Proceedings of the
47th IEEE Conference on Decision and Control
Cancun, Mexico, Dec. 9-11, 2008

The Rate of Convergence for a Pseudospectral Optimal Control Method

Wei Kang

Abstract— **Over the last decade, pseudospectral (PS) methods have emerged as a popular computational solution for the problem of nonlinear constrained optimal control. They have been applied to many industrial-strength problems, notably the recent zero-propellant-maneuvering of the International Space Station performed by NASA. In this paper, we prove a theorem on the rate of convergence for the approximate optimal cost computed using PS methods. This paper contains several essential differences from existing papers on PS optimal control as well as some other direct computational methods. First, the proofs do not use necessary conditions of optimal control. Secondly, we do not make coercivity type of assumptions. As a result, the theory does not require the local uniqueness of optimal solutions. The proof is not build on the bases of consistency approximation theory. Thus, we can remove some restrictive assumptions in the previous results on PS optimal control methods.**

I. INTRODUCTION

As a result of significant progress in large-scale computational algorithms and nonlinear programming, the so-called direct computational methods have become popular for solving nonlinear optimal control problems [1], [2], [3], particularly in aerospace applications [4], [5]. In simple terms, in a direct method, the continuous-time problem of optimal control is discretized, and the resulting discretized optimization problem is solved by nonlinear programming algorithms. Over the last decade, pseudospectral (PS) techniques have emerged as a popular direct method for optimal control. They have been applied to many industrial-strength problems, notably the recent attitude maneuvers of the International Space Station (ISS) performed by NASA. By following an attitude trajectory developed using PS optimal control, ISS was maneuvered 180 degrees on March 3, 2007, by using the gyroscopes equipped on the ISS without propellant consumption. This single maneuver have saved NASA about one million dollars' worth of fuel [6]. The Legendre PS optimal control method has already been developed into software named DIDO, a MATLAB based package commercially available [7]. In addition, the next generation of the OTIS software package [8] will have the Legendre PS method as a problem solving option.

For the last decade, active research has been carried out in the effort of developing a theoretical foundation for PS optimal control methods. Among the research focuses, there are three fundamental issues, namely the state and costate approximation, the existence and convergence of approximate solutions, and the convergence rate. The general importance of these issues is not limited to PS optimal control. They are essential to other computational optimal control methods suchlike those based on Euler [11] and Runge-Kutta [12] discretization. For PS methods, a covector mapping was established and proved in [13] and [14]; several theorems on the existence and convergence were published in [15] and [16]; and then the results were generalized in [17] to problems with non-smooth control.

In this paper, we prove a rate of convergence for the approximate optimal cost computed using PS methods, which is a first result on the rate of convergence proved for PS optimal control. This paper contains several essential differences from existing papers on PS optimal control as well as some other direct computational methods. First, the proof does not use necessary conditions of optimal control. Furthermore, we do not make coercivity type of assumptions. As a result, the theory does not require the local uniqueness of optimal solutions. Therefore, it is applicable to problems with multiple optimal solutions that exist in a small neighborhood. Secondly, the proof is not build on the bases of consistent approximation theory [3]. Thus, we can remove the assumption in [15] and [17] on the existence of cluster points for the derivatives of discrete solutions. The key that makes these differences possible is that we regulate the region in which a solution is being searched. This regulation does not add restrictions to the original control problem. It is different from existing approaches in which the desired convergence is achieved by making harsh assumptions on the original control problem. However, the approach in the current paper does require that the system is feedback linearizable. A proof for general control systems is one of the tasks for future research.

II. PROBLEM FORMULATION

In this paper, we address the following Bolza problem of control systems in the feedback linearizable normal form.

Problem B: Determine the state-control function pair $(x(t), u(t))$, $x \in \Re^r$ and $u \in \Re$, that minimizes the cost function

$$J(x(\cdot), u(\cdot)) = \int_{-1}^{1} F(x(t), u(t))\ dt + E(x(-1), x(1)) \quad (1)$$

subject to the state equation

$$\begin{cases} \dot{x}_1 = x_2, \cdots, \dot{x}_{r-1} = x_r \\ \dot{x}_r = f(x) + g(x)u \end{cases} \quad (2)$$

$$x(-1) = x_0 \quad (3)$$

where $x \in \Re^r$, $u \in \Re$, and $F : \Re^r \times \Re \to \Re$, $E : \Re^r \times \Re^r \to \Re$, $f : \Re^r \to \Re$, and $g : \Re^r \to \Re$ are all Lipschitz continuous

This work was supported in part by AFOSR and AFRL.

Wei Kang is with the Department of Applied Mathematics, Naval Postgraduate School, Monterey, CA 93943, USA wkang@nps.edu

978-1-4244-3124-3/08/$25.00 ©2008 IEEE

functions with respect to their arguments. In addition, we assume $g(x) \neq 0$ for all x. In this paper, we only consider the problems that have at least one optimal solution in which $x_r^*(t)$ has bounded m-th order weak derivative. For some results, we assume $m \geq 3$. For others, m is smaller. Unless the term 'strong derivative' is emphasized, all derivatives in the paper are in the weak sense.

In typical direct methods, the original optimal control problem, not the associated necessary conditions, is discretized to formulate a nonlinear programming problem. Given any function $f(t) : [a, b] \rightarrow \Re$, a conventional method of approximation is to interpolate at uniformly spaced nodes: $t_0 = a$, $t_1 = (b - a)/N$, $\cdots$, $t_N = b$. However, uniform spacing is not efficient in approximation. More sophisticated node selection methods are able to achieve significantly improved accuracy with fewer nodes. In a PS approximation, a function $f(t)$ is approximated by N-th order Lagrange polynomials based on the interpolation at the Legendre-Gauss-Lobatto (LGL) quadrature nodes. The LGL nodes, $t_0 = -1 < t_1 < \cdots < t_N = 1$, are defined by

$$\begin{array}{l} t_0 = -1, \ \ t_N = 1, \text{ and} \\ \text{for } k = 1, 2, \ldots, N-1, t_k \text{ are the roots of } \dot{L}_N(t) \end{array}$$

where $\dot{L}_N(t)$ is the derivative of the N-th order Legendre polynomial $L_N(t)$. The discretization works in the interval of $[-1, 1]$. It was proved in approximation theory that the polynomial interpolation at the LGL nodes converges to $f(t)$ under L^2 norm at the rate of $1/N^m$, where m is the smoothness of $f(t)$. If $f(t)$ is C^∞, then the polynomial interpolation at the LGL nodes converges at a spectral rate, i.e. it is faster than any given polynomial rate. This is a very impressive convergence rate.

In a PS optimal control method, the state and control functions, $x(t)$ and $u(t)$, are approximated by N-th order Lagrange polynomials based on the interpolation at the LGL quadrature nodes. In the discretization, the state variables are approximated by the vectors $\bar{x}^{Nk} \in \Re^r$, i.e.

$$\bar{x}^{Nk} = \left[\begin{array}{cccc} \bar{x}_1^{Nk} & \bar{x}_2^{Nk} & \cdots & \bar{x}_r^{Nk} \end{array} \right]^T$$

is an approximation of $x(t_k)$. Similarly, $\bar{u}^{Nk}$ is the approximation of $u(t_k)$. Thus, a discrete approximation of the function $x_i(t)$ is the vector

$$\bar{x}_i^N = \left[\begin{array}{cccc} \bar{x}_i^{N1} & \bar{x}_i^{N2} & \cdots & \bar{x}_i^{NN} \end{array} \right]$$

A continuous approximation is defined by its polynomial interpolation, denoted by $x_i^N(t)$, i.e.

$$x_i(t) \ \approx \ x_i^N(t) = \sum_{k=0}^{N} \bar{x}_i^{Nk} \phi_k(t), \tag{4}$$

where $\phi_k(t)$ is the Lagrange interpolating polynomial [9]. Instead of polynomial interpolation, the control input is approximated by the following non-polynomial interpolation

$$u^N(t) = \frac{\dot{x}_r^N(t) - f(x^N(t))}{g(x^N(t))} \tag{5}$$

In the notations, the discrete variables are denoted by letters with an upper bar, such as $\bar{x}_i^{Nk}$ and $\bar{u}^{Nk}$. If k in the superscript and/or i in the subscript are missing, it represents the corresponding vector or matrix in which the indices run from minimum to maximum. For example,

$$\begin{array}{rcl} \bar{x}^N & = & \left[\begin{array}{cccc} \bar{x}_1^{N0} & \bar{x}_1^{N1} & \cdots & \bar{x}_1^{NN} \\ \bar{x}_2^{N0} & \bar{x}_2^{N1} & \cdots & \bar{x}_2^{NN} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{x}_r^{N0} & \bar{x}_r^{N1} & \cdots & \bar{x}_r^{NN} \end{array} \right] \\ \bar{u}^N & = & \left[\begin{array}{cccc} \bar{u}^{N0} & \bar{u}^{N1} & \cdots & \bar{u}^{NN} \end{array} \right] \end{array}$$

Given a discrete approximation of a continuous function, the interpolation is denoted by the same notation without the upper bar. For example, $x_i^N(t)$ in (4), $u^N(t)$ in (5). The superscript N represents the number of LGL nodes used in the approximation. Throughout this paper, the interpolation of $(\bar{x}^N, \bar{u}^N)$ is defined by (4)-(5), in which $u^N(t)$ is not necessarily a polynomial. It is proved in Lemma 5 that (5) is indeed an interpolation.

For differentiation, the derivative of $x_i^N(t)$ at the LGL node t_k is computed using matrix multiplications [9]

$$\left[\begin{array}{cccc} \dot{x}_i^N(t_0) & \dot{x}_i^N(t_1) & \cdots & \dot{x}_i^N(t_N) \end{array} \right]^T = D(\bar{x}_i^N)^T \tag{6}$$

where D is the $(N + 1) \times (N + 1)$ differentiation matrix, which depends on N and can be computed off-line ([9]). The cost functional $J[x(\cdot), u(\cdot)]$ is approximated by the Gauss-Lobatto integration rule,

$$\bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{k=0}^{N} F(\bar{x}^{Nk}, \bar{u}^{Nk}) w_k + E(\bar{x}^{N0}, \bar{x}^{NN})$$

where w_k are the LGL weights[9]. The approximation is so accurate that it has zero error if the integrand function is a polynomial of degree less than or equal to $2N - 1$, a degree that almost double the number of nodes [9]. Now, we are ready to define a PS discretization of Problem B.

For any integer $m_1 > 0$, let $\{a_0^N(m_1), a_1^N(m_1), \cdots, a_{N-r-m_1+1}^N(m_1)\}$ denote the sequence of spectral coefficients (the coefficients in the Legendre series) for the interpolation polynomial of the vector $\bar{x}_r^N(D^T)^{m_1}$. Note that the interpolation of $\bar{x}_r^N(D^T)^{m_1}$ equals the polynomial of $\frac{d^{m_1} x_r^N(t)}{dt^{m_1}}$. Thus, there are only $N - r - m_1 + 2$ nonzero spectral coefficients because it can be proved that the order of $\frac{d^{m_1} x_r^N(t)}{dt^{m_1}}$ is at most degree of $N - r - m_1 + 1$. They depend linearly on $\bar{x}_r^N$ [18]. The PS discretization of Problem $\text{B}^\text{N}$is defined as follows.

Problem B^N Find $\bar{x}^{Nk} \in \Re^r$ and $\bar{u}^{Nk} \in \Re$, $k = 0, 1, \ldots, N$, that minimize

$$\bar{J}^N(\bar{x}^N, \bar{u}^N) = \sum_{k=0}^{N} F(\bar{x}^{Nk}, \bar{u}^{Nk}) w_k + E(\bar{x}^{N0}, \bar{x}^{NN}) \tag{7}$$

subject to

$$\left\{\begin{array}{rcl} D(\bar{x}_1^N)^T & = & (\bar{x}_2^N)^T \\ D(\bar{x}_2^N)^T & = & (\bar{x}_3^N)^T \\ & \vdots & \\ D(\bar{x}_{r-1}^N)^T & = & (\bar{x}_r^N)^T \\ D(\bar{x}_r^N)^T & = & \left[\begin{array}{c} f(\bar{x}^{N0})+g(\bar{x}^{N0})\bar{u}^{N0} \\ \vdots \\ f(\bar{x}^{NN})+g(\bar{x}^{NN})\bar{u}^{NN} \end{array}\right] \end{array}\right. \quad (8)$$

$$\bar{x}^{N0} = x_0 \quad (9)$$

$$\underline{\boldsymbol{b}} \leq \left[\begin{array}{c} \bar{x}^{Nk} \\ \bar{u}^{Nk} \end{array}\right] \leq \bar{\boldsymbol{b}}, \ \ 0 \leq k \leq N \quad (10)$$

$$\underline{\boldsymbol{b}}_j \leq \left[\begin{array}{cccc} 1 & 0 & \cdots & 0 \end{array}\right] D^j(\bar{x}_r^N)^T \leq \bar{\boldsymbol{b}}_j \quad (11)$$
$$1 \leq j \leq m_1 - 1, m_1 \geq 2$$

$$\sum_{n=0}^{N-r-m_1+1} |a_n^N(m_1)| \leq \boldsymbol{d} \quad (12)$$

Comparing to Problem B, (8) is the discretization of the control system defined by the differential equation. The constraint (11) assures that the derivative of the interpolation up to the order of m_1 is bounded. It is proved in the following sections that the integer m_1 is closely related to the convergence rate. The inequalities (10), (11) and (12) are additional constraints that do not exist in Problem B. It can be proved that these additional constraints do not affect the feasibility of Problem $\mathrm{B^N}$. It is important to emphasize that these constraints do not impose assumptions or limitations to Problem B. They simply regulate the region in which an approximate solution is being searched. The family of problems that can be solved is not limited by these constraints. These additional constraints are necessary for several reasons. In practical computation, nonlinear programming solvers always require a bounded region (10) for the searching of an optimal solution. The constraints (11) and (12) are necessary to avoid the restrictive consistent approximation assumption made in [15]. At a more fundamental level, the order of derivatives, m_1 in (11) and (12), determines the convergence rate of the approximate optimal control. Another interesting fact that amply justify these additional constraints is that Problem $\mathrm{B^N}$ may not even have an optimal solution if we do not enforce (10). This is shown by the following counter example.

Example 1: Consider the following problem of optimal control.

$$\min_{(x(\cdot),u(\cdot))} \int_{-1}^{1} \frac{(x(t)-u(t))^2}{u(t)^4} dt$$
$$\dot{x} = u \quad (13)$$
$$x(-1) = e^{-1}$$

It is easy to check that the optimal solution is $u = e^t, \ x(t) = e^t$ and the optimal cost value is zero. Although the solution to the problem (13) is simple and analytic, the PS discretization of (13) does not have an optimal solution if the constraint (10) is not enforced. To prove this claim, consider the PS discretization,

$$\min_{(\bar{x}^N,\bar{u}^N)} \bar{J}^N(\bar{x}^N,\bar{u}^N) = \sum_{k=0}^{N} \frac{(\bar{x}^{Nk} - D_k(\bar{x}^N)^T)^2}{(D_k(\bar{x}^N)^T)^4} w_k$$
$$D(\bar{x}^N)^T = (\bar{u}^N)^T \quad (14)$$
$$\bar{x}^{N0} = e^{-1}$$

where D_k is the kth row of the differentiation matrix D. Let $x^N(t)$ be the interpolation polynomial of $\bar{x}^N$, then it is obvious that $\bar{x}^N(t) - \dot{\bar{x}}^N(t) \not\equiv 0$. Thus, there exists k so that $\bar{x}^{Nk} - D_k(\bar{x}^N)^T \neq 0$. So,

$$\bar{J}^N(\bar{x}^N,\bar{u}^N) > 0 \quad (15)$$

for all feasible pairs $(\bar{x}^N,\bar{u}^N)$. For any $\alpha > 0$, define $\bar{x}^{Nk} = e^{-1} + \alpha(t_k+1)$. The interpolation of $\bar{x}^N$ is the linear polynomial $x^N(t) = e^{-1} + \alpha(t+1)$. Then,

$$D_k(\bar{x}^N)^T = \dot{x}^N(t_k) = \alpha$$

The cost function is

$$\begin{array}{rcl} \bar{J}^N(\bar{x}^N,\bar{u}^N) & = & \displaystyle\sum_{k=0}^{N} \frac{(e^{-1}+\alpha t_k)^2}{\alpha^4} w_k \\ & \leq & \displaystyle\sum_{k=0}^{N} \frac{(e^{-1}+\alpha)^2}{\alpha^4} w_k = 2\frac{(e^{-1}+\alpha)^2}{\alpha^4} \end{array}$$

Therefore, $\bar{J}^N(\bar{x}^N,\bar{u}^N)$ can be arbitrarily small as α approaches ∞. However, $\bar{J}^N(\bar{x}^N,\bar{u}^N)$ is always positive as shown by (15). We conclude that the discretization (14) has no minimum value for $\bar{J}^N(\bar{x}^N,\bar{u}^N)$.

III. Rate of Convergence

Problem $\mathrm{B^N}$ has several bounds in its definition, $\underline{\boldsymbol{b}}$, $\bar{\boldsymbol{b}}$, $\underline{\boldsymbol{b}}_j$, $\bar{\boldsymbol{b}}_j$, and $\boldsymbol{d}$. These bounds can be selected from a range determined by Problem B. The constraints $\underline{\boldsymbol{b}}$ and $\bar{\boldsymbol{b}}$ are lower and upper bounds so that the optimal trajectory of Problem B is contained in the interior of the region. Suppose Problem B has an optimal solution $(x^*(t), u^*(t))$ in which $(x_r^*(t))^{(m)}$ has bounded variation for some $m \geq 3$, where $x_r^*(t)$ is the rth component of the optimal trajectory. Suppose m_1 in Problem $\mathrm{B^N}$ satisfies $2 \leq m_1 \leq m-1$. Then, we can select the bounds $\underline{\boldsymbol{b}}_j$ and $\bar{\boldsymbol{b}}_j$ so that $(x_r^*(t))^{(j)}$ is contained in the interior of the region. For $\boldsymbol{d}$, we assume

$$\boldsymbol{d} \ > \ \frac{6}{\sqrt{\pi}}(U(x_r^{*\,(m_1+1)}) + V(x_r^{*\,(m_1+1)}))\zeta(3/2) \quad (16)$$

where $U(x_r^{*\,(m_1+1)})$ is the upper bound and $V(x_r^{*\,(m_1+1)})$ is the total variation of $x_r^{*\,(m_1+1)}(t)$; and $\zeta(s)$ is the ζ function. If all the bounds are selected as above, then it can be proved that Problem $\mathrm{B^N}$ is always feasible provided $m \geq 2$ (the proof is omitted for the reason of space). Note that in practical computation, $\underline{\boldsymbol{b}}$, $\bar{\boldsymbol{b}}$, $\underline{\boldsymbol{b}}_j$, $\bar{\boldsymbol{b}}_j$, and $\boldsymbol{d}$ are unknown. They must be estimated based upon experience or other information about the system. However, we would like to point out that in most engineering problems that we solved using PS methods, the sequence of optimal solutions satisfy all the bounds without specifically enforcing the constraints (11) and (12) in the computation. This is because that the nonlinear programming seeks for optimal solutions with minimum cost that generally rules out unbounded discrete trajectories automatically. The constraint (10) is always enforced in the computation as a required bound for nonlinear programming solvers.

Theorem 1: Suppose Problem B has an optimal solution $(x^*(t), u^*(t))$ in which the strong derivative $(x_r^*(t))^{(m)}$ has a bounded variation for some $m \geq 3$. In Problem B^{N}, select m_1 and α so that $1 \leq m_1 \leq m-1$ and $0 < \alpha < m_1 - 1$. Suppose $f(\cdot)$, $g(\cdot)$, and $F(\cdot)$ are C^m. Suppose all other bounds in Problem B^{N} are large enough. Given any sequence

$$\{(\bar{x}^{*N}, \bar{u}^{*N})\}_{N \geq N_1} \tag{17}$$

of optimal solutions of Problem B^{N}. Then the approximate cost converge to the optimal value at the following rate

$$\left| J(x^*(\cdot), u^*(\cdot)) - J(x^{*N}(\cdot), u^{*N}(\cdot)) \right| \leq \frac{M_1}{N^{2m-2m_1-1}} + \frac{M_2}{N^\alpha} \tag{18}$$

$$\left| J(x^*(\cdot), u^*(\cdot)) - \bar{J}^N(\bar{x}^{*N}, \bar{u}^{*N}) \right| \leq \frac{M_1}{N^{2m-2m_1-1}} + \frac{M_2}{N^\alpha} \tag{19}$$

where M_1 and M_2 are some constants independent of N. In (18), $(x^{*N}(t), u^{*N}(t))$ is the interpolation of (17) defined by (4)-(5). In fact, $x^{*N}(t)$ is the trajectory of (2) under the control input $u^{*N}(t)$.

Remark 3.1: Theorem 1 implies that the costs of any sequence of discrete optimal solutions must converge to the optimal cost of Problem B, no matter the sequence of the discrete state and control trajectories converge or not. In other words, it is possible that the sequence of discrete optimal controls does not converge to a unique continuous-time control; meanwhile the costs using these approximate optimal controls converge to the true optimal cost of Problem B. Therefore, this theorem does not require the local uniqueness of solutions for Problem B.

Remark 3.2: If $f(\cdot)$, $g(\cdot)$, $F(\cdot)$ and $x_r^*(t)$ are C^∞, then we can select m and m_1 arbitrarily large. In this case, we can make the optimal cost of Problem B^{N}converge faster than any given polynomial rate.

The proof is convoluted that involves results from several different areas, including nonlinear functional analysis, orthogonal polynomials, and approximation theory. The following lemma is standard in functional analysis [19].

Lemma 1: Suppose $\mathcal{J}$ takes a local minimum value at u^*. Suppose $\mathcal{J}$ has second order Fréchet derivative at u^*. Then,

$$\mathcal{J}(u^* + \Delta u) = (\mathcal{J}''(u^*)\Delta u)\Delta u + o(||\Delta u||^2)$$

The rate of convergence for the spectral coefficients can be estimated by the following Jackson's Theorem.

Lemma 2: (Jackson's Theorem [20]) Let $h(t)$ be of bounded variation in $[-1, 1]$. Define

$$H(t) = H(-1) + \int_{-1}^{t} h(s)ds$$

then $\{a_n\}_{n=0}^\infty$, the sequence of spectral coefficients of $H(t)$, satisfies the following inequality

$$a_n < \frac{6}{\sqrt{\pi}}(U(h(t)) + V(h(t)))\frac{1}{n^{3/2}}$$

for $n \geq 1$.

Given a continuous function $h(t)$ defined on $[-1, 1]$. Let $\hat{p}^N(t)$ be the best polynomial of degree N, i.e. the Nth order polynomial with the smallest distance to $h(t)$ under $||\cdot||_\infty$ norm. Let $I_N h(t)$ be the polynomial interpolation using the value of $h(t)$ at the LGL nodes. Then,

Lemma 3: ([9] or [10])

$$||h(t) - I_N h||_\infty \leq (1 + \Lambda_N)||h(t) - \hat{p}^N(t)||_\infty$$

where Λ_N is called Lebesgue constant. It satisfies

$$\Lambda_N \leq \frac{2}{\pi}log(N+1) + 0.685\cdots$$

The best polynomial approximation represents the closest polynomial to a function under $||\cdot||_\infty$. The error can be estimated by the following Lemma [9].

Lemma 4: (1) Suppose $h(t) \in W^{m,\infty}$. Let $\hat{p}^N(t)$ be the best polynomial approximation. Then

$$||\hat{p}^N(t) - h(t)||_\infty \leq \frac{C}{N^m}||h(t)||_{W^{m,\infty}}$$

for some constant C independent of $h(t)$, m and N.
(2) If $h(t) \in W^{m,2}$, then

$$||h(t) - P_N h(t)||_\infty \leq \frac{C||h(t)||_{W^{m,2}}}{N^{m-3/4}}$$

where $P_N h$ is the N-th order truncation of the Legendre series of $h(t)$.
(3) If $h(t)$ has the m-th order strong derivative with a bounded variation, then

$$||h(t) - P_N h(t)||_\infty \leq \frac{CV(h^{(m)}(t))}{N^{m-1/2}}$$

The following are lemmas proved specifically for PS optimal control methods.

Lemma 5: (i) For any trajectory, $(\bar{x}^N, \bar{u}^N)$, of the dynamics (8), the pair $(x^N(t), u^N(t))$ defined by (4)-(5) satisfies the differential equations defined in (2). Furthermore,

$$\bar{x}^{Nk} = x^N(t_k),\ \bar{u}^{Nk} = u^N(t_k), k = 0, \cdots, N \tag{20}$$

(ii) For any pair $(x^N(t), u^N(t))$ in which $x^N(t)$ consists of polynomials of degree less than or equal to N and $u^N(t)$ is a function, if $(x^N(t), u^N(t))$ satisfies the differential equations in (2), then $(\bar{x}^N, \bar{u}^N)$ defined by (20) satisfies (8).

(iii) If $(\bar{x}^N, \bar{u}^N)$ satisfies (8), then the degree of $x_i^N(t)$ is less than or equal to $N - i + 1$.

Proof: (i) Suppose $(\bar{x}^N, \bar{u}^N)$ satisfies the equations in (8). Because $x^N(t)$ is the polynomial interpolation of $\bar{x}^N$, and because of equations (6), we have

$$\begin{aligned} &\left[\begin{array}{cccc} \dot{x}_i^N(t_0) & \dot{x}_i^N(t_1) & \cdots & \dot{x}_i^N(t_N) \end{array}\right] = \bar{x}_i^N D^T = \bar{x}_{i+1}^N \\ &= \left[\begin{array}{cccc} x_{i+1}^N(t_0) & x_{i+1}^N(t_1) & \cdots & x_{i+1}^N(t_N) \end{array}\right] \end{aligned}$$

Therefore, the polynomials $\dot{x}_i^N(t)$ and $x_{i+1}^N(t)$ must equal each other because they coincide at $N+1$ points and because the degrees of $x_i^N(t)$ and $x_{i+1}^N(t)$ are less than or equal to N. In addition, (5), the definition of $u^N(t)$, implies the last equation in (2). So, the pair $(x^N(t), u^N(t))$ satisfies all equations in (2). Now, we prove (20). Because $x^N(t)$ is an interpolation of $\bar{x}^N$, we know $\bar{x}^{Nk} = x^N(t_k)$ for $0 \le k \le N$. From (5),

$$u^N(t_k) = \frac{\dot{x}_r^N(t_k) - f(x^N(t_k))}{g(x^N(t_k))} = \frac{\dot{x}_r^N(t_k) - f(\bar{x}^{Nk})}{g(\bar{x}^{Nk})} \tag{21}$$

Because of (6), (21) is equivalent to

$$\begin{bmatrix} u^N(t_0) & u^N(t_1) & \cdots & u^N(t_N) \end{bmatrix}^T = \text{diag}\left(\frac{1}{g(\bar{x}^{N0})}, \cdots, \frac{1}{g(\bar{x}^{NN})}\right)\left(D(\bar{x}_r^N)^T - \begin{bmatrix} f(\bar{x}^{N0}) \\ \vdots \\ f(\bar{x}^{NN}) \end{bmatrix}\right)$$

Comparing to the last equation in (8), it is obvious that $u^N(t_k) = \bar{u}^{Nk}$. So, (20) holds true. Part (i) is proved.

(ii) Assume $(x^N(t), u^N(t))$ satisfies the differential equations in (2). Because $x^N(t)$ are polynomials, (6) implies

$$\begin{aligned} \bar{x}_i^N D^T &= \begin{bmatrix} \dot{x}_i^N(t_0) & \dot{x}_i^N(t_1) & \cdots & \dot{x}_i^N(t_N) \end{bmatrix} \\ &= \begin{bmatrix} x_{i+1}^N(t_0) & x_{i+1}^N(t_1) & \cdots & x_{i+1}^N(t_N) \end{bmatrix} = \bar{x}_{i+1}^N \end{aligned} \tag{22}$$

Furthermore,

$$\begin{aligned} \bar{x}_r^N D^T &= \begin{bmatrix} \dot{x}_r^N(t_0) & \dot{x}_r^N(t_1) & \cdots & \dot{x}_r^N(t_N) \end{bmatrix} \\ &= \begin{bmatrix} f(x^N(t_0)) + g(x^N(t_0))u^N(t_0) & \cdots \end{bmatrix} \end{aligned}$$

Equations (22) and (23) imply that $(\bar{x}^N, \bar{u}^N)$ satisfies (8). Part (ii) is proved.

(iii) We know that the degree of $x_1^N(t)$, the interpolation polynomial, is less than or equal to N. From (i), we know $x_2^N(t) = \dot{x}_1^N(t)$. Therefore, the degree of $x_2^N(t)$ must be less than or equal to $N-1$. In general, the degree of $x_i^N(t)$ is less than or equal to $N - i + 1$. ■

Lemma 6: Suppose $\{(\bar{x}^N, \bar{u}^N)\}_{N=N_1}^{\infty}$ is a sequence satisfying (8), (10), (11) and (12), where $m_1 \ge 1$. Then,

$$\left\{ ||(x^N(t))^{(l)}||_\infty \,\middle|\, N \ge N_1,\, l = 0, 1, \cdots, m_1 \right\}$$

is bounded. If $f(x)$ and $g(x)$ are C^{m_1-1}, then

$$\left\{ ||(u^N(t))^{(l)}||_\infty \,\middle|\, N \ge N_1,\, l = 0, 1, \cdots, m_1 - 1 \right\}$$

is bounded.

Proof: Consider $(x_r^N(t))^{(m_1)}$. From Lemma 5, it is a polynomial of degree less than or equal to $N - r - m_1 + 1$.

$$(x_r^N(t))^{(m_1)} = \sum_{n=0}^{N-r-m_1+1} a_n^N(m_1) L_n(t)$$

where $L_n(t)$ is the Legendre polynomial of degree n. It is known that $|L_n(t)| \le 1$. Therefore, (12) implies that $||(x_r^N(t))^{(m_1)}||_\infty$ is bounded by $\boldsymbol{d}$ for all $N \ge N_1$. Now, let us consider $(x_r^N(t))^{(m_1-1)}$. From (6) we have,

$$\begin{aligned} (x_r^N(t))^{(m_1-1)} &= (x_r^N(t))^{(m_1-1)}|_{t=-1} + \textstyle\int_0^t (x_r^N(s))^{(m_1)} ds \\ &= \begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} D^{m_1-1}(\bar{x}_r^N)^T + \textstyle\int_0^t (x_r^N(s))^{(m_1)} ds \end{aligned}$$

So, $||(x_r^N(t))^{(m_1-1)}||_\infty$, $N \ge N_1$, is bounded because of (11). Similarly, we can prove all derivatives of $x_r^N(t)$ of order less than m_1 are bounded. The same approach can also be applied to prove the boundedness of$u^N(t)$. Because $f(x)$ and $g(x)$ have continuous derivatives of order less than or equal to $m_1 - 1$, the boundedness of

$$\left\{ ||(u^N(t))^{(l)}||_\infty \,\middle|\, N \ge N_1,\, j = 0, 1, \cdots, m_1 - 1 \right\}$$

follows the boundedness of $(x_r^N(t))^{(l)}$ proved above. ■

Given any function $h(t)$ defined on $[-1, 1]$. In the following, $U(h)$ represents an upper bound of $h(t)$ and $V(h)$ represents the total variation.

Lemma 7: Let $(x(t), u(t))$ be a solution of the differential equation (2). Suppose $x_r^{(m)}(t)$ has bounded variation for some $m \ge 2$. Let m_1 be an integer satisfying $1 \le m_1 \le m - 1$. Then, there exist constants $M > 0$ and $N_1 > 0$ so that for each integer $N \ge N_1$ the differential equation (2) has a solution $(x^N(t), u^N(t))$ in which $x^N(t)$ consists of polynomials of degree less than or equal to N. Furthermore, for $i = 1, \cdots, r; l = 1, \cdots, m_1$, the pair $(x^N(t), u^N(t))$ satisfies

$$||x_i^N(t) - x_i(t)||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-3/4}}, \tag{23}$$

$$||(x_r^N(t))^{(l)} - (x_r(t))^{(l)}||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-3/4}}, \tag{24}$$

$$||u^N(t) - u(t)||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-3/4}} \tag{25}$$

Furthermore, the spectral coefficients of $(x_r^N)^{(m_1)}(t)$ satisfy

$$|a_n^N(m_1)| \le \frac{6(U(x_r^{(m_1+1)}) + V(x_r^{(m_1+1)}))}{\sqrt{\pi} n^{3/2}}, \tag{26}$$

for $n = 1, 2, \cdots, N - r - 1$. If $f(x)$ and $g(x)$ have Lipschitz continuous Lth order derivatives for some $L \le m_1 - 1$, then

$$||(u^N(t))^{(l)} - (u(t))^{(l)}||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-3/4}}, l = 1 : L \tag{27}$$

Furthermore,

$$\begin{aligned} x^N(-1) &= x(-1) \\ u^N(-1) &= u(-1), \quad \text{If } m_1 \ge 2 \end{aligned} \tag{28}$$

Remark 3.3: In this lemma, if $x_r(t)$ has the m-th order strong derivative and $x_r^{(m)}(t)$ has bounded variation for some $m \ge 2$, then the inequalities (23), (24), and (25) are slightly tighter.

$$||x_i^N(t) - x_i(t)||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-1/2}}, \tag{29}$$

$$||(x_r^N(t))^{(l)} - (x_r(t))^{(l)}||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-1/2}}, \tag{30}$$

$$||u^N(t) - u(t)||_\infty \le \frac{M||x_r||_{W^{m,2}}}{N^{(m-m_1)-1/2}} \tag{31}$$

The proof is identical as that of Lemma 7 except that the error estimation in (3) of Lemma 4 is used.

Proof: Consider the Legendre series

$$(x_r)^{(m_1)}(t) \sim \sum_{n=0}^{N-r-m_1+1} a_n^N(m_1) L_n(t)$$

Define a sequence of polynomials $x_1^N(t), \cdots, x_{r+m_1}^N(t)$,

$$\begin{aligned} x_{r+m_1}^N(t) &= \sum_{n=0}^{N-r-m_1+1} a_n^N(m_1)L_n(t) \\ x_{r+m_1-1}^N(t) &= (x_r)^{(m_1-1)}(-1) + \int_{-1}^t x_{r+m_1}^N(s)ds \\ &\vdots \\ x_{r+1}^N(t) &= \dot{x}_r(-1) + \int_{-1}^t x_{r+2}^N(s)ds \end{aligned}$$

$$x_i^N(t) = x_i(-1) + \int_{-1}^t x_{i+1}^N(s)ds, \quad \text{for } 1 \le i \le r$$

$$u^N(t) = \frac{x_{r+1}^N(t) - f(x^N(t))}{g(x^N(t))}$$

From the definition of $x^N(t)$, we have $x^N(-1) = x(-1)$. If $m_1 \ge 2$, then $x_{r+1}(-1) = \dot{x}_r(-1)$. From the definition of $u^N(t)$, we know $u^N(-1) = u(-1)$ provided $m_1 \ge 2$. Therefore, $(x^N(t), u^N(t))$ satisfies (28). It is obvious that $x_i^N(t)$ is a polynomial of degree less than or equal to N; and $(x^N(t), u^N(t))$ satisfies the differential equation (2). Because we assume $V(x_r^{(m)}) < \infty$, we have $x_r^{(m)} \in L^2$. From Lemma 4

$$||x_{r+m_1}^N(t) - x_r^{(m_1)}(t)||_\infty = ||x_r^{(m_1)}(t) - \sum_{n=0}^{N-r-m_1+1} a_n^N(m_1)L_n(t)||_\infty \le \frac{C_1||x_r||_{W^{m,2}}}{N^{-(m-m_1)+3/4}}$$

for some constant $C_1 > 0$. Therefore,

$$\begin{aligned} &|x_{r+m_1-1}^N(t) - (x_r)^{(m_1-1)}(t)| \\ &\le \int_{-1}^t |x_{r+m_1}^N(s) - (x_r)^{(m_1)}(s)|ds \\ &\le 2C_1||x_r||_{W^{m,2}} N^{-(m-m_1)+3/4} \end{aligned}$$

Similarly, we can prove (23) and (24).

To prove (26), note that the spectral coefficient $a_n^N(m_1)$ of $(x_r^N)^{(m_1)}(t)$ is the same as the spectral coefficients of $(x_r)^{(m_1)}(t)$. From Jackson's Theorem (Lemma 2), we have

$$|a_n^N(m_1)| < \frac{6}{\sqrt{\pi}}(U(x_r^{(m_1+1)}) + V(x_r^{(m_1+1)}))\frac{1}{n^{3/2}}$$

Because f and g are Lipschitz continuous. In a bounded set around $x(t)$, $g(x) > \alpha > 0$ for some $\alpha > 0$. Therefore, the function

$$\frac{s - f(x)}{g(x)}$$

is Lipschitz in a neighborhood of (x, s), i.e. there exists a constant C_2 independent of N such that

$$\begin{aligned} &|u^N(t) - u(t)| = \\ &\left|\frac{x_{r+1}^N(t) - f(x^N(t))}{g(x^N(t))} - \frac{\dot{x}_r(t) - f(x(t))}{g(x(t))}\right| \le C_2(|x_{r+1}^N(t) \\ &-\dot{x}_r(t)| + |x_1^N(t) - x_1(t)| + \cdots + |x_r^N(t) - x_r(t)|) \end{aligned} \quad (32)$$

Hence, (25) follows (23), (24) and (32) when $l = 0$. Similarly, we can prove (27) for $l \le L$. ■

Now, we are ready to prove Theorem 1.

Proof of Theorem 1: Let $(x^*(t), u^*(t))$ be an optimal solution to Problem B. According to Lemma 7 and Remark 3.3, for any positive integer N that is large enough, there exists a pair of functions $(\hat{x}^N(t), \hat{u}^N(t))$ in which $\hat{x}^N(t)$ consists of polynomials of degree less than or equal to N. Furthermore, the pair satisfies the differential equation with initial conditions in Problem B and

$$||\hat{x}^N(t) - x^*(t)||_\infty < \frac{L}{N^{m-m_1-1/2}} \quad (33)$$

$$||\hat{u}^N(t) - u^*(t)||_\infty < \frac{L}{N^{m-m_1-1/2}} \quad (34)$$

$$||(\hat{x}_r^N(t))^{(l)} - (x_r^*(t))^{(l)}||_\infty < \frac{L}{N^{m-m_1-1/2}}, \quad (35)$$

where $1 \le l \le m_1$. If we define

$$\hat{\bar{u}}^{Nk} = \hat{u}^N(t_k), \hat{\bar{x}}^{Nk} = \hat{x}^N(t_k)$$

Then $\{(\hat{\bar{x}}^N, \hat{\bar{u}}^N)\}$ satisfies (8) and (9) (Lemma 5 and 7). Because $\hat{x}_r^N(t)$ is a polynomial of degree less than or equal to N and because of (6), we know $(\hat{x}_r^N(t))^{(j)}$ equals the interpolation polynomial of $\hat{\bar{x}}_r^N(D^T)^j$. So,

$$\begin{bmatrix} 1 & 0 & \cdots & 0 \end{bmatrix} D^j(\hat{\bar{x}}_r^N)^T = (\hat{x}_r^N(t))^{(j)}|_{t=-1}$$

Therefore, (35) implies (11) if the bounds $\underline{\boldsymbol{b}}_j$ and $\boldsymbol{b}_j$ are large enough. In addition, the spectral coefficients of $\hat{\bar{x}}_r^N(D^T)^{m_1}$ is the same as the spectral coefficients of $(\hat{x}_r^N(t))^{(m_1)}$. From (26), (16), and the definition of ζ function, we have

$$\sum_{n=0}^{N-r-m_1+1} |a_n^N(m_1)| \le \boldsymbol{d}$$

So, the spectral coefficients of $(\hat{x}_r^N)^{(m_1)}$ satisfies (12). Because we select $\underline{\boldsymbol{b}}$ and $\bar{\boldsymbol{b}}$ large enough so that the optimal trajectory of the original continuous-time problem is contained in the interior of the region, then (33) and (34) imply (10) for N large enough. In summary, we have proved that $(\hat{\bar{x}}^N, \hat{\bar{u}}^N)$ is a discrete feasible trajectory satisfying all constraints, (8)-(12), in Problem $\mathrm{B^N}$.

The cost $J(x^*(\cdot), u^*(\cdot))$ can be considered as a functional, denoted by $\mathcal{J}(u)$. Because all the functions in Problem B are C^m with $m \ge 2$, we know that $\mathcal{J}(u)$ has second order Fréchet derivative. By Lemma 1

$$\begin{aligned} &|J(x^*(\cdot), u^*(\cdot)) - J(\hat{x}^N(\cdot), \hat{u}^N(\cdot))| = |\mathcal{J}(u^*) - \mathcal{J}(\hat{u}^N)| \\ &\le C_1(||u^* - \hat{u}^N||^2_{W^{m_1-1,\infty}}) \le \frac{C_2}{N^{2m-2m_1-1}} \end{aligned} \quad (36)$$

for some constant numbers C_1 and C_2 (34).

Now, consider $F(\hat{x}^N(t), \hat{u}^N(t))$ as a function of t. Let $F^N(t)$ represent the polynomial interpolation of this function at $t = t_0, t_1, \cdots, t_N$. Let $\hat{p}(t)$ be the best polynomial approximation of $F(\hat{x}^N(t), \hat{u}^N(t))$ under the norm of $L^\infty[-1, 1]$.

$$\begin{aligned} &|J(\hat{x}^N(\cdot), \hat{u}^N(\cdot)) - \bar{J}^N(\hat{\bar{x}}^N, \hat{\bar{u}}^N)| \\ &= \left|\int_{-1}^1 F(\hat{x}^N(t), \hat{u}^N(t))dt - \int_{-1}^1 F^N(t)dt\right| \\ &\le \int_{-1}^1 |F(\hat{x}^N(t), \hat{u}^N(t)) - F^N(t)|dt \\ &\le 2(1 + \Lambda_N)||\hat{p}(t) - F(\hat{x}^N(t), \hat{u}^N(t))||_\infty \end{aligned}$$

where

$$\Lambda_N \leq \frac{2}{\pi} log(N+1) + 0.685\cdots \quad (37)$$

is the Lebesgue constant. The inequality (III) is a corollary of Lemma 3. Because $f(\cdot)$, $g(\cdot)$, and $F(\cdot)$ are C^m, it is known (Lemma 4) that the best polynomial approximation satisfies

$$||\hat{p}(t) - F(\hat{x}^N(t), \hat{u}^N(t))||_\infty \leq \frac{C_3||F(\hat{x}^N(t), \hat{u}^N(t))||_{W^{m_1-1,\infty}}}{N^{m_1-1}}$$

Because of Lemma 6, $\{||F(\hat{x}^N(t), \hat{u}^N(t))||_{W^{m_1-1,\infty}} | N \geq N_1\}$ is bounded. Therefore,

$$|J(\hat{x}^N(\cdot), \hat{u}^N(\cdot)) - \bar{J}^N(\hat{\bar{x}}^N, \hat{u}^N)| \leq \frac{(1+\Lambda_N)C_4}{N^{m_1-1}} \leq \frac{C_5}{N^\alpha} \quad (38)$$

for some constant numbers C_4 and C_5 independent of N and any $\alpha < m_1 - 1$. Let

$$\{(\bar{x}^{*N}, \bar{u}^{*N})\}_{N=N_0}^{\infty} \quad (39)$$

be a sequence of optimal discrete solutions. Its interpolation is denoted by $(x^{*N}(t), u^{*N}(t))$. Then, similar to the derivation above, we can prove

$$\begin{aligned} & |J(x^{*N}(\cdot), u^{*N}(\cdot)) - \bar{J}^N(\bar{x}^{*N}, \bar{u}^{*N})| \\ \leq\ & 2(1+\Lambda_N)||p^N(t) - F(x^{*N}(t), u^{*N}(t))||_\infty \quad (40) \\ \leq\ & \frac{C_6(1+\Lambda_N)}{N^{m_1-1}}||F(x^{*N}(t), u^{*N}(t))||_{W^{m_1-1,\infty}} \end{aligned}$$

where $p^N(t)$ is the best polynomial approximation of $F(x^{*N}(t), u^{*N}(t))$ with degree less than or equal to N. Because of Lemma 6, $||F(x^{*N}(t), u^{*N}(t))||_{W^{m_1-1,\infty}} | N \geq N_1\}$ is bounded. So

$$|J(x^{*N}(\cdot), u^{*N}(\cdot)) - \bar{J}^N(\bar{x}^{*N}, \bar{u}^{*N})| \leq \frac{C_7}{N^\alpha} \quad (41)$$

for some constant $C_7 > 0$. Now, we are ready to piece together the puzzle of inequalities and finalize the proof.

$$\begin{aligned} & J(x^*(\cdot), u^*(\cdot)) \\ & \leq J(x^{*N}(\cdot), u^{*N}(\cdot)) \quad \left(\begin{array}{l} (x^{*N}(t), u^{*N}(t)) \text{ is a feasible} \\ \text{trajectory (Lemma 5)} \end{array} \right) \\ & \leq \bar{J}^N(\bar{x}^{*N}, \bar{u}^{*N}) + \frac{C_7}{N^\alpha} \quad (\text{ inequality } (41)) \\ & \leq \bar{J}^N(\hat{x}^N, \hat{u}^N) + \frac{C_7}{N^\alpha} \quad \left(\begin{array}{l} (\hat{x}^N, \hat{u}^N) \text{ is a feasible discrete} \\ \text{trajectory and } (\bar{x}^{*N}, \bar{u}^{*N}) \text{ is optimal} \end{array} \right) \\ & \leq J(\hat{x}^N(\cdot), \hat{u}^N(\cdot)) + \frac{C_5}{N^\alpha} + \frac{C_7}{N^\alpha} \quad (\text{ inequality } (38)) \\ & \leq J(x^*(\cdot), u^*(\cdot)) + \frac{C_2}{N^{2m-2m_1-1}} \\ & + \frac{C_5}{N^\alpha} + \frac{C_7}{N^\alpha} \quad (\text{ inequality } (36)) \end{aligned}$$

Therefore,

$$\begin{aligned} & 0 \leq J(x^{*N}(\cdot), u^{*N}(\cdot)) - J(x^*(\cdot), u^*(\cdot)) \\ & \leq \frac{C_2}{(N-r-m_1-1)^{2m-2m_1-1}} + \frac{C_5}{N^\alpha} + \frac{C_7}{N^\alpha} \end{aligned}$$

This inequality implies (18). Furthermore, (18) and (41) imply (19). □

IV. Conclusion

According to Theorem 1, the rate of convergence for the optimal cost is determined by both m and m_1. It can be proved that the optimal selection of m_1 results in a convergence rate of at least $\frac{1}{N^{\frac{2(m-1)}{3}-1}}$. The proof is based on basic algebra; and it is omitted in this paper for the reason of space.

References

[1] J. T. Betts, *Practical Methods for Optimal Control Using Nonlinear Programming*, SIAM, Philadelphia, PA, 2001.

[2] J. T. Betts, "Survey of Numerical Methods for Trajectory Optimization," *Journal of Guidance, Control, and Dynamics,* Vol. 21, No. 2, 1998, pp. 193-207.

[3] E. Polak, *Optimization: Algorithms and Consistent Approximations*, Springer-Verlag, Heidelberg, 1997.

[4] S. W. Paris, J. P. Riehl, and W. K. Sjauw, "Enhanced Procedures for Direct Trajectory Optimization Using Nonlinear Programming and Implicit Integration," *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, 21-24 August 2006, Keystone, CO. AIAA Paper No. 2006-6309.

[5] J. P. Riehl, S. W. Paris, and W. K. Sjauw, "Comparision of Implicit Integration Methods for Solving Aerospace Trajectory Optimization Problems," *Proceedings of the AIAA/AAS Astrodynamics Specialist Conference and Exhibit*, 21-24 August 2006, Keystone, CO. AIAA Paper No. 2006-6033.

[6] W. Kang, N. Bedrossian, Pseudospectral Optimal Control Theory Makes Debut Flight - Saves NASA $1M in under 3 hrs, SIAM News, September, 2007.

[7] Ross, I. M., User's Manual for DIDO: A MATLAB Application Package for Solving Optimal Control Problems, Tomlab Optimization, February, 2004.

[8] S. W. Paris and C. R. Hargraves, *OTIS 3.0 Manual*, Boeing Space and Defense Group, Seattle, WA, 1996.

[9] C. Canuto, M. Y. Hussaini, A. Quarteroni and T. A. Zang, *Spectral Method in Fluid Dynamics.* New York: Springer-Verlag, 1988.

[10] J. Hesthaven, S. Gottlieb, and D. Gottlieb, Spectral Methods for Time-Dependent Problems, Cambridge University Press, 2007.

[11] A. L. Dontchev and W. W. Hager, The Euler approximation in state constrained optimal control, *Mathematics of Computation,* Vol. 70, pp. 173-203, 2000.

[12] W. W. Hager, Runge-Kutta methods in optimal control and the transformed adjoint system, *Numerische Mathematik,* Vol. 87, pp. 247-282, 2000.

[13] Fahroo, F., Ross, I. M., "Costate Estimation by a Legendre Pseudospectral Method," Proceedings of the AIAA Guidance, Navigation and Control Conference, 10-12 August 1998, Boston, MA.

[14] Q. Gong, M. Ross, W. Kang, F. Fahroo, Connections Between the Covector Mapping Theorem and Convergence of Pseudospectral Methods for Optimal Control, Computational Optimization and Applications, to appear.

[15] Q. Gong, W. Kang, and I. M. Ross, A Pseudospectral Method for the Optimal Control of Constrained Feedback Linearizable Systems, IEEE Trans. Automat. Contr., Vol. 51, No. 7, pp. 1115-1129, 2006.

[16] W. Kang, I. M. Ross, and Q. Gong, Pseudospectral Optimal Control and Its Convergence Theorems,*Analysis and Design of Nonlinear Control Systems - In Honor of Alberto Isidori*, Alessandro Astolfi and Lorenzo Marconi eds., Springer, to appear.

[17] W. Kang, Q. Gong, and I. M. Ross, On the Convergence of Nonlinear Optimal Control using Pseudospectral Methods for Feedback Linearizable Systems, *International Journal of Robust and Nonlinear Control*, Vol. 17, 1251-1277, online publication 3 January, 2007.

[18] J. P. Boyd, Chebyshev and Fourier Spectral Methods, second edition, Dover, 2001.

[19] A. Wouk, A Course of Applied Functional Analysis, John Wiley & Sons, New York, 1979.

[20] G. Sansone, A. H. Diamond, and E. Hille, Orthogonal Functions, Robert E. Krieger Publishing Co., Huntington, New York, 1977.