_N(1) = 1$ and $\phi_N(1) = 0$. Moreover, since $w_N(-1) = 0$ and $e_N(-1) = e_N(\tau_0) = 0$, we conclude that $\phi_N(-1) = 0$ and $\phi_N \in \mathcal{P}_N^0$. In [3, Lem. 4.4] it is shown that any $\phi_N \in \mathcal{P}_N^0$ satisfies

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

Hence, by (3.5), we have

$$(3.6) \quad |e_N|_{\mathcal{H}^1(\Omega)} \leq cN \left(\int_{\Omega} \frac{\phi_N^2(\tau)}{1 - \tau^2} d\tau \right)^{1/2} + |w_N|_{\mathcal{H}^1(\Omega)} |e_N(1)|.$$

Rodrigues' formula for P_N and integration by parts give

$$\|P'_N\|_{\mathcal{L}^2(\Omega)} = \sqrt{N(N + 1)}.$$

It follows that

$$\|w_N\|_{\mathcal{L}^2(\Omega)} \leq \frac{2}{\sqrt{N(N + 1)}} \leq \frac{2}{N}.$$

Bellman's [1] inequality

$$\int_{\Omega} p'(\tau)^2 dt \leq \frac{(N + 1)^4}{2} \int_{\Omega} p(\tau)^2 dt \quad \text{for all } p \in \mathcal{P}_N$$

implies that

$$|w_N|_{\mathcal{H}^1(\Omega)} = \|w'_N\|_{\mathcal{L}^2(\Omega)} \leq \frac{\sqrt{2}(N + 1)^2}{N} \leq cN.$$

We combine this bound for $|w_N|_{\mathcal{H}^1(\Omega)}$ with (3.6) to obtain

$$(3.7) \quad |e_N|_{\mathcal{H}^1(\Omega)} \leq cN \left[\left(\int_{\Omega} \frac{\phi_N^2(\tau)}{1 - \tau^2} d\tau \right)^{1/2} + |e_N(1)| \right].$$

Since $e_N = E_N^I$ and $E_N(-1) = 0$, the interpolant can be expressed

$$e_N(\tau) = E_N^I(\tau) = \sum_{i=1}^N E_N(\tau_i) \left(\frac{(\tau + 1)P_N(\tau)}{(\tau_i + 1)P'_N(\tau_i)(\tau - \tau_i)} \right),$$

where the expression in parentheses is the Lagrange interpolating polynomial; it vanishes at $\tau = \tau_j$ for $0 \leq j \leq N$ and $j \neq i$ since the numerator vanishes, while it is one at $\tau = \tau_i$ since

$$\frac{P_N(\tau)/(\tau - \tau_i)}{P'_N(\tau)} \Big|_{\tau=\tau_i} = \frac{\prod_{j \neq i} (\tau_i - \tau_j)}{\prod_{j \neq i} (\tau_i - \tau_j)} = 1.$$

At $\tau = 1$, it follows from the Schwarz inequality that

$$|e_N(1)| \leq \sum_{i=1}^N \frac{2|E_N(\tau_i)|}{(1 - \tau_i^2)|P'_N(\tau_i)|} \leq 2\sqrt{N} \left(\sum_{i=1}^N \frac{E_N^2(\tau_i)}{(1 - \tau_i^2)^2 P'_N(\tau_i)^2} \right)^{1/2}.$$

Replace $2/[(1 - \tau_i^2)P'_N(\tau_i)^2]$ by ω_i using (1.10) to obtain

$$(3.8) \quad |e_N(1)| \leq \sqrt{2N} \left(\sum_{i=1}^N \frac{\omega_i E_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2}.$$

Since $E_N \in \mathcal{H}_0^1(\Omega)$, it follows from [3, Lem. 4.3] that

$$(3.9) \quad \left(\sum_{i=1}^N \frac{\omega_i E_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2} \leq c \left[\left(\int_{\Omega} \frac{E_N^2(\tau)}{1 - \tau^2} d\tau \right)^{1/2} + N^{-1} |E_N|_{\mathcal{H}^1(\Omega)} \right].$$

By Proposition 10.1 in Appendix 2,

$$(3.10) \quad N \left[\left(\int_{\Omega} \frac{E_N^2(\tau)}{1 - \tau^2} d\tau \right)^{1/2} + N^{-1} |E_N|_{\mathcal{H}^1(\Omega)} \right] \leq 2|E_N|_{\mathcal{H}^1(\Omega)}.$$

Together, (3.9) and (3.10) give

$$(3.11) \quad \left(\sum_{i=1}^N \frac{\omega_i E_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2} \leq (c/N) |E_N|_{\mathcal{H}^1(\Omega)}.$$

Combine (3.8) and (3.11) to obtain

$$(3.12) \quad |e_N(1)| \leq (c/\sqrt{N}) |E_N|_{\mathcal{H}^1(\Omega)}.$$

Since $\phi_N \in \mathcal{P}_N^0$, we deduce that $\phi_N^2(\tau)/(1 - \tau^2) \in \mathcal{P}_{2N-2}$. Consequently, N -point Gaussian quadrature is exact, and we have

$$\begin{aligned} \left(\int_{\Omega} \frac{\phi_N^2(\tau)}{1 - \tau^2} d\tau \right)^{1/2} &= \left(\sum_{i=1}^N \frac{\omega_i \phi_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2} \\ &\leq \left(\sum_{i=1}^N \frac{\omega_i e_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2} + |e_N(1)| \left(\sum_{i=1}^N \frac{\omega_i w_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2} \\ &= \left(\sum_{i=1}^N \frac{\omega_i E_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2} + |e_N(1)| \left(\sum_{i=1}^N \frac{\omega_i w_N^2(\tau_i)}{1 - \tau_i^2} \right)^{1/2}. \end{aligned}$$

The last equality holds since $e_N = E_N$ at the collocation points τ_i , $1 \leq i \leq N$. In [3, eq. (4.15)], it is proved that

$$(3.13) \quad \sum_{i=1}^N \frac{\omega_i w_N^2(\tau_i)}{1 - \tau_i^2} \leq c.$$

Combine (3.11), (3.12), and (3.13) to obtain

$$(3.14) \quad \left(\int_{\Omega} \frac{\phi_N^2(\tau)}{(1 - \tau^2)} d\tau \right)^{1/2} \leq (c/\sqrt{N})|E_N|_{\mathcal{H}^1(\Omega)}.$$

Finally, (3.7), (3.12), and (3.14) yield (3.4), which completes the proof. □

4. Analysis of the residual. In this section, we establish a bound for the distance from $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ to $\mathcal{F}(\mathbf{U}^*)$. This bound ultimately enters into the right-hand side of the error estimate (1.9).

LEMMA 4.1. *If \mathbf{x}^* and $\mathbf{\lambda}^* \in \mathcal{H}^\eta(\Omega; \mathbb{R}^n)$ for some $\eta \geq 2$, then there exists a constant c , independent of N and η , such that*

$$(4.1) \quad \text{dist}[\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*), \mathcal{F}(\mathbf{U}^*)] \leq \left(\frac{c}{N}\right)^{p-3/2} (\|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)} + \|\mathbf{\lambda}^*\|_{\mathcal{H}^p(\Omega)}),$$

where $p = \min\{\eta, N + 1\}$. The left-hand side of (4.1) denotes the distance from $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ to $\mathcal{F}(\mathbf{U}^*)$ relative to $\|\cdot\|_{\mathcal{Y}}$.

Proof. Since $\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ appears throughout the analysis, it is abbreviated \mathcal{T}^* . The feasibility of \mathbf{x}^* in (1.1) implies that $\mathbf{X}_0^* = \mathbf{x}_0$, or $\mathcal{T}_0^* = \mathbf{0}$. By the costate equation (1.3), $\mathbf{\Lambda}_{N+1}^* = \mathbf{\lambda}^*(1) = \nabla C(\mathbf{x}^*(1)) = \nabla C(\mathbf{X}_{N+1}^*)$, which implies that $\mathcal{T}_5^* = \mathbf{0}$. By the Pontryagin minimum principle (1.4),

$$-\nabla_u H(\mathbf{X}_i^*, \mathbf{U}_i^*, \mathbf{\Lambda}_i^*) = -\nabla_u H(\mathbf{x}^*(\tau_i), \mathbf{u}^*(\tau_i), \mathbf{\lambda}^*(\tau_i)) \in \mathcal{F}(\mathbf{u}^*(\tau_i)) = \mathcal{F}(\mathbf{U}_i^*),$$

$1 \leq i \leq N$. Thus $\mathcal{T}_0^* = \mathcal{T}_5^* = \mathbf{0}$, $\mathcal{T}_6^* \in \mathcal{F}_6(\mathbf{U}^*)$.

Now let us consider \mathcal{T}_1 . Since \mathbf{D} is a differentiation matrix associated with the collocation points, we have

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

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

$$(4.3) \quad \mathbf{f}(\mathbf{X}_i^*, \mathbf{U}_i^*) = \mathbf{f}(\mathbf{x}^*(\tau_i), \mathbf{u}^*(\tau_i)) = \dot{\mathbf{x}}^*(\tau_i).$$

Combine (4.2) and (4.3) to obtain

$$(4.4) \quad \mathcal{T}_{1i}^* = \dot{\mathbf{x}}^I(\tau_i) - \dot{\mathbf{x}}^*(\tau_i), \quad 1 \leq i \leq N.$$

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

$$\|\mathcal{T}_1^*\|_{\omega} = \|\dot{\mathbf{x}}^I - (\dot{\mathbf{x}}^*)^J\|_{\mathcal{L}^2(\Omega)} \leq \|\dot{\mathbf{x}}^I - \dot{\mathbf{x}}^*\|_{\mathcal{L}^2(\Omega)} + \|\dot{\mathbf{x}}^* - (\dot{\mathbf{x}}^*)^J\|_{\mathcal{L}^2(\Omega)}.$$

By Lemma 3.1, $\|\dot{\mathbf{x}}^I - \dot{\mathbf{x}}^*\|_{\mathcal{L}^2(\Omega)} \leq (c/N)^{p-3/2} \|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)}$. By [3, Cor. 3.2] and [13, Prop. 3.1], it follows that $\|\dot{\mathbf{x}}^* - (\dot{\mathbf{x}}^*)^J\|_{\mathcal{L}^2(\Omega)} \leq (c/N)^{p-1} \|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)}$. Hence, we have

$$(4.5) \quad \|\mathcal{T}_1^*\|_{\omega} = \|\dot{\mathbf{x}}^I - (\dot{\mathbf{x}}^*)^J\|_{\mathcal{L}^2(\Omega)} \leq (c/N)^{p-3/2} \|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)}.$$

The analysis of \mathcal{T}_4 is identical to that of \mathcal{T}_1 ; the only adjustment is that $\boldsymbol{\lambda}^I$ is the interpolating polynomial that passes through $\boldsymbol{\lambda}^*(\tau_j)$ for $1 \leq j \leq N + 1$. Next, let us consider

$$(4.6) \quad \mathcal{T}_2^* = \mathbf{x}^*(1) - \mathbf{x}^*(-1) - \sum_{j=1}^N \omega_j \mathbf{f}(\mathbf{x}^*(\tau_j), \mathbf{u}^*(\tau_j)).$$

By the fundamental theorem of calculus and the exactness of Gaussian quadrature, we have

$$(4.7) \quad \mathbf{0} = \mathbf{x}^I(1) - \mathbf{x}^I(-1) - \int_{\Omega} \dot{\mathbf{x}}^I(t) dt = \mathbf{x}^I(1) - \mathbf{x}^I(-1) - \sum_{j=1}^N \omega_j \dot{\mathbf{x}}^I(\tau_j).$$

Subtract (4.7) from (4.6) and substitute $\dot{\mathbf{x}}^*(\tau_j) = \mathbf{f}(\mathbf{x}^*(\tau_j), \mathbf{u}^*(\tau_j))$ to obtain

$$(4.8) \quad \mathcal{T}_2^* = (\mathbf{x}^* - \mathbf{x}^I)(1) + \sum_{j=1}^N \omega_j (\dot{\mathbf{x}}^I(\tau_j) - \dot{\mathbf{x}}^*(\tau_j)).$$

Since the ω_i are positive and sum to 2, it follows from the Schwarz inequality and (4.5) that

$$(4.9) \quad \left| \sum_{j=1}^N \omega_j (\dot{\mathbf{x}}^I(\tau_j) - \dot{\mathbf{x}}^*(\tau_j)) \right| \leq \left(\sum_{j=1}^N \omega_j \right)^{1/2} \left(\sum_{j=1}^N \omega_j |\dot{\mathbf{x}}^I(\tau_j) - \dot{\mathbf{x}}^*(\tau_j)|^2 \right)^{1/2} = \sqrt{2} \|\dot{\mathbf{x}}^I - (\dot{\mathbf{x}}^*)^J\|_{\mathcal{L}^2(\Omega)} \leq (c/N)^{p-3/2} \|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)}.$$

Also, writing $(\mathbf{x}^* - \mathbf{x}^I)(1)$ as the integral of the derivative from -1 to 1 and applying the Schwarz inequality yields

$$(4.10) \quad |\mathbf{x}^*(1) - \mathbf{x}^I(1)| \leq \sqrt{2} \|\dot{\mathbf{x}}^* - \dot{\mathbf{x}}^I\|_{\mathcal{L}^2(\Omega)} \leq (c/N)^{p-3/2} \|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)},$$

where the last equality is by Lemma 3.1. Combine (4.8), (4.9), and (4.10) to obtain $|\mathcal{T}_2^*| \leq (c/N)^{p-3/2} \|\mathbf{x}^*\|_{\mathcal{H}^p(\Omega)}$. The analysis of \mathcal{T}_3 is the same as that of \mathcal{T}_2 . This completes the proof. \square

5. Invertibility of linearized dynamics. In this section, we introduce the linearized inclusion and established the invertibility of the linearized dynamics for both the state and costate. Given $\mathbf{Y} \in \mathcal{Y}$, the linearized problem is to find $(\mathbf{X}, \mathbf{U}, \boldsymbol{\Lambda})$ such that

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

Here $\nabla \mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \boldsymbol{\Lambda}^*)[\mathbf{X}, \mathbf{U}, \boldsymbol{\Lambda}]$ denotes the derivative of \mathcal{T} evaluated at $(\mathbf{X}^*, \mathbf{U}^*, \boldsymbol{\Lambda}^*)$ operating on $[\mathbf{X}, \mathbf{U}, \boldsymbol{\Lambda}]$. Since $\nabla \mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \boldsymbol{\Lambda}^*)$ appears frequently in the analysis, it is abbreviated $\nabla \mathcal{T}^*$. This derivative involves the following matrices:

$$\mathbf{A}_i = \mathbf{A}(\tau_i), \quad \mathbf{B}_i = \mathbf{B}(\tau_i), \quad \mathbf{Q}_i = \mathbf{Q}(\tau_i), \quad \mathbf{S}_i = \mathbf{S}(\tau_i), \quad \mathbf{R}_i = \mathbf{R}(\tau_i).$$

With this notation, the 7 components of $\nabla \mathcal{T}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}]$ are as follows:

$$\begin{aligned} \nabla \mathcal{T}_0^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{X}_0, \\ \nabla \mathcal{T}_{1i}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \left(\sum_{j=1}^N D_{ij} \mathbf{X}_j \right) - \mathbf{A}_i \mathbf{X}_i - \mathbf{B}_i \mathbf{U}_i, \quad 1 \leq i \leq N, \\ \nabla \mathcal{T}_2^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{X}_{N+1} - \mathbf{X}_0 - \sum_{j=1}^N \omega_j (\mathbf{A}_j \mathbf{X}_j + \mathbf{B}_j \mathbf{U}_j), \\ \nabla \mathcal{T}_3^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{\Lambda}_{N+1} - \mathbf{\Lambda}_0 + \sum_{j=1}^N \omega_j (\mathbf{A}_j^\top \mathbf{\Lambda}_j + \mathbf{Q}_j \mathbf{X}_j + \mathbf{S}_j \mathbf{U}_j), \\ \nabla \mathcal{T}_{4i}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \left(\sum_{j=1}^{N+1} D_{ij}^\dagger \mathbf{\Lambda}_j \right) + \mathbf{A}_i^\top \mathbf{\Lambda}_i + \mathbf{Q}_i \mathbf{X}_i + \mathbf{S}_i \mathbf{U}_i, \quad 1 \leq i \leq N, \\ \nabla \mathcal{T}_5^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= \mathbf{\Lambda}_{N+1} - \mathbf{T} \mathbf{X}_{N+1}, \\ \nabla \mathcal{T}_{6i}^*[\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}] &= -(\mathbf{S}_i^\top \mathbf{X}_i + \mathbf{R}_i \mathbf{U}_i + \mathbf{B}_i^\top \mathbf{\Lambda}_i), \quad 1 \leq i \leq N. \end{aligned}$$

Let us first study the invertibility of the linearized dynamics. This amounts to solving for the state in (5.1) for given values of the control.

LEMMA 5.1. *If (P1), (P2), and (A2) hold, then for each \mathbf{q}_0 and $\mathbf{q}_1 \in \mathbb{R}^n$ and $\mathbf{p} \in \mathbb{R}^{nN}$ with $\mathbf{p}_i \in \mathbb{R}^n$, $1 \leq i \leq N$, the linear system*

$$(5.2) \quad \left(\sum_{j=0}^N D_{ij} \mathbf{X}_j \right) - \mathbf{A}_i \mathbf{X}_i = \mathbf{p}_i \quad 1 \leq i \leq N,$$

$$(5.3) \quad \mathbf{X}_{N+1} - \mathbf{X}_0 - \sum_{j=1}^N \omega_j \mathbf{A}_j \mathbf{X}_j = \mathbf{q}_1, \quad \mathbf{X}_0 = \mathbf{q}_0,$$

has a unique solution $\mathbf{X} \in \mathbb{R}^{n(N+2)}$. Moreover, there exists a constant c , independent of N , such that

$$(5.4) \quad \|\mathbf{X}\|_\infty \leq c(|\mathbf{q}_0| + |\mathbf{q}_1| + \|\mathbf{p}\|_\omega).$$

Proof. Let $\bar{\mathbf{X}}$ be the vector obtained by vertically stacking \mathbf{X}_1 through \mathbf{X}_N , let \mathbf{A} be the block diagonal matrix with i th diagonal block \mathbf{A}_i , $1 \leq i \leq N$, and define $\bar{\mathbf{D}} = \mathbf{D}_{1:N} \otimes \mathbf{I}_n$, where \otimes is the Kronecker product. With this notation, the linear system (5.2) can be expressed

$$(5.5) \quad (\bar{\mathbf{D}} - \mathbf{A}) \bar{\mathbf{X}} = \mathbf{p} - (\mathbf{D}_0 \otimes \mathbf{I}_n) \mathbf{q}_0.$$

Here \mathbf{D}_0 is the first column of \mathbf{D} and the $\mathbf{X}_0 = \mathbf{q}_0$ component of \mathbf{X} has been moved to the right side of the equation. By (P1) $\mathbf{D}_{1:N}$ is invertible, which implies that $\bar{\mathbf{D}}$ is invertible with $\bar{\mathbf{D}}^{-1} = \mathbf{D}_{1:N}^{-1} \otimes \mathbf{I}_n$. Moreover, $\|\bar{\mathbf{D}}^{-1}\|_\infty = \|\mathbf{D}_{1:N}^{-1}\|_\infty \leq 2$ by (P1). By (A2) $\|\mathbf{A}\|_\infty \leq \beta$ and $\|\bar{\mathbf{D}}^{-1} \mathbf{A}\|_\infty \leq \|\bar{\mathbf{D}}^{-1}\|_\infty \|\mathbf{A}\|_\infty \leq 2\beta < 1$ since $\beta < 1/2$. By [29, p. 351], $\mathbf{I} - \bar{\mathbf{D}}^{-1} \mathbf{A}$ is invertible and

$$(5.6) \quad \left\| \left(\mathbf{I} - \bar{\mathbf{D}}^{-1} \mathbf{A} \right)^{-1} \right\|_\infty \leq 1/(1 - 2\beta).$$

Multiply (5.5) first by $\bar{\mathbf{D}}^{-1}$ and then by $(\mathbf{I} - \bar{\mathbf{D}}^{-1}\mathbf{A})^{-1}$ to obtain

$$\bar{\mathbf{X}} = \left(\mathbf{I} - \bar{\mathbf{D}}^{-1}\mathbf{A}\right)^{-1} \left(\bar{\mathbf{D}}^{-1}\mathbf{p} + \bar{\mathbf{D}}^{-1}(\mathbf{D}_0 \otimes \mathbf{I}_n)\mathbf{q}_0\right).$$

We take the norm of $\bar{\mathbf{X}}$ and utilize (5.6) to find that

$$(5.7) \quad \|\bar{\mathbf{X}}\|_\infty \leq \left(\frac{1}{1-2\beta}\right) \left[\|\bar{\mathbf{D}}^{-1}\mathbf{p}\|_\infty + \|\bar{\mathbf{D}}^{-1}(\mathbf{D}_0 \otimes \mathbf{I}_n)\mathbf{q}_0\|_\infty\right].$$

Since the polynomial that is identically equal to $\mathbf{1}$ has derivative $\mathbf{0}$ and since \mathbf{D} is a differentiation matrix, we have $\mathbf{D}\mathbf{1} = \mathbf{0}$, which implies that $\mathbf{D}_{1:N}\mathbf{1} = -\mathbf{D}_0$. Hence, $\mathbf{D}_{1:N}^{-1}\mathbf{D}_0 = -\mathbf{1}$. It follows that

$$(\bar{\mathbf{D}})^{-1}[\mathbf{D}_0 \otimes \mathbf{I}_n] = [(\mathbf{D}_{1:N})^{-1} \otimes \mathbf{I}_n][\mathbf{D}_0 \otimes \mathbf{I}_n] = -\mathbf{1} \otimes \mathbf{I}_n.$$

We make this substitution in (5.7) and use the bound for the sup-norm in terms of the Euclidean norm to obtain

$$\|\bar{\mathbf{X}}\|_\infty \leq \left(\frac{1}{1-2\beta}\right) \left(\|\bar{\mathbf{D}}^{-1}\mathbf{p}\|_\infty + |\mathbf{q}_0|\right).$$

Observe that

$$\bar{\mathbf{D}}^{-1}\mathbf{p} = (\mathbf{D}_{1:N}^{-1} \otimes \mathbf{I}_n)\mathbf{p} = \left(\mathbf{D}_{1:N}^{-1}\mathbf{W}^{-1/2} \otimes \mathbf{I}_n\right) \left[\left(\mathbf{W}^{1/2} \otimes \mathbf{I}_n\right)\mathbf{p}\right],$$

where \mathbf{W} is the diagonal matrix with the quadrature weights on the diagonal. Based on this identity, an element of $\bar{\mathbf{D}}^{-1}\mathbf{p}$ is the dot product between a row of $(\mathbf{D}_{1:N}^{-1}\mathbf{W}^{-1/2} \otimes \mathbf{I}_n)$ and the column vector $(\mathbf{W}^{1/2} \otimes \mathbf{I}_n)\mathbf{p}$.

By the Schwarz inequality, this dot product is bounded by the product between the largest Euclidean length of the rows of the matrix and the Euclidean length of the vector. By (P2), the Euclidean lengths of the rows of $[\mathbf{W}^{1/2}\mathbf{D}_{1:N}]^{-1}$ are bounded by $\sqrt{2}$, and by the definition of the ω -norm, we have $|(\mathbf{W}^{1/2} \otimes \mathbf{I}_n)\mathbf{p}| = \|\mathbf{p}\|_\omega$. Hence, we have

$$(5.8) \quad \|\bar{\mathbf{D}}^{-1}\mathbf{p}\|_\infty \leq \sqrt{2}\|\mathbf{p}\|_\omega \quad \text{and} \quad \|\bar{\mathbf{X}}\|_\infty \leq \left(\frac{1}{1-2\beta}\right) \left(\sqrt{2}\|\mathbf{p}\|_\omega + |\mathbf{q}_0|\right).$$

By the first equation in (5.3),

$$\|\mathbf{X}_{N+1}\|_\infty \leq \|\mathbf{q}_0\|_\infty + \|\mathbf{q}_1\|_\infty + \sum_{j=1}^N \omega_j \|\mathbf{A}_j\|_\infty \|\mathbf{X}_j\|_\infty.$$

Since the ω_j sum to 2, $\|\mathbf{A}_j\| \leq \beta < 1/2$, and the sup-norm is bounded by the Euclidean norm, it follows that

$$(5.9) \quad \|\mathbf{X}_{N+1}\|_\infty \leq |\mathbf{q}_0| + |\mathbf{q}_1| + \|\bar{\mathbf{X}}\|_\infty.$$

Combine (5.8) and (5.9) to obtain (5.4). □

Next, let us consider the linearized costate dynamics.

LEMMA 5.2. *If (P1), (P2), and (A2) hold, then for each \mathbf{q}_0 and $\mathbf{q}_1 \in \mathbb{R}^n$ and $\mathbf{p} \in \mathbb{R}^{nN}$ with $\mathbf{p}_i \in \mathbb{R}^n$, $1 \leq i \leq N$, the linear system*

$$(5.10) \quad \left(\sum_{j=1}^{N+1} D_{ij}^\dagger \Lambda_j \right) + \mathbf{A}_i^\top \Lambda_i = \mathbf{p}_i \quad 1 \leq i \leq N,$$

$$(5.11) \quad \Lambda_{N+1} - \Lambda_0 + \sum_{j=1}^N \omega_j \mathbf{A}_j^\top \Lambda_j = \mathbf{q}_0, \quad \Lambda_{N+1} = \mathbf{q}_1,$$

has a unique solution $\Lambda \in \mathbb{R}^{n(N+2)}$. Moreover, there exists a constant c , independent of N , such that

$$(5.12) \quad \|\Lambda\|_\infty \leq c(|\mathbf{q}_0| + |\mathbf{q}_1| + \|\mathbf{p}\|_\omega).$$

Proof. As noted in (2.10), \mathbf{D}^\dagger is a differentiation matrix, analogous to \mathbf{D} , except that \mathbf{D}^\dagger operates on function values at $\tau_1, \dots, \tau_{N+1}$, while \mathbf{D} operates on function values at τ_0, \dots, τ_N . The proof is identical to that of Lemma 5.1 except that Λ_{N+1} plays the role of \mathbf{X}_0 , while Λ_0 plays the role of \mathbf{X}_{N+1} . \square

6. Invertibility of $\mathcal{F} - \nabla \mathcal{T}^*$ and Lipschitz continuity of the inverse. The invertibility of $\mathcal{F} - \nabla \mathcal{T}^*$ is now established.

PROPOSITION 6.1. *If (A1)–(A2) and (P1)–(P2) hold, then for each $\mathbf{Y} \in \mathcal{Y}$, there is a unique solution $(\mathbf{X}, \mathbf{U}, \Lambda)$ to (5.1).*

Proof. As in our earlier work [5, 6, 7, 10, 21, 24, 25, 26], we formulate a strongly convex quadratic programming problem whose first-order optimality conditions reduce to (5.1). Let us consider the problem

$$(6.1) \quad \begin{aligned} & \text{minimize} && \frac{1}{2} \mathcal{Q}(\mathbf{X}, \mathbf{U}) + \mathcal{L}(\mathbf{X}, \mathbf{U}, \mathbf{Y}) \\ & \text{subject to} && \sum_{j=1}^N D_{ij} \mathbf{X}_j = \mathbf{A}_i \mathbf{X}_i + \mathbf{B}_i \mathbf{U}_i - \mathbf{y}_{1i}, \quad \mathbf{U}_i \in \mathcal{U}, \quad 1 \leq i \leq N, \\ & && \mathbf{X}_0 = -\mathbf{y}_0, \quad \mathbf{X}_{N+1} = \mathbf{X}_0 - \mathbf{y}_2 + \sum_{j=1}^N \omega_j (\mathbf{A}_j \mathbf{X}_j + \mathbf{B}_j \mathbf{U}_j). \end{aligned}$$

Here the quadratic and linear terms in the objective are

$$\begin{aligned} \mathcal{Q}(\mathbf{X}, \mathbf{U}) &= \mathbf{X}_{N+1}^\top \mathbf{T} \mathbf{X}_{N+1} + \sum_{i=1}^N \omega_i (\mathbf{X}_i^\top \mathbf{Q}_i \mathbf{X}_i + 2\mathbf{X}_i^\top \mathbf{S}_i \mathbf{U}_i + \mathbf{U}_i^\top \mathbf{R}_i \mathbf{U}_i), \\ \mathcal{L}(\mathbf{X}, \mathbf{U}, \mathbf{Y}) &= \mathbf{X}_0^\top \left(\mathbf{y}_3 - \sum_{i=1}^N \omega_i \mathbf{y}_{4i} \right) - \mathbf{y}_5^\top \mathbf{X}_{N+1} + \sum_{i=1}^N \omega_i (\mathbf{y}_{4i}^\top \mathbf{X}_i - \mathbf{y}_{6i}^\top \mathbf{U}_i). \end{aligned}$$

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 program (6.1) is feasible for any choice of \mathbf{y}_0 , \mathbf{y}_1 , and \mathbf{y}_2 . Since $\mathbf{X}_0 = -\mathbf{y}_0$, \mathbf{X}_0 can be eliminated from the quadratic program (6.1). By (A1), the quadratic program is strongly convex with respect to $\mathbf{X}_1, \dots, \mathbf{X}_{N+1}$, and $\mathbf{U}_1, \dots, \mathbf{U}_N$. Hence, there exists a unique state and control solving (6.1). Next, we will show that the first-order optimality conditions for (6.1) reduce to (5.1). These conditions hold since \mathcal{U} has nonempty interior and the state dynamics have full row rank by Lemma 5.1. Due to the convexity of the objective and constraints, the first-order optimality conditions are both necessary and sufficient for optimality. Uniqueness of \mathbf{X} and \mathbf{U} is due to (A1) and the strong convexity of (6.1). Uniqueness of Λ is by Lemma 5.2.

Now let us show that (5.1) corresponds to the optimality conditions for (6.1). Components 0, 1, and 2 of (5.1) are simply the constraints of (6.1). The remaining optimality conditions are associated with the Lagrangian L given by

$$\begin{aligned} L(\boldsymbol{\mu}, \mathbf{X}, \mathbf{U}) &= \frac{1}{2}\mathcal{Q}(\mathbf{X}, \mathbf{U}) + \mathcal{L}(\mathbf{X}, \mathbf{U}, \mathbf{Y}) + \sum_{i=1}^N \left\langle \boldsymbol{\mu}_i, \mathbf{A}_i \mathbf{X}_i + \mathbf{B}_i \mathbf{U}_i - \mathbf{y}_{1i} - \sum_{j=0}^N D_{ij} \mathbf{X}_j \right\rangle \\ &\quad - \langle \boldsymbol{\mu}_0, \mathbf{X}_0 + \mathbf{y}_1 \rangle + \left\langle \boldsymbol{\mu}_{N+1}, \mathbf{X}_0 - \mathbf{X}_{N+1} - \mathbf{y}_2 + \sum_{j=1}^N \omega_j (\mathbf{A}_j \mathbf{X}_j + \mathbf{B}_j \mathbf{U}_j) \right\rangle. \end{aligned}$$

The negative derivative of the Lagrangian with respect to \mathbf{U}_i is

$$\omega_i (\mathbf{y}_{6i} - \mathbf{S}_i^T \mathbf{X}_i - \mathbf{R}_i \mathbf{U}_i - \mathbf{B}_i^T \boldsymbol{\mu}_{N+1}) - \mathbf{B}_i^T \boldsymbol{\mu}_i.$$

Substitute $\boldsymbol{\mu}_{N+1} = \boldsymbol{\Lambda}_{N+1}$ and $\boldsymbol{\mu}_i = \omega_i (\boldsymbol{\Lambda}_i - \boldsymbol{\Lambda}_{N+1})$, $1 \leq i \leq N$. The requirement that the resulting vector lies in $N_{\mathcal{U}}(\mathbf{U}_i)$ is the sixth component of (5.1). Equate to zero the derivative of the Lagrangian with respect to \mathbf{X}_{N+1} to obtain

$$\mathbf{0} = \mathbf{T} \mathbf{X}_{N+1} - \mathbf{y}_5 - \boldsymbol{\mu}_{N+1} = \mathbf{T} \mathbf{X}_{N+1} - \mathbf{y}_5 - \boldsymbol{\Lambda}_{N+1}.$$

This is the fifth component of (5.1). The derivative of the Lagrangian with respect to \mathbf{X}_j , $1 \leq j \leq N$, gives the relation

$$\sum_{i=1}^N D_{ij} \boldsymbol{\mu}_i = \mathbf{A}_j^T (\boldsymbol{\mu}_j + \omega_j \boldsymbol{\mu}_{N+1}) + \omega_j (\mathbf{Q}_j \mathbf{X}_j + \mathbf{S}_j \mathbf{U}_j + \mathbf{y}_{4j}).$$

Change variables from $\boldsymbol{\mu}$ to $\boldsymbol{\Lambda}$ and substitute for D_{ij} using (2.6) to obtain the fourth component of (5.1). Finally, differentiate the Lagrangian with respect to \mathbf{X}_0 to obtain

$$\boldsymbol{\mu}_{N+1} - \boldsymbol{\mu}_0 + \mathbf{y}_3 - \sum_{i=1}^N \omega_i \mathbf{y}_{4i} - \sum_{i=1}^N D_{i0} \boldsymbol{\mu}_i = \mathbf{0}.$$

Substitute for the D_{i0} sum using both (2.11) and the fourth component of (5.1) to obtain the third component of (5.1). \square

We now wish to bound the change in the solution of (6.1) in terms of the change in \mathbf{Y} . Let $\boldsymbol{\chi}(\mathbf{Y})$ denote the solution of the state dynamics (5.2)–(5.3) associated with $\mathbf{p} = \mathbf{y}_1$, $\mathbf{q}_0 = \mathbf{y}_0$, and $\mathbf{q}_1 = \mathbf{y}_2$. In (6.1) we make the change of variables $\mathbf{X} = \mathbf{Z} + \boldsymbol{\chi}(\mathbf{Y})$. The dynamics of (6.1) become

$$(6.2) \quad \sum_{j=1}^N D_{ij} \mathbf{Z}_j = \mathbf{A}_i \mathbf{Z}_i + \mathbf{B}_i \mathbf{U}_i, \quad \mathbf{Z}_0 = \mathbf{0}, \quad \mathbf{Z}_{N+1} = \sum_{j=1}^N \omega_j (\mathbf{A}_j \mathbf{Z}_j + \mathbf{B}_j \mathbf{U}_j).$$

Hence, the effect of the variable change is to remove \mathbf{Y} from the constraints. After the change of variables, the linear term in the objective of (6.1) reduces to

$$\begin{aligned} \hat{\mathcal{L}}(\mathbf{Z}, \mathbf{U}, \mathbf{Y}) &= \mathbf{y}_5^T \mathbf{Z}_{N+1} - \sum_{i=1}^N \omega_i (\mathbf{y}_{4i}^T \mathbf{Z}_i + \mathbf{y}_{6i}^T \mathbf{U}_i) \\ &\quad + \mathbf{Z}_{N+1}^T \mathbf{T} \boldsymbol{\chi}_{N+1}(\mathbf{Y}) + \sum_{i=1}^N \omega_i [\mathbf{Z}_i^T \mathbf{Q}_i \boldsymbol{\chi}_i(\mathbf{Y}) + \mathbf{U}_i^T \mathbf{S}_i^T \boldsymbol{\chi}_i(\mathbf{Y})], \end{aligned}$$

since $\mathbf{Z}_0 = \mathbf{0}$. Let $(\mathbf{Z}^j, \mathbf{U}^j)$ denote the solution of (6.1) corresponding to $\mathbf{Y}^j \in \mathcal{Y}$, $j = 1$, and 2. By [6, Lem. 4], the solution change satisfies the relation

$$(6.3) \quad \mathcal{Q}(\Delta\mathbf{Z}, \Delta\mathbf{U}) \leq |\hat{\mathcal{L}}(\Delta\mathbf{Z}, \Delta\mathbf{U}, \Delta\mathbf{Y})|,$$

where $\Delta\mathbf{Z} = \mathbf{Z}^1 - \mathbf{Z}^2$, $\Delta\mathbf{U} = \mathbf{U}^1 - \mathbf{U}^2$, and $\Delta\mathbf{Y} = \mathbf{Y}^1 - \mathbf{Y}^2$.

By (A1) we have the lower bound

$$(6.4) \quad \mathcal{Q}(\Delta\mathbf{Z}, \Delta\mathbf{U}) \geq \alpha(|\Delta\mathbf{Z}_{N+1}|^2 + \|\Delta\bar{\mathbf{Z}}\|_\omega^2 + \|\Delta\mathbf{U}\|_\omega^2),$$

where $\Delta\bar{\mathbf{Z}}$ is the subvector of $\Delta\mathbf{Z}$ corresponding to components 1 through N . The Schwarz inequality applied to the linear terms in (6.3) yields the upper bound

$$\begin{aligned} & \left| \hat{\mathcal{L}}(\Delta\mathbf{Z}, \Delta\mathbf{U}, \Delta\mathbf{Y}) \right| \\ & \leq c \left(|\Delta\mathbf{Z}_{N+1}| + \|\Delta\bar{\mathbf{Z}}\|_\omega + \|\Delta\mathbf{U}\|_\omega \right) \left(\|\Delta\mathbf{Y}\|_{\mathcal{Y}} + \|\bar{\chi}(\Delta\mathbf{Y})\|_\omega + |\chi_{N+1}(\Delta\mathbf{Y})| \right). \end{aligned}$$

By (2.15), $\|\bar{\chi}(\Delta\mathbf{Y})\|_\omega \leq \sqrt{2n}\|\bar{\chi}(\Delta\mathbf{Y})\|_\infty$, and by Lemma 5.1, $\|\chi(\Delta\mathbf{Y})\|_\infty \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}$. Hence, the upper bound simplifies to

$$(6.5) \quad \left| \hat{\mathcal{L}}(\Delta\mathbf{Z}, \Delta\mathbf{U}, \Delta\mathbf{Y}) \right| \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}} \left(|\Delta\mathbf{Z}_{N+1}| + \|\Delta\bar{\mathbf{Z}}\|_\omega + \|\Delta\mathbf{U}\|_\omega \right).$$

Combine (6.3)–(6.5) to obtain the Lipschitz result

$$(6.6) \quad |\Delta\mathbf{Z}_{N+1}| + \|\Delta\bar{\mathbf{Z}}\|_\omega + \|\Delta\mathbf{U}\|_\omega \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.$$

By (6.2), we see that $\Delta\mathbf{Z}$ is the solution of (5.2)–(5.3) corresponding to

$$\mathbf{q}_0 = \mathbf{0}, \quad \mathbf{p}_i = \mathbf{B}_i \Delta\mathbf{U}_i, \quad \mathbf{q}_1 = \sum_{j=1}^N \omega_j \mathbf{B}_j \Delta\mathbf{U}_j.$$

By (6.6), it follows that

$$(6.7) \quad \|\mathbf{B}\Delta\mathbf{U}\|_\omega \leq c\|\Delta\mathbf{U}\|_\omega \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}},$$

where \mathbf{B} is the block diagonal matrix with i th diagonal block \mathbf{B}_i . Moreover, by the Schwarz inequality and (6.7), we have

$$(6.8) \quad \left| \sum_{j=1}^N \omega_j \mathbf{B}_j \Delta\mathbf{U}_j \right| \leq \left(\sum_{j=1}^N \omega_j \right)^{1/2} \left[\sum_{j=1}^N \omega_j |\mathbf{B}_j \Delta\mathbf{U}_j|^2 \right]^{1/2} \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.$$

Hence, this choice for \mathbf{q}_0 , \mathbf{q}_1 , and \mathbf{p} together with Lemma 5.1 and the bounds (6.7) and (6.8) imply that $\|\Delta\mathbf{Z}\|_\infty \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}$. Since $\Delta\mathbf{X} = \Delta\mathbf{Z} + \chi(\Delta\mathbf{Y})$ where $\|\chi(\Delta\mathbf{Y})\|_\infty \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}$ by Lemma 5.1, we conclude that

$$(6.9) \quad \|\Delta\mathbf{X}\|_\infty \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.$$

Now consider the costate dynamics (5.10)–(5.11) with

$$\begin{aligned} \mathbf{q}_0 &= - \left(\Delta\mathbf{y}_3 + \sum_{j=1}^N \omega_j [\mathbf{Q}_j \Delta\mathbf{X}_j + \mathbf{S}_j \Delta\mathbf{U}_j] \right), \\ \mathbf{p}_i &= - (\Delta\mathbf{y}_{4i} + \mathbf{Q}_i \Delta\mathbf{X}_i + \mathbf{S}_i \Delta\mathbf{U}_i), \\ \mathbf{q}_1 &= -\Delta\mathbf{y}_5 + \mathbf{T} \Delta\mathbf{X}_{N+1}. \end{aligned}$$

By (2.15) and (6.9), we have

$$(6.10) \quad \|\mathbf{Q}\Delta\bar{\mathbf{X}}\|_\omega \leq c\|\Delta\bar{\mathbf{X}}\|_\omega \leq c\|\Delta\bar{\mathbf{X}}\|_\infty \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}},$$

where \mathbf{Q} is the block diagonal matrix with i th diagonal block \mathbf{Q}_i . The $\mathbf{S}_i\Delta\mathbf{U}_i$ term associated with \mathbf{p}_i can be analyzed as in (6.7) and the $\mathbf{S}_j\Delta\mathbf{U}_j$ terms in \mathbf{q}_0 can be analyzed as in (6.8). Analogous to the state dynamics, it follows from Lemma 5.2 that

$$(6.11) \quad \|\Delta\mathbf{\Lambda}\|_\infty \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.$$

Let $[\mathbf{X}(\mathbf{Y}), \mathbf{U}(\mathbf{Y}), \mathbf{\Lambda}(\mathbf{Y})]$ denote the solution of (5.1) for given $\mathbf{Y} \in \mathcal{Y}$. From the last component of the inclusion (5.1) and for any i between 1 and N , we have

$$[\mathbf{y}_6 - \mathbf{S}_i^\top \mathbf{X}_i(\mathbf{Y}) - \mathbf{R}_i \mathbf{U}_i(\mathbf{Y}) - \mathbf{B}_i^\top \mathbf{\Lambda}_i(\mathbf{Y})]^\top (\mathbf{V} - \mathbf{U}_i(\mathbf{Y})) \leq 0 \quad \text{for all } \mathbf{V} \in \mathcal{U}.$$

We add the inequality corresponding to $\mathbf{Y} = \mathbf{Y}^1$ and $\mathbf{V} = \mathbf{U}_i(\mathbf{Y}^2)$ to the inequality corresponding to $\mathbf{Y} = \mathbf{Y}^2$ and $\mathbf{V} = \mathbf{U}_i(\mathbf{Y}^1)$ to obtain the inequality

$$\Delta\mathbf{U}_i^\top \mathbf{R}_i \Delta\mathbf{U}_i \leq [-\Delta\mathbf{y}_6 + \mathbf{S}^\top \Delta\mathbf{X}_i + \mathbf{B}_i^\top \Delta\mathbf{\Lambda}_i]^\top \Delta\mathbf{U}_i.$$

By (A1) and the Schwarz inequality, it follows that

$$\|\Delta\mathbf{U}_i\|_\infty \leq |\Delta\mathbf{U}_i| \leq c(|\Delta\mathbf{y}_6| + |\Delta\mathbf{X}_i| + |\Delta\mathbf{\Lambda}_i|).$$

We utilize the previously established bounds (6.9) and (6.11) to obtain $\|\Delta\mathbf{U}\|_\infty \leq \|\Delta\mathbf{Y}\|_{\mathcal{Y}}$. The following lemma summarizes these observations.

LEMMA 6.2. *If (A1)–(A2) and (P1)–(P2) hold, then there exists a constant c , independent of N , such that the change $(\Delta\mathbf{X}, \Delta\mathbf{U}, \Delta\mathbf{\Lambda})$ in the solution of (5.1) corresponding to a change $\Delta\mathbf{Y}$ in $\mathbf{Y} \in \mathcal{Y}$ satisfies*

$$\max\{\|\Delta\mathbf{X}\|_\infty, \|\Delta\mathbf{U}\|_\infty, \|\Delta\mathbf{\Lambda}\|_\infty\} \leq c\|\Delta\mathbf{Y}\|_{\mathcal{Y}}.$$

Theorem 1.1 follows from Lemma 6.2 and Proposition 2.2; the proof is a small modification of the analysis in [26, Thm. 2.1]. The Lipschitz constant μ of Proposition 2.2 is the constant c of Lemma 6.2. Choose ε small enough that $\varepsilon\mu < 1$. When we compute the difference $\nabla\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) - \nabla\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$ for $(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda})$ near $(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)$, the \mathbf{D} and \mathbf{D}^\dagger constant terms cancel, and we are left with terms involving the difference of derivatives of \mathbf{f} or C up to second order at nearby points. By the smoothness assumption, these second derivatives are uniformly continuous on the closure of \mathcal{O} and on a ball around $\mathbf{x}^*(1)$. Utilizing (2.15), it follows that for r sufficiently small,

$$\|\nabla\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) - \nabla\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)\|_{\mathcal{Y}} \leq \varepsilon$$

whenever

$$(6.12) \quad \max\{\|\mathbf{X} - \mathbf{X}^*\|_\infty, \|\mathbf{U} - \mathbf{U}^*\|_\infty, \|\mathbf{\Lambda} - \mathbf{\Lambda}^*\|_\infty\} \leq r.$$

Since the smoothness $\eta \geq 2$ in Theorem 1.1, let us choose $\eta = 2$ in Lemma 4.1 and then take \bar{N} large enough that $\|\mathcal{T}(\mathbf{X}^*, \mathbf{U}^*, \mathbf{\Lambda}^*)\|_{\mathcal{Y}} \leq (1 - \mu\varepsilon)r/\mu$ for all $N \geq \bar{N}$. Hence, by Proposition 2.2, there exists a solution to $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$ satisfying (6.12). Moreover, by (2.14) and (4.1), the estimate (1.9) holds. We can use exactly the same argument given in [26] to show that this solution to the first-order condition $\mathcal{T}(\mathbf{X}, \mathbf{U}, \mathbf{\Lambda}) \in \mathcal{F}(\mathbf{U})$ is a local minimizer of (1.2) or equivalently, of (2.1).

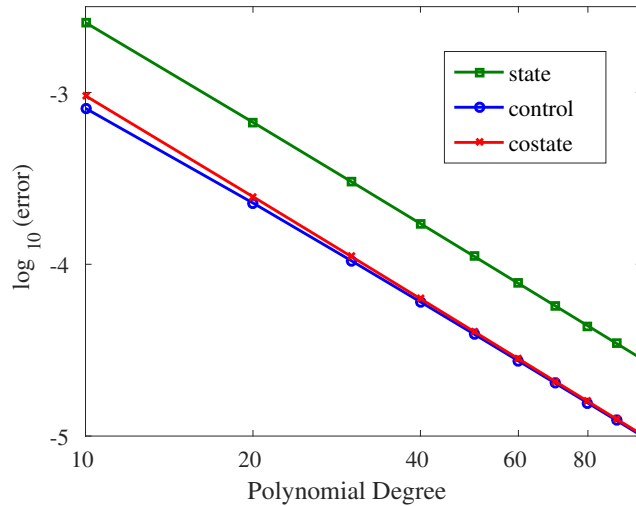


FIG. 1. Logarithm of sup-norm error in state, control, and costate versus polynomial degree.

7. Numerical experiments. We consider the problem from [27] given by

$$(7.1) \quad \begin{aligned} & \text{minimize} && \frac{1}{2} \int_0^1 [x^2(t) + u^2(t)] dt \\ & \text{subject to} && \dot{x}(t) = u(t), \quad u(t) \leq 1, \quad t \in \Omega, \quad x(0) = \frac{1+3e}{2(1-e)}. \end{aligned}$$

The optimal state and control are

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

The associated costate is the integral of the state from t to 1. Since the objective of the test problem is quadratic and the constraints are linear equalities and inequalities, the discrete problem (2.1) is a quadratic programming problem, which we solved using the MATLAB routine QUADPROG. In Figure 1, we plot in base 10 the logarithm of the sup-norm error in the state, control, and costate versus the logarithm of the degree of the polynomial in the discrete problem. Since the optimal state has a discontinuity in its second derivative at $t = 1/2$, x^* lies in $H^2([0, 1])$ as well as in the fractional Sobolev space $H^{2.5-\epsilon}([0, 1])$ for any $\epsilon > 0$. Theorem 1.1 implies that the error is $O(N^{\epsilon-1})$. On the other hand, the observed convergence rate in Figure 1 is $O(N^{-2})$, so the error bound given in Theorem 1.1 is not tight, at least for this particular test problem.

8. Conclusions. An estimate is obtained for the sup-norm error in an approximation to a control constrained variational problem where the state is approximated by a polynomials of degree N and the dynamics is enforced at the Gauss quadrature points. The error is bounded by $(c/N)^{p-3/2}$ times the \mathcal{H}^p norms of the state and costate, where p is the minimum of $N + 1$ and the smoothness η ; it is assumed

that $\eta \geq 2$. In [26], an unconstrained control problem was considered and the corresponding bound was $(c/N)^{p-3}$ with $\eta \geq 4$. The new work advances the convergence theory by significantly improving the exponent in the convergence rate, by relaxing the smoothness requirement, and by including control constraints. When control constraints are present, η is often at most 2, so the relaxation in the smoothness condition is needed to treat control constrained problems. When control constraints are introduced, the first-order optimality conditions lead to a variational inequality, and the analysis centers on the stability of the linearized variational problem under perturbations. The improvements in the convergence theory were achieved by analyzing the effect of perturbations in an $L^2(\Omega)$ setting rather in $L^\infty(\Omega)$, and by analyzing interpolation errors in the Sobolev space $\mathcal{H}^p(\Omega)$ rather than in $L^\infty(\Omega)$. A numerical example indicates that further tightening of the convergence theory may be possible.

9. Appendix 1: Proof of (P1). Let $p \in \mathcal{P}_N$ be any polynomial for which $p(-1) = 0$ and let \mathbf{p} and $\dot{\mathbf{p}} \in \mathbb{R}^N$ denote the vectors with components $p(\tau_i)$ and $\dot{p}(\tau_i)$, respectively, $1 \leq i \leq N$. Since $p(-1) = 0$, the differentiation matrix \mathbf{D} satisfies $\mathbf{D}_{1:N} \mathbf{p} = \dot{\mathbf{p}}$, or equivalently, $\mathbf{D}_{1:N}^{-1} \dot{\mathbf{p}} = \mathbf{p}$. Let \mathbf{r}^T denote the j th row of $\mathbf{D}_{1:N}^{-1}$ for any j between 1 and N , and let $\dot{\mathbf{p}}$ have components $+1$ or -1 , where the sign is chosen so that

$$\mathbf{r}^T \dot{\mathbf{p}} = \sum_{i=1}^N |r_i|.$$

Due to the identity $\mathbf{D}_{1:N}^{-1} \dot{\mathbf{p}} = \mathbf{p}$, we conclude that

$$\sum_{i=1}^N |r_i| = p(\tau_j).$$

Hence, (P1) holds if $|p(\tau_j)| \leq 2$ whenever $p \in \mathcal{P}_N$ is a polynomial that satisfies $p(-1) = 0$ and $|\dot{p}(\tau_i)| \leq 1$ for all $1 \leq i \leq N$. We will prove the following stronger result.

PROPOSITION 9.1. *For any $p \in \mathcal{P}_N$ with $p(-1) = 0$ and $|\dot{p}(\tau_i)| \leq 1$ for all $1 \leq i \leq N$, we have $|p(\tau)| \leq 2$ for all $\tau \in [-1, 1]$.*

Proof. Let l_i , $1 \leq i \leq N$, be the Lagrange interpolating polynomials defined by

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

Let $p \in \mathcal{P}_N$ be any polynomial with $p(-1) = 0$ and $|\dot{p}(\tau_i)| \leq 1$ for all $1 \leq i \leq N$. Since $\dot{p} \in \mathcal{P}_{N-1}$, we can write

$$\dot{p}(\tau) = \sum_{i=1}^N \dot{p}(\tau_i) l_i(\tau).$$

Since $|\dot{p}(\tau_i)| \leq 1$, it follows that

$$(9.1) \quad |p(t)| = \left| \int_{-1}^t \dot{p}(\tau) d\tau \right| = \left| \sum_{i=1}^N \dot{p}(\tau_i) \int_{-1}^t l_i(\tau) d\tau \right| \leq \sum_{i=1}^N \left| \int_{-1}^t l_i(\tau) d\tau \right|.$$

Let $q \in \mathcal{P}_{N-1}$ be defined by

$$q(\tau) = \sum_{i=1}^N a_i l_i(\tau), \quad \text{where } a_i = \begin{cases} 1 & \text{if } \int_{-1}^t l_i(\tau) d\tau > 0, \\ -1 & \text{otherwise.} \end{cases}$$

Hence, we have

$$(9.2) \quad \sum_{i=1}^N \left| \int_{-1}^t l_i(\tau) d\tau \right| = \sum_{i=1}^N a_i \int_{-1}^t l_i(\tau) d\tau = \int_{-1}^t q(\tau) d\tau.$$

Since $|q(\tau_i)| = |a_i| = 1$ for each i , it follows that $q^2(\tau) - 1$ vanishes at $\tau = \tau_i$, $1 \leq i \leq N$. Since $q^2 \in \mathcal{P}_{2N-2}$, we have the factorization

$$(9.3) \quad q^2(\tau) - 1 = r(\tau)P_N(\tau),$$

where $r \in \mathcal{P}_{N-2}$ and P_N is the Legendre polynomial of degree N . Since P_N is orthogonal to polynomials of degree at most $N - 1$, the integral of (9.3) yields the identity

$$\int_{-1}^1 q^2(\tau) d\tau = 2.$$

By the Schwarz inequality,

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

Combine this with (9.1) and (9.2) to obtain

$$|p(t)| \leq \int_{-1}^1 |q(\tau)| d\tau \leq 2,$$

which completes the proof. □

Although this paper has focused on the Gauss abscissa, Proposition 9.1 holds when the Gauss abscissa are replaced by the Radau abscissa.

COROLLARY 9.2. *If τ_i , $1 \leq i \leq N$, are the Radau abscissa with $\tau_N = 1$, then for any $p \in \mathcal{P}_N$ with $p(-1) = 0$ and $|\dot{p}(\tau_i)| \leq 1$ for all $1 \leq i \leq N$, we have $|p(\tau)| \leq 2$ for all $\tau \in [-1, 1]$.*

Proof. Recall that the interior Radau abscissa τ_i , $1 \leq i \leq N - 1$, are the roots of the Jacobi polynomial $P_{N-1}^{(1,0)}$ associated with the weight function $1 - \tau$. The proof of the corollary is identical to the proof of Proposition 9.1 until (9.3), which is replaced by

$$(9.4) \quad q^2(\tau) - 1 = r(\tau)P_{N-1}^{(1,0)}(\tau)(\tau_N - \tau),$$

where $r \in \mathcal{P}_{N-2}$. Since $P_{N-1}^{(1,0)}$ is orthogonal to all polynomials in \mathcal{P}_{N-2} with respect to the weight function $1 - \tau = \tau_N - \tau$, the integral of (9.4) again yields the identity

$$\int_{-1}^1 q^2(\tau) d\tau = 2.$$

The remainder of the proof is identical to that of Proposition 9.1. □

Remark 9.1. The polynomial $p(\tau) = 1 + \tau$ satisfies the conditions of Proposition 9.1 and Corollary 9.2, and $p(1) = 2$. Hence, the upper bound 2 is tight.

Remark 9.2. For the Radau abscissa with $\tau_1 = -1$ and $\tau_N < 1$, the condition $p(-1) = 0$ in the statement of Corollary 9.2 should be replaced by $p(1) = 0$.

10. Appendix 2: \mathcal{L}^2 approximation with a singular weight by Maday. In (3.9) we integrate the error $u - \pi_N u$ in best $\mathcal{H}^1(\Omega)$ approximation using a singular weight $1/(1 - \tau^2)$. Here we relate this singular integral of the error to the error in the $\mathcal{H}_0^1(\Omega)$ norm.

PROPOSITION 10.1. *If $u \in \mathcal{H}_0^1(\Omega)$, then*

$$(10.1) \quad \|u - \pi_N u\|_0 \leq N^{-1} |u - \pi_N u|_{\mathcal{H}^1(\Omega)}, \quad \text{where } \|v\|_0 = \left(\int_{\Omega} \frac{v^2(\tau)}{1 - \tau^2} d\tau \right)^{1/2},$$

and π_N is the projection into \mathcal{P}_N^0 relative to the the norm $|\cdot|_{\mathcal{H}^1(\Omega)}$.

Proof. Let $\langle \cdot, \cdot \rangle_1$ denote the standard $\mathcal{H}_0^1(\Omega)$ inner product defined by

$$\langle u, v \rangle_1 = \int_{\Omega} u'(\tau)v'(\tau) d\tau.$$

By the Legendre equation, the polynomials $\psi_k(\tau) := (1 - \tau^2)P_k'(\tau)$ are orthogonal with respect to the $\mathcal{H}_0^1(\Omega)$ inner product and

$$\langle \psi_k, \psi_k \rangle_1 = \langle (1 - \tau^2) P_k', (1 - \tau^2) P_k' \rangle_1 = k^2(k + 1)^2 \langle P_k, P_k \rangle_{\mathcal{L}^2(\Omega)} = \frac{2k^2(k + 1)^2}{2k + 1}.$$

Consequently, $\{\psi_k : 1 \leq k \leq N - 1\}$ is an orthogonal basis for \mathcal{P}_N^0 , and the orthogonal projection of u into \mathcal{P}_N^0 is given by

$$\pi_N u = \sum_{k=1}^{N-1} u_k \psi_k, \quad u_k = \frac{\langle u, \psi_k \rangle_1}{\langle \psi_k, \psi_k \rangle_1}.$$

Let $\langle \cdot, \cdot \rangle_0$ denote the inner product on $\mathcal{H}_0^1(\Omega)$ defined by

$$\langle u, v \rangle_0 = \int_{\Omega} \frac{u(\tau)v(\tau)}{1 - \tau^2} d\tau.$$

By the Schwarz and Hardy inequalities, $\|u\|_0^2 \leq 2|u|_{\mathcal{H}^1(\Omega)}\|u\|_{\mathcal{L}^2(\Omega)}$. By the Legendre equation, the ψ_k are also orthogonal in the $\langle \cdot, \cdot \rangle_0$ inner product and

$$\begin{aligned} \langle \psi_k, \psi_k \rangle_0 &= \langle (1 - \tau^2) P_k', (1 - \tau^2) P_k' \rangle_0 = \langle (1 - \tau^2) P_k', P_k' \rangle_{\mathcal{L}^2(\Omega)} \\ &= k(k + 1) \langle P_k, P_k \rangle_{\mathcal{L}^2(\Omega)} = \frac{2k(k + 1)}{2k + 1}. \end{aligned}$$

Due to orthogonality, we have

$$\begin{aligned} \|u - \pi_N u\|_0^2 &= \sum_{k \geq N} u_k^2 \langle \psi_k, \psi_k \rangle_0 = \sum_{k \geq N} \left(\frac{2k(k + 1)}{2k + 1} \right) u_k^2, \\ |u - \pi_N u|_{\mathcal{H}^1(\Omega)}^2 &= \sum_{k \geq N} u_k^2 \langle \psi_k, \psi_k \rangle_1 = \sum_{k \geq N} \left(\frac{2k^2(k + 1)^2}{2k + 1} \right) u_k^2. \end{aligned}$$

Comparing these norms, we see that (10.1) holds. □

Acknowledgments. We thank the reviewers for their careful reading of the manuscript and their constructive comments and suggestions. In particular, one reviewer suggested a better arrangement for the proof of Lemma 3.1. The authors deeply appreciate Yvon Maday's contribution of Proposition 10.1, a key step in Lemma 3.1.

REFERENCES

- [1] R. BELLMAN, *A note on an inequality of E. Schmidt*, Bull. Amer. Math. Soc., 50 (1944), pp. 734–737.
- [2] D. A. BENSON, G. T. HUNTINGTON, T. P. THORVALDSEN, AND A. V. RAO, *Direct trajectory optimization and costate estimation via an orthogonal collocation method*, J. Guid. Control Dyn., 29 (2006), pp. 1435–1440.
- [3] C. BERNARDI AND Y. MADAY, *Polynomial interpolation results in Sobolev spaces*, J. Comput. Appl. Math., 43 (1992), pp. 53–82.
- [4] J. F. BONNANS, *Lipschitz solutions of optimal control problems with state constraints of arbitrary order*, Ann. Acad. Rom. Sci. Ser. Math. Appl., 2 (2010), pp. 78–98.
- [5] A. DONTCHEV, W. W. HAGER, A. POORE, AND B. YANG, *Optimality, stability, and convergence in nonlinear control*, Appl. Math. Optim., 31 (1995), pp. 297–326.
- [6] A. L. DONTCHEV AND W. W. HAGER, *Lipschitzian stability in nonlinear control and optimization*, SIAM J. Control Optim., 31 (1993), pp. 569–603.
- [7] A. L. DONTCHEV AND W. W. HAGER, *A new approach to Lipschitz continuity in state constrained optimal control*, Systems Control Lett., 35 (1998), pp. 137–143.
- [8] A. L. DONTCHEV AND W. W. HAGER, *The Euler approximation in state constrained optimal control*, Math. Comp., 70 (2001), pp. 173–203.
- [9] A. L. DONTCHEV, W. W. HAGER, AND K. MALANOWSKI, *Error bounds for Euler approximation of a state and control constrained optimal control problem*, Numer. Funct. Anal. Optim., 21 (2000), pp. 653–682.
- [10] A. L. DONTCHEV, W. W. HAGER, AND V. M. VELIOV, *Second-order Runge–Kutta approximations in constrained optimal control*, SIAM J. Numer. Anal., 38 (2000), pp. 202–226.
- [11] G. ELNAGAR, M. KAZEMI, AND M. RAZZAGHI, *The pseudospectral Legendre method for discretizing optimal control problems*, IEEE Trans. Automat. Control, 40 (1995), pp. 1793–1796.
- [12] G. N. ELNAGAR AND M. A. KAZEMI, *Pseudospectral Chebyshev optimal control of constrained nonlinear dynamical systems*, Comput. Optim. Appl., 11 (1998), pp. 195–217.
- [13] J. ELSCHNER, *The h - p -version of spline approximation methods for Melin convolution equations*, J. Integral Equations Appl., 5 (1993), pp. 47–73.
- [14] F. FAHROO AND I. M. ROSS, *Costate estimation by a Legendre pseudospectral method*, J. Guid. Control Dyn., 24 (2001), pp. 270–277.
- [15] F. FAHROO AND I. M. ROSS, *Direct trajectory optimization by a Chebyshev pseudospectral method*, J. Guid. Control Dyn., 25 (2002), pp. 160–166.
- [16] F. FAHROO AND I. M. ROSS, *Pseudospectral methods for infinite-horizon nonlinear optimal control problems*, J. Guid. Control Dyn., 31 (2008), pp. 927–936.
- [17] D. GARG, M. A. PATTERSON, C. L. DARBY, C. FRANÇOLIN, G. T. HUNTINGTON, W. W. HAGER, AND A. V. RAO, *Direct trajectory optimization and costate estimation of finite-horizon and infinite-horizon optimal control problems using a Radau pseudospectral method*, Comput. Optim. Appl., 49 (2011), pp. 335–358.
- [18] D. GARG, M. A. PATTERSON, W. W. HAGER, A. V. RAO, D. A. BENSON, AND G. T. HUNTINGTON, *A unified framework for the numerical solution of optimal control problems using pseudospectral methods*, Automatica, 46 (2010), pp. 1843–1851.
- [19] Q. GONG, I. M. ROSS, W. KANG, AND F. FAHROO, *Connections between the covector mapping theorem and convergence of pseudospectral methods for optimal control*, Comput. Optim. Appl., 41 (2008), pp. 307–335.
- [20] W. W. HAGER, *Lipschitz continuity for constrained processes*, SIAM J. Control Optim., 17 (1979), pp. 321–337.
- [21] W. W. HAGER, *Multiplier methods for nonlinear optimal control*, SIAM J. Numer. Anal., 27 (1990), pp. 1061–1080.
- [22] W. W. HAGER, *Runge–Kutta methods in optimal control and the transformed adjoint system*, Numer. Math., 87 (2000), pp. 247–282.
- [23] W. W. HAGER, *Numerical analysis in optimal control*, in International Series of Numerical Mathematics, Vol. 139, K.-H. Hoffmann, I. Lasiecka, G. Leugering, J. Sprekels, and F. Tröltzsch, eds., Birkhäuser-Verlag, Basel, Switzerland, 2001, pp. 83–93.

- [24] W. W. HAGER, H. HOU, S. MOHAPATRA, AND A. V. RAO, *Convergence Rate for an hp-Collocation Method Applied to Constrained Optimal Control*, preprint, arXiv:1605.02121, 2016.
- [25] W. W. HAGER, H. HOU, AND A. V. RAO, *Convergence Rate for a Radau Collocation Method Applied to Unconstrained Optimal Control*, preprint, arXiv:1508.03783, 2015.
- [26] W. W. HAGER, H. HOU, AND A. V. RAO, *Convergence rate for a Gauss collocation method applied to unconstrained optimal control*, *J. Optim. Theory Appl.*, 169 (2016), pp. 801–824.
- [27] W. W. HAGER AND G. IANULESCU, *Dual approximations in optimal control*, *SIAM J. Control Optim.*, 22 (1984), pp. 423–465.
- [28] A. HERMANT, *Stability analysis of optimal control problems with a second-order state constraint*, *SIAM J. Optim.*, 22 (2009), pp. 104–129.
- [29] R. A. HORN AND C. R. JOHNSON, *Matrix Analysis*, Cambridge University Press, Cambridge, 2013.
- [30] S. KAMESWARAN AND L. T. BIEGLER, *Convergence rates for direct transcription of optimal control problems using collocation at Radau points*, *Comput. Optim. Appl.*, 41 (2008), pp. 81–126.
- [31] W. KANG, *The rate of convergence for a pseudospectral optimal control method*, in *Proceedings of the 47th IEEE Conference on Decision and Control*, IEEE, 2008, pp. 521–527.
- [32] W. KANG, *Rate of convergence for the Legendre pseudospectral optimal control of feedback linearizable systems*, *J. Control Theory Appl.*, 8 (2010), pp. 391–405.
- [33] F. LIU, W. W. HAGER, AND A. V. RAO, *Adaptive mesh refinement method for optimal control using nonsmoothness detection and mesh size reduction*, *J. Franklin Inst.*, 352 (2015), pp. 4081–4106.
- [34] V. A. MARKOV, *Über Polynome, die in einem gegebenen Intervalle möglichst wenig von Null abweichen*, *Math. Ann.*, 77 (1916), pp. 185–191.
- [35] J. NOCEDAL AND S. J. WRIGHT, *Numerical Optimization*, 2nd ed., Springer, New York, 2006.
- [36] M. A. PATTERSON, W. W. HAGER, AND A. V. RAO, *A ph mesh refinement method for optimal control*, *Optim. Control Appl. Meth.*, 36 (2015), pp. 398–421.
- [37] G. W. REDDIEN, *Collocation at Gauss points as a discretization in optimal control*, *SIAM J. Control Optim.*, 17 (1979), pp. 298–306.
- [38] J. STOER AND R. BULIRSCH, *Introduction to Numerical Analysis*, 3rd ed., Springer, New York, 2002.
- [39] P. WILLIAMS, *Jacobi pseudospectral method for solving optimal control problems*, *J. Guid. Control Dyn.*, 27 (2004), pp. 293–297.