JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 17, No. 6, November–December 1994

Optimization of Very-Low-Thrust, Many-Revolution Spacecraft Trajectories

Wayne A. Scheel*
Spectrum Astro, Inc., Gilbert, Arizona 85233
and
Bruce A. Conway†
University of Illinois, Urbana, Illinois 61801

Optimal minimum flight time solutions are obtained for continuous, very-low-thrust orbit transfers using a direct-transcription approach to convert the continuous optimal control problem into a nonlinear programming problem. The thrust accelerations used are characteristic of solar electric and nuclear electric propulsion, resulting in trajectories that require many revolutions of Earth to achieve the desired final orbits. Among the problems examined are transfers from low Earth orbit to geosynchronous orbit (GEO) and orbit raising from GEO to a specified radius. All initial and terminal orbits are circular, with motion constrained to the equatorial plane. Motion of the spacecraft is described using the equinoctial orbit elements. The variation of spacecraft mass and acceleration due to fuel consumption is modeled. The orbit transfers include the effect of Earth's oblateness through first order as well as third-body perturbations from the moon.

Introduction

THERE are essentially two approaches to solving optimal control problems. The "indirect" approach uses the calculus of variations to derive the necessary conditions of optimality, i.e., the Euler-Lagrange equations. This results in a two-point boundary-value problem (TPBVP), which may then be solved numerically, although this becomes a very difficult task for significant problems. The "direct" approach is to discretize the original problem, transforming it into a parameter optimization problem, which is then solved using mathematical programming.[1] Numerous methods have been used to solve the TPBVP that results from the indirect approach to the optimal control problem. One approach to solving this problem is the initial-value "shooting" methods. Redding and Breakwell[2] used this method for low-thrust orbit transfers. The shooting method requires a guess of either the initial or terminal boundary conditions, from which the Euler-Lagrange system is then integrated forward or backward, iteratively adjusting the guess in an attempt to satisfy the boundary conditions at the other end. Common difficulties with shooting methods are the requirement of an initial guess for the adjoint variables of the TPBVP, the sensitivity of the adjoint equations to variations in the initial values of the adjoint variables and the Euler-Lagrange equations, and possible occurrences of discontinuities in the optimal control.[1] This sensitivity can make the numerical computation for an improved estimate difficult to determine.[3] Another approach in solving the TPBVP avoids explicitly integrating the system differential equations. Instead, algebraic relationships approximate the differential equations locally. These algebraic relationships, along with the boundary conditions, form a system of nonlinear simultaneous equations. This system is then solved using variants of Newton's method. Examples of these types of approaches include finite difference and collocation techniques, in which difference quotients for derivatives and polynomials, respectively, approximate the TPBVP differential equations.[3]

The direct approach ignores the calculus of variations first-order necessary conditions and approximates the original optimal control problem with a nonlinear programming (NLP) problem that often consists of a large number of variables and constraints. Without the aid of sophisticated NLP codes capable of handling such large problems, the NLP approach was unattractive for many years. However, with recent advances in the development of NLP codes, there has been a revived interest in this approach.

In 1987, Hargraves and Paris[4] developed a "direct-collocation" method that converts the original optimal control problem, by discretizing the equations of motion using a collocation scheme, into an NLP problem. The collocation scheme employed was one that Dickmanns and Well[5] used to solve the TPBVP of the indirect approach, which was a particular implementation of a general collocation method for boundary-value problems analyzed by Russell and Shampine.[6] In the direct-collocation approach, the state histories are discretized and represented by piecewise Hermite cubic polynomials whereas the controls are linearly represented. A collocation scheme enforces the satisfaction of the constraints (i.e., the differential equations at the center of each cubic), and a sequential quadratic programming algorithm is used to solve the resulting NLP problem.

The direct-collocation and nonlinear programming (DCNLP) method has recently been used successfully in many trajectory optimization problems. Enright and Conway,[1,7] Tang,[8] Herman,[9] and Downey and Conway[10] have obtained optimal trajectories for numerous problems using the DCNLP method. These include optimal low-thrust Earth escape, optimal multiburn transfers and rendezvous, interplanetary transfers from Earth to Mars, optimal spacecraft attitude control, and optimal interception of Earth-orbiting spacecraft. The largest NLP problem attempted thus far by our research group at the University of Illinois was in Tang's master's thesis[8] on interplanetary transfer from Earth to Mars. Tang's DCNLP problem had 918 variables and 728 constraints.

Enright and Conway[1] also developed a Runge-Kutta "parallel-shooting" scheme for the NLP problem. The Runge-Kutta (RK) parallel-shooting method was developed as a way to decrease the size of the NLP problem. This new transcription method is essentially an adaptation of the parallel-shooting method for boundary-value problems to the discretization of the equation of motion (EOM) constraints in the direct approach to the optimal control problem.[3] An explicit four-stage RK integration rule is used to propagate the EOMs across the discretized segments of the trajectory. Requiring continuity at the nodes of the segment boundaries generates a set of discrete nonlinear algebraic constraints involving the states and controls on the boundaries of the segment. These nonlinear constraints replace the constraints from the original Hargraves-Paris[4] method. Because the RK method is explicit, it can be incorporated into a parallel-shooting scheme; i.e., the propagation of the states can proceed for several segments before a constraint need be formed.

Received Oct. 2, 1993; revision received March 27, 1994; accepted for publication March 22, 1994.

*Engineer. Member AIAA.
†Professor. Associate Fellow AIAA.

The parallel-shooting method thus allows for the use of larger intervals and results in smaller NLP problems. Although additional control variables must be introduced in each segment to accommodate the multiple integration steps, the intermediate state variables (resulting from the forward propagation) are used to reinitialize values for the next step and never appear explicitly in the NLP problem. Since most optimal control problems have many more state variables than control variables, the trade-off of introducing more control variables and reducing the number of state variables is a favorable one. This improvement on the original collocation method will be necessary to solve problems with very small thrust accelerations and correspondingly long times of flight.

This work presents the optimal solutions for minimum-time orbit raising transfers using continuous, very-low-thrust spacecraft. Direct transcription and nonlinear programming are used to optimize the trajectory, implementing the RK parallel-shooting method previously discussed. Optimal transfers from an initial circular orbit to a final circular orbit are obtained for the two cases of low Earth orbit (LEO) to geosynchronous orbit (GEO) transfer and from GEO to an outer orbit transfer of specified radius. Both Earth oblateness effects and third-body perturbations from the moon are modeled. Thrust accelerations of 325 μg and 4.95 μg are used for the LEO-to-GEO and GEO-to-outer-orbit cases, respectively, and were selected as representative of current and/or future continuous very-low-thrust spacecraft. For simplicity, the transfer is only two-dimensional, taking place in the equatorial plane. Similarly, the Moon is assumed to be in this plane. Variation of the spacecraft's mass due to fuel consumption is also modeled.

Equations of Motion

Although previous work[3,4] by our research group in the area of optimal spacecraft orbit transfer has utilized polar coordinates (i.e., radius, velocity, angular position, and angular velocity) to describe the motion of a spacecraft, this work employs orbit elements to describe the motion of the spacecraft. This has two advantages: The governing equations become those of celestial mechanics perturbation theory and, for low-thrust trajectories, all of the orbit elements but one vary slowly. A choice of coordinates such that the coordinates change slowly and within a restricted range should improve the speed of convergence of the solution algorithm and make it more robust.

For the coplanar two-body problem, the disturbing perturbation is simply the very-low-thrust acceleration of the spacecraft and may be resolved into components as

$$\boldsymbol{F} = R\hat{e}_R + T\hat{e}_T \tag{1}$$

where the unit vector $\hat{e}_R$ points radially outward and unit vector $\hat{e}_T$ is normal to $\hat{e}_R$, lies in the orbit plane, and is positive in the direction of motion, as shown in Fig. 1. There will be no normal force upon the spacecraft, as thrust acceleration and perturbing forces such as the attraction of the moon are assumed to act in the original orbit plane. In Fig. 1, α is the thrust acceleration vector and the angle β defines the spacecraft thrust direction measured from $\hat{e}_T$. When no external perturbations are assumed to act, we have $T = \alpha \cos\beta$ and $R = \alpha \sin\beta$. Figure 1 also shows how L, defined later as the true longitude, measures the spacecraft's angular position from the fixed line of Aries.

For motion confined to one plane, with circular initial and terminal orbits, it is beneficial to use equinoctial elements to describe the orbit. The name "equinoctial elements" were first introduced by Arsenault et al.[11] in 1970, and recent authors such as Broucke and Cefola[12] and Battin[13] have continued this usage. Brouke and Cefola[12] proved that the variational equations in terms of equinoctial elements are valid when the eccentricity or inclination is zero. In the set of equinoctial variables, elements P_1, P_2, Q_1, and Q_2 replace the classical elements e, i, ω, and Ω in the following manner:

$$P_1 = e\sin\tilde{\omega} \tag{2}$$

$$P_2 = e\cos\tilde{\omega} \tag{3}$$

$$Q_1 = \tan\tfrac{1}{2}i\sin\Omega \tag{4}$$

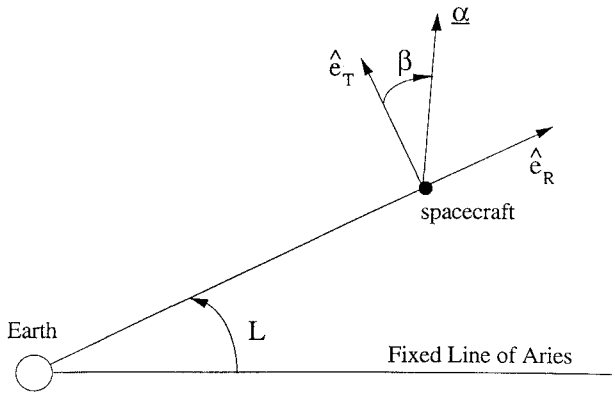

Fig. 1 Illustrating the body-fixed unit-vector basis.

$$Q_2 = \tan\tfrac{1}{2}i\cos\Omega \tag{5}$$

where $\tilde{\omega}$ is the longitude of pericenter and is defined as

$$\tilde{\omega} = \Omega + \omega \tag{6}$$

For coplanar transfers in the equatorial plane, $i = 0°$ and Q_1 and Q_2 will equal zero throughout the duration of the transfer. Thus, the time rates of change for Q_1 and Q_2 will not be required in the set of governing equations.

The equations for the time rates of change of the equinoctial elements are given by Battin.[13] The time rates of change of the semimajor axis a, element P_1, element P_2, and mean longitude l, are, respectively,

$$\frac{da}{dt} = \frac{2a^2}{h}\left[(P_2\sin L - P_1\cos L)R + \frac{p}{r}T\right] \tag{7}$$

$$\frac{dP_1}{dt} = \frac{r}{h}\left\{-\frac{p}{r}(\cos L)R + \left[P_1 + \left(1+\frac{p}{r}\right)\sin L\right]T\right\} \tag{8}$$

$$\frac{dP_2}{dt} = \frac{r}{h}\left\{\frac{p}{r}(\sin L)R + \left[P_2 + \left(1+\frac{p}{r}\right)\cos L\right]T\right\} \tag{9}$$

$$\frac{dl}{dt} = n - \frac{r}{h}\left\{\left[\frac{a}{a+b}\frac{p}{r}(P_1\sin L + P_2\cos L) + \frac{2b}{a}\right]R + \frac{a}{a+b}\left(1+\frac{p}{r}\right)(P_1\cos L - P_2\sin L)T\right\} \tag{10}$$

where

$$b = a\sqrt{1 - P_1^2 - P_2^2} \tag{11}$$

$$h = nab \tag{12}$$

$$\frac{p}{r} = 1 + P_1\sin L + P_2\cos L \tag{13}$$

$$\frac{r}{h} = \frac{h}{\mu(1 + P_1\sin L + P_2\cos L)} \tag{14}$$

$$\frac{a}{a+b} = \frac{1}{1+\sqrt{1 - P_1^2 - P_2^2}} \tag{15}$$

and n is the mean motion of the spacecraft about Earth:

$$n = \sqrt{\mu/a^3} \tag{16}$$

The mean longitude l is defined as

$$l = \tilde{\omega} + M \tag{17}$$

Although the mean longitude does not appear explicitly in the EOMs, true longitude L does. True longitude is defined as

$$L = \tilde{\omega} + f \tag{18}$$

where f is the true anomaly of the spacecraft. When L is needed in the right-hand sides of the EOMs, it must be calculated in the following manner. First, a form of Kepler's equation,

$$l = K + P_1 \cos K - P_2 \sin K \tag{19}$$

must be solved for the eccentric longitude K. Eccentric longitude is defined as:

$$K = \tilde{\omega} + E \tag{20}$$

where E is the eccentric anomaly. Then L is calculated from the eccentric longitude according to the relations[13]

$$\sin L = \frac{a}{r}\left[\left(1 - \frac{a}{a+b}P_2^2\right)\sin K + \frac{a}{a+b}P_1 P_2 \cos K - P_1\right] \tag{21}$$

$$\cos L = \frac{a}{r}\left[\left(1 - \frac{a}{a+b}P_1^2\right)\sin K + \frac{a}{a+b}P_1 P_2 \sin K - P_2\right] \tag{22}$$

The spacecraft's mass variation due to fuel consumption will constitute the fifth and final governing equation. The equation for the time rate of change of the spacecraft's thrust acceleration due to the decrease in spacecraft mass with time is given by

$$\frac{d\alpha}{dt} = \frac{\alpha^2}{c} \tag{23}$$

where α is the thrust acceleration and c is the effective exhaust velocity.[3] The classical orbit elements are easily recoverable from the following relations[13]:

$$e^2 = P_1^2 + P_2^2 \tag{24}$$

$$\tan^2 \tfrac{1}{2}i = Q_1^2 + Q_2^2 \tag{25}$$

$$\tan \tilde{\omega} = \frac{P_1}{P_2} \tag{26}$$

$$\tan \Omega = \frac{Q_1}{Q_2} \tag{27}$$

$$\omega = \tilde{\omega} - \Omega \tag{28}$$

Through first order (i.e., J_2), the perturbing acceleration upon the spacecraft due to the nonuniformity of Earth's gravitational field is

$$\boldsymbol{F}_{\text{obl}} = -\frac{3\mu J_2 R_e^2}{r^4}\left[\hat{e}_R\left(\frac{1}{2} - \frac{3\sin^2 i \sin^2\theta}{2}\right) + \hat{e}_T \sin^2 i \sin\theta\cos\theta + \hat{e}_N \sin i \sin\theta \cos i\right] \tag{29}$$

where $\theta = \omega + f$, R_e is the radius of Earth, and r is the instantaneous radius of the spacecraft's orbit. For an equatorial orbit the perturbing acceleration due to Earth's oblateness through first order reduces to

$$\boldsymbol{F}_{\text{obl}} = -\frac{3\mu J_2 R_e^2}{2r^4}\hat{e}_R = R_{\text{obl}}\hat{e}_R \tag{30}$$

Perturbations upon the spacecraft due to the moon have also been modeled. The EOM of the spacecraft can be written as

$$\frac{d^2\boldsymbol{r}}{dt^2} + \frac{\mu}{r^3}\boldsymbol{r} = -\mu_m\left(\frac{\boldsymbol{r}_{ms}}{r_{ms}^3} + \frac{\boldsymbol{r}_m}{r_m^3}\right) \tag{31}$$

where μ_m is the gravitational constant for the moon, $\boldsymbol{r}$ is the position vector from Earth to the spacecraft, $\boldsymbol{r}_m$ is the position vector from Earth to the moon, and $\boldsymbol{r}_{ms}$ is the position vector of the spacecraft with respect to the moon. The radial and tangential components of the disturbing acceleration due to the attraction of the moon are thus

$$R_{\text{mper}} = \mu_m\left(\frac{\boldsymbol{r}_{ms}\cdot\hat{e}_R}{r_{ms}^3} + \frac{\boldsymbol{r}_m\cdot\hat{e}_R}{r_m^3}\right) \tag{32}$$

$$T_{\text{mper}} = \mu_m\left(\frac{\boldsymbol{r}_{ms}\cdot\hat{e}_T}{r_{ms}^3} + \frac{\boldsymbol{r}_m\cdot\hat{e}_T}{r_m^3}\right) \tag{33}$$

where

$$\boldsymbol{r}_m\cdot\hat{e}_R = r_m(\cos\gamma\cos L + \sin\gamma\sin L) \tag{34}$$

$$\boldsymbol{r}_m\cdot\hat{e}_T = r_m(\sin\gamma\cos L - \cos\gamma\sin L) \tag{35}$$

$$\boldsymbol{r}_{ms}\cdot\hat{e}_R = -\boldsymbol{r}_m\cdot\hat{e}_R + r \tag{36}$$

$$\boldsymbol{r}_{ms}\cdot\hat{e}_T = -\boldsymbol{r}_m\cdot\hat{e}_T \tag{37}$$

and γ, the angular position of the moon with respect to Earth, measured from the fixed line of Aries, is given by

$$\gamma = n_m t + \gamma_0 \tag{38}$$

where n_m is the mean motion of the moon in its assumed circular orbit about Earth, t is elapsed time, and γ_0 is initial angular displacement of the moon from the fixed line of Aries at $t = t_0$.

NZSOL, the NLP problem solver used for the problems solved in this work, performs better when the parameters of the problem have similar magnitudes.[14] Although scaling the parameters can accomplish this, as is later described, using canonical units can make the scaling process easier. The canonical distance unit (DU) used here is one Earth radius. To yield a value of unity for the gravitational parameter μ, the period of a circular orbit at Earth's surface must be 2π times the corresponding time unit. Thus, one time unit is equivalent to 13.45 min. The normalized value for the moon's gravitational constant, μ_m, is 1.23×10^{-2}.

The objective function is the flight time. With constant, continuous thrust, minimizing the flight time is equivalent to minimizing the required fuel, an important objective in most missions.

Method of Solution

The goal of this work is to develop a method for obtaining optimal orbit raising trajectories for values of thrust so low that a large number (~100) of revolutions must be used. It is easily determined that using the DCNLP method as employed by Hargraves and Paris[4] and Tang[8] would yield a problem too large for the NLP code. With a minimum of three segments per orbit, the discretization of a 100-revolution transfer requires 300 segments. The number of state variables alone would be 1505 (301 nodes times 5 state variables), with another 601 variables required to represent the controls, and one more variable to represent the final time. This results in an NLP problem consisting of more than 2100 variables, or approximately twice as large as any problem solved to date to our knowledge using the NZSOL code. Thus, an alternative method needs to be implemented in order to reduce the size of this very large DCNLP problem.

The alternative that was successfully used employs the RK parallel-shooting algorithm of Enright and Conway.[1,3] In this algorithm, illustrated schematically in Fig. 2, the previously defined optimal control problem is first discretized into a sequence of stages. The partition $[t_0, t_1, \ldots, t_N]$ is introduced, with $t_0 = 0$, $t_N = t_f$, where $t_0 < t_1 < \cdots < t_N$, and letting $h_i = t_i - t_{i-1}$ for $i = 2, \ldots, N$. The h_i may or may not be uniform. The mesh points t_i are referred to as "nodes," whereas the intervals $[t_{i-1}, t_i]$ are referred to as "segments" [1, 4]. The state variables $x_i = x(t_i)$ are approximated by values at the nodes, for $i = 0, 1, \ldots, N$. Control variables are provided at the nodes t_i as well as the segment center points, $t_i + \frac{1}{2}h$, by $u_i = u(t_i)$ for $i = 0, \ldots, N$ and $v_i = u(t_{i-1} + \frac{1}{2}h)$ for $i = 1, \ldots, N$. From a given node t_{i-1}, the equations of motion are integrated forward

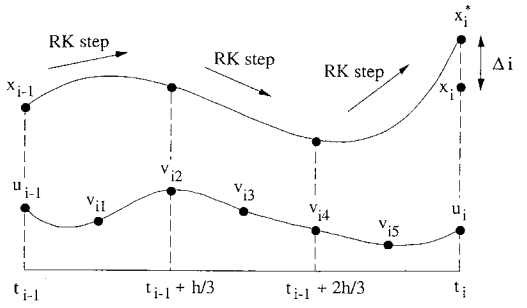

Fig. 2 Illustration of parallel-shooting method ($p = 3$).

from x_{i-1} to the next node t_i using the controls u_{i-1}, v_i, and u_i by a step of a four-stage RK formula[3]:

$$y_i^1 = x_{i-1} + \tfrac{1}{2}hf(x_{i-1}, u_{i-1}) \tag{39}$$

$$y_i^2 = x_{i-1} + \tfrac{1}{2}hf\left(y_i^1, v_i\right) \tag{40}$$

$$y_i^3 = x_{i-1} + hf\left(y_i^2, v_i\right) \tag{41}$$

$$y_i^4 = x_{i-1} + \tfrac{1}{6}h\left[f(x_{i-1}, u_{i-1}) + 2f\left(y_i^1, v_i\right) + 2f\left(y_i^2, v_i\right) + f\left(y_i^3, u_i\right)\right] \tag{42}$$

The state at the next node is estimated by y_i^4; thus, for continuity it is necessary that

$$\Delta_i = y_i^4 - x_i = 0 \tag{43}$$

for $i = 1, \ldots, N$. These are referred to as the "Runge-Kutta" defects.[3] The RK procedure has order h^5 local truncation errors.[3] Perhaps the biggest advantage of this RK method is that, since it is explicit, it can be incorporated into a parallel-shooting approach. The single step of the RK procedure previously described is replaced with multiple steps. This allows the use of larger intervals, resulting in smaller NLP problems. In each segment, additional control variables must be introduced to accommodate the multiple integration steps.

Let p be the number of integration steps per segment, $[t_{i-1}, t_i]$. In the usual manner, states and controls are provided at the nodes, $x_i = x(t_i)$ and $u_i = u(t_i)$. Now, the "center" controls $v_{ij} = u(t_{i-1} + jh/2p)$ for $j = 1, 2, \ldots, 2p - 1$ and for $i = 1, 2, \ldots, N$ must, be provided, as shown in Fig. 2 for $p = 3$. Then, using Eqs. (39–42) the states are integrated from t_{i-1} forward one step to $t_{i-1} + h/p$ with controls u_{i-1}, v_{i1}, and v_{i2}. Using the resulting estimate of the state at $t_{i-1} + h/p$, the state integration is continued forward using the four-stage RK procedure and the controls v_{i2}, v_{i3}, and v_{i4} yielding the estimate of the state at $t_{i-1} + 2h/p$. The process is repeated once more using controls v_{i4}, v_{i5}, and u_i resulting in an estimate of the state, x_i^*, at node t_i which then replaces y_i^4 in the defect formula, Eq. (43). Note that the $(p - 1)$ intermediate estimates of the state vector do not appear explicitly in the NLP problem; this is the source of the great economy in NLP variables which results from the parallel-shooting approach.

It is interesting to note that the parallel-shooting approach is somewhere between direct-transcription methods and the state elimination approach recently taken by Kelley and Sachs.[15] If only one integration step per segment is used (with many segments), the method becomes direct transcription. If only one segment is used (with many integration steps), the method is state elimination. Sensitivity problems may appear as the segments get large; however, the parallel-shooting approach allows one to control accuracy and robustness independently.[1] The most efficient strategy, as Enright and Conway[1] point out, is to use as many segments as necessary to yield robust behavior and to use as many integration steps per segment to achieve the required accuracy.

The RK defects constitute a set of nonlinear "defect" equations. These defect equations become the nonlinear constraints in the NLP problem. Collecting all the independent variables into a single vector $\boldsymbol{P}$ defined as

$$\boldsymbol{P}^T = [\boldsymbol{Z}^T] \tag{44}$$

where

$$\boldsymbol{Z}^T = \left(\boldsymbol{x}_1^T, \boldsymbol{u}_1^T, \boldsymbol{x}_2^T, \boldsymbol{u}_2^T, \ldots, \boldsymbol{x}_{N+1}^T, \boldsymbol{u}_{N+1}^T, t_f\right)$$

and where, in this problem, the cost function is the final time t_f (the last parameter in the vector $\boldsymbol{Z}$), and similarly collecting all the nonlinear constraints into a vector $\boldsymbol{C}^T$, the optimal control problem can then be restated as an NLP problem of the form

$$\text{Minimize } \varphi(\boldsymbol{P}) \qquad \text{subject to} \qquad \boldsymbol{b}_L \le \begin{Bmatrix} \boldsymbol{P} \\ \boldsymbol{AP} \\ \boldsymbol{C}(\boldsymbol{P}) \end{Bmatrix} \le \boldsymbol{b}_U$$

where $\boldsymbol{AP}$ contains all the linear relationships of the stated problem and $\boldsymbol{b}_L$ and $\boldsymbol{b}_U$ are the lower and upper bounds on the variables and constraints.[4]

The direct-transcription methods, including this RK parallel-shooting method, do not use the calculus of variations first-order necessary conditions, i.e., the Euler-Lagrange equations of the continuous problem.[16] However, Enright[3] and Enright and Conway[1] have shown that, at the solution of the NLP problem, the Lagrange multipliers (or adjoint variables) corresponding to the defects are a discrete approximation to the adjoint variables of the continuous optimal control problem. It is possible to exploit this information to assess the adequacy of the discrete-approximation solution.

One simple test is to extract the adjoints at the final time, as well as the states, and to numerically integrate the state differential equations (7–10) and adjoint differential equations,[16]

$$\dot{\lambda}^T = -\frac{\partial H}{\partial x} \tag{45}$$

backward to the initial time, using the control generated by the control optimality condition $\partial H/\partial u = 0$, where H is the system Hamiltonian

$$H = \lambda_a^T \frac{\mathrm{d}a}{\mathrm{d}t} + \lambda_{P_1}^T \frac{\mathrm{d}P_1}{\mathrm{d}t} + \lambda_{P_2}^T \frac{\mathrm{d}P_2}{\mathrm{d}t} + \lambda_l^T \frac{\mathrm{d}l}{\mathrm{d}t} + \lambda_\alpha^T \frac{\mathrm{d}\alpha}{\mathrm{d}t} \tag{46}$$

If the initial state $\boldsymbol{x}_0$ is accurately recovered, then the Euler-Lagrange TPBVP has been accurately solved, thus verifying the optimality as well as the feasibility of the discrete-approximate solution. Optimality can further be checked by evaluating the system Hamiltonian (46) throughout the trajectory, which for this minimum-time problem with autonomous governing equations should be constant and equal to -1.

The nonlinear programming code used for this work is NZSOL, an improved version of the program NPSOL.[17] Subroutines coded in FORTRAN-77 were used to evaluate the objective function and its gradient as well as the constraints and their Jacobian. The derivatives required for the constraint Jacobian were calculated using finite differencing. This information was then supplied to the NZSOL program, along with a required initial guess of the discrete solution for the state and control variables, via the main program. These programs were then executed on a CRAY Y-MP supercomputer.

Programming Considerations

An initial guess of the solution (i.e., all state and control variables at all discrete times as well as the final time) must be provided to NZSOL. Boltz[18] has shown, for constant thrusting in the tangential direction at a thrust-to-weight ratio less than a critical factor, that the trajectory oscillates around a logarithmic spiral. Using this approximation, i.e., a thrust angle of 0 deg, a thrust acceleration can be found that requires 100 revolutions to reach a prescribed final semimajor axis, thus generating a nominal initial guess of the solution for the use of the NLP code. (Since the eccentricity for the Boltz-type solutions is always small, the approximation $P_1 = P_2 = \text{const} = 0$ was assumed for the initial guess.) For large number of revolution, LEO-to-GEO transfer, much of the change in the semimajor axis happens in the last few revolutions out of the hundreds of revolutions required for the transfer. A revolution near LEO requires much less time than a revolution at GEO. Thus, for equal time segments,

a segment at the beginning of the trajectory (near LEO) may represent several revolutions,whreas toward the end of the trajectory (near GEO) each segment may represent only a small fraction of the revolution. With the use of a constant number of parallel-shooting steps per segment, there will be a large disparity in the step sizes in mean longitude. Since true longitude is found indirectly from mean longitude, and all of the governing EOMs are a function of true longitude, the potential for serious difficulties in the discretization of the problem exists. The solution to this potential problem was to use unequal time segments, yielding step sizes of near equal true longitude, obtained by interpolation using information provided by the initial guess of the solution, thus allowing for a more consistent discretization of the problem.

In the determination of true longitude from mean longitude, it is necessary to solve Kepler's equation (19). The Laguerre-Conway method is used to solve Kepler's equation, and convergence is obtained for any choice of starting approximation.[19,20]

In the example optimal trajectories described in the next section, 60 time segments were used ($N = 60$), with six parallel-shooting steps per segment ($p = 6$). The total number of variables is 1027 [$(N + 1)$ nodes $\times$ (5 states + 1 control) and 60 segments $\times$ $(2p - 1)$ interior controls + final time t_f], and the total number of constraints is 300(N segments $\times$ 5 defects per segment). The execution time for the various cases ranged between 8 and 51 min, with all but two cases rquiring less than 25 min. Four of the eight cases required less than 15 min.

The solution strategy was to implement an upper limit of 30 major iterations for the NLP code. If the optimal solution was not found in 30 iterations, the final values for all the states and controls were retained upon exit from the NLP solver. The NLP problem could then be restarted using this point as an initial guess. This is referred to as "cold starting."[17] An alternative upon reaching the maximum number of iterations is to retain not only the final values for the states and controls but also the NLP algorithm matrices ISTATE, R, and CLAMDA as well. The ISTATE array corresponds to the bounds and linear constraints that define the initial working set for the procedure used to determine feasible points; the R matrix contains the upper-triangular Cholesky factor of the initial Hessian approximation; and CLAMDA contains a multiplier estimate for each nonlinear constraint17. Restarting the NLP problem using this additional information is referred to as "warm starting."[17] As a result, NZSOL need not compute the multipliers; i.e., the Lagrange multipliers conjugate to the NLP constraints, anew, thus improving the likelihood of convergence.

The scaling of the NLP variables is an important issue. As Hargraves and Paris[4] point out, convergence of the NLP problem can be improved if the variables are scaled so that they all have comparable magnitudes. Since the thrust accelerations used in this work are of the order of 10^{-4}–10^{-6}, the mean longitude is of the order of 10^2, and the final time can be of the order of 10^4 or more, scaling problems were considered likely. Thus, scaling was used in this work to yield variables of comparable magnitude. This was done manually by reading in scale factors for the semimajor axis, P_1, P_2, thrust acceleration, mean longitude, and control variables. Scale factors were chosen to yield values between -1 and 1 for the NLP parameters.

Example Optimal Trajectories

Two types of coplanar orbit raising transfers are now described. The two cases are GEO-to-outer-circular-orbit transfer and LEO-to-GEO transfer. Both initial and final orbits are constrained to be circular, though there is nothing in the preceding analysis that requires circular terminal orbits, and the transfer takes place in the equatorial plane. The optimal trajectories are determined with lunar attraction and Earth oblateness perturbations both neglected and included, so that the effect of perturbations may be determined. Only two examples are presented; in both cases the thrust acceleration is chosen so that approximately 100 orbits of Earth will be required to achieve the desired final orbit. The choice of 100 orbits is entirely arbitrary but is clearly sufficient to be characterized as a "very many" revolution transfer. The intention in these examples is not to provide a comprehensive description of low-thrust orbit raising; the space of initial and final orbits and thrust levels is too large to allow this. Rather these examples simply demonstrate that the method we have applied to the solution of the problem is effective and robust.

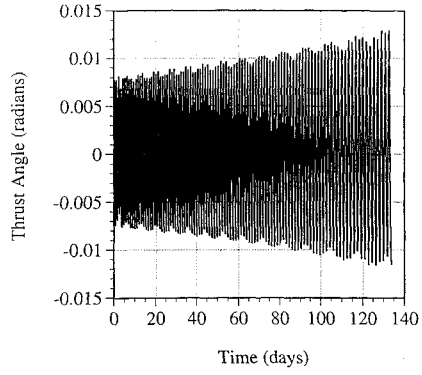

Fig. 3 GEO-to-outer orbit transfer optimal thrust angle history.

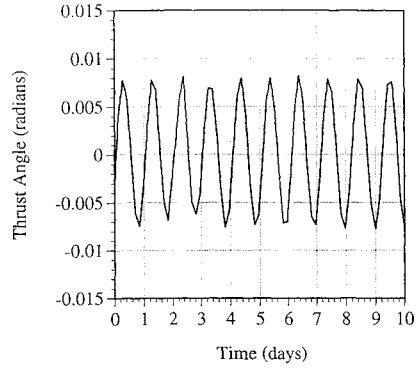

Fig. 4 GEO-to-outer orbit transfer optimal thrust angle history (near GEO).

Orbit Raising from GEO

In this problem, the spacecraft starts in a (circular) geosynchronous orbit and is required to transfer to another coplanar circular orbit of specified radius in an unspecified final time. In canonical units, the semimajor axis at GEO is 6.63 DU. The final semimajor axis was arbitrarily specified to be 10 DU. The initial true longitude of the spacecraft's orbit was chosen to be 0 deg, and because the initial orbit is circular, the initial mean longitude is equal to the initial true longitude. The initial thrust acceleration is 4.95μg. In Eq. (23), the time rate of change of thrust acceleration, the effective exhaust velocity is required. Stationary plasma thrusters, a form of electrical propulsion, are being developed with specific impulses (I_{sp}) between 1500 and 2000 s.[21] Using the relation

$$I_{sp} = \frac{c}{g} \tag{47}$$

yields a range for the exhaust velocity between 14.715 and 19.620 km/s or between 1.86 and 2.48 in normalized units. A value of 2.0 has been chosen for the exhaust velocity. Thus both thrust acceleration and thruster efficiency used in this analysis are representative of current capability.

For this choice of thrust acceleration and exhaust velocity the total transfer time, which was minimized, was 133.7 days. Figure 3 shows the optimal control history for this transfer, which required 100.8 revolutions. The control angle varies sinusoidally about a zero mean and has a constantly increasing amplitude. The thrust angle remains small, indicating the thrust is nearly all tangential. Using an initial guess of purely tangential thrusting was thus very reasonable. Figure 4 shows the thrust history for the first 10 days for the transfer, which equates to 9.7 revolutions.

Figure 5 shows that the optimal transfer orbit does not deviate greatly from a circular orbit. This is reasonable, since the thrust is nearly all tangential and is very small in comparison to the local gravity. The thrust acceleration (not shown) varies almost linearly

Table 1 Results of posteriori backward integration of system equations for GEO-to-outer-orbit transfer

State	Value at $t_f = 133.7$ days	Value at $t_i = 0$ days (from integration)	Exact value at $t_i = 0$ days
a, DU	10.0	6.62998	6.63
P_1	0.0	-1.291×10^{-5}	0.0
P_2	0.0	-1.378×10^{-5}	0.0
l, rads	633.62	-1.74×10^{-3}	0.0
α, g	5.1318×10^{-6}	4.95×10^{-6}	4.95×10^{-6}

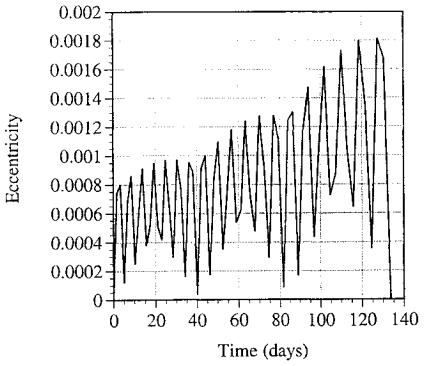

Fig. 5 GEO-to-outer-orbit transfer eccentricity history.

from 4.95 μg at the beginning of the transfer to its final value of 5.132 μg. This nearly constant rate of variation of the thrust acceleration during the orbit raising is what would be expected from Eq. (23) with such a small initial value of α.

The discrete solution obtained by nonlinear programming can now be verified by checking the necessary conditions for the original optimal control problem. Because the Lagrange multipliers approximate the adjoints at the nodes for the RK parallel-shooting method, these values were available directly from the NLP problem solution output. Using a continuous control history approximated with piecewise continuous cubics fitted through the discrete controls, the Euler-Lagrange equations, i.e., the system EOMs and adjoint equations, were integrated backward in time. The system Hamiltonian, as well as the control optimality condition, were evaluated at the discrete points, i.e., the boundaries of the nodes. The initial states were reacquired within an acceptable margin, as shown in Table 1. As mentioned previously, the Hamiltonian should have a value of -1 on the optimal trajectory. Figure 6 shows the variation of the Hamiltonian for the trajectory. The optimality condition $\partial H/\partial u = 0$ is similarly well satisfied. It was verified that as smaller step sizes are used, the accuracy of the solution improves. For the GEO-to-outer-orbit case presented, 60 segments and six RK steps per segment were used. This yields discrete controls that are roughly 50 deg apart in mean longitude.

The optimization problem was solved for the same initial conditions but with an additional perturbation modeled, that of Earth oblateness due to the spherical harmonic J_2 (through first order). The resulting time of flight increased by 0.35 time units (282 s) from the 14314.5 time units of the first case, a negligible change. The resulting time histories of the thrust angle, semimajor axis, eccentricity, thrust acceleration, and true longitude were indistinguishable from those of the case without oblateness and for the sake of brevity are not shown. Thus, with motion constrained to the equatorial plane, the Earth oblateness effect on orbit raising transfers from GEO that require hundreds of revolutions is negligible.

The optimization problem was similarly solved for the same initial conditions, but with an additional perturbation, lunar attraction, included. The initial position of the moon was chosen arbitrarily to be 25 deg east of the first point of Aries (and 25 deg east of the spacecraft, which has an initial true longitude of 0 deg). The resulting minimum time of flight increased from that of the unperturbed case by 3.1 time units (41.7 min). Again, there is no distinguishable difference (to the eye) between the time histories for the control

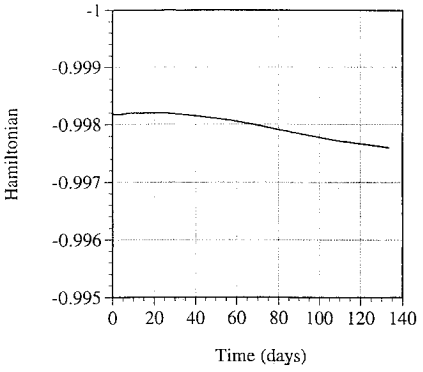

Fig. 6 Variation of system Hamiltonian during a posteriori backward integration of system equations for GEO to outer orbit transfer.

and states from the corresponding results for the unperturbed case. Although the spacecraft makes 100.86 revolutions in 133.73 days, the moon makes 4.8 revolutions. A partial explanation of the insignificance of the lunar perturbation is that there is no prevalent geometry between the Earth-spacecraft-moon; the geometry is continually changing over the transfer of the spacecraft and the period of the moon.

When oblateness perturbations and lunar perturbations are simultaneously modeled, the time of flight increases from that of the unperturbed case by 3.6 time units (52.2 min). Again, there is no distinguishable difference to the eye between the time histories of the control and states from those of the unperturbed case.

LEO-to-GEO Transfer

The spacecraft now starts in a circular LEO and is required to transfer to a coplanar (circular) GEO in an unspecified final time. The initial altitude in LEO is chosen to be 270 km. Thus, in canonical units, the initial semimajor axis at LEO is 1.0423 DU and the final semimajor axis at GEO will be 6.63 DU. As in the previously described cases, the spacecraft's true longitude is 0 deg at the initial time. The thrust acceleration is now 325 μg, determined a priori to yield approximately 100 revolutions to reach GEO, and the exhaust velocity of 2.0 in normalized units is used again. The resulting minimum transfer time was 14.75 days. Figure 7 shows the optimal control history for this transfer, which requires 100.6 revolutions. Figure 8 shows the optimal control history near GEO.

The control history is similar in form to that of the previous transfers from GEO to outer circular orbit. However, although the thrust direction was almost purely tangential in the GEO-to-outer-orbit cases, the thrust angle becomes quite large toward the end of the LEO-to-GEO trajectory. The explanation for this is straightforward. At lower altitudes the spacecraft burns almost in the tangential direction in order to get out of the strong "gravity well" at the LEO altitudes. The larger thrust angle toward the end of the trajectory is to circularize the spacecraft into its terminal orbit. The eccentricity time history in Fig. 9 shows that the transfer trajectory deviates significantly more from circularity than the GEO-to-outer-circular-orbit transfers did. Although the largest eccentricity for this transfer is only 0.04, indicating the transfer orbit is always nearly circular, the magnitude of this largest eccentricity is approximately 25 times greater than the largest eccentricity for the GEO-to-outer-orbit transfers. The thrust acceleration is significantly larger than that used in the previous example so there is correspondingly greater change in the magnitude of the thrust acceleration, from 325 to 437 μg, over the 100 revolution trajectory. Verification of the optimal control problem necessary conditions for the discrete solution was again performed. This trajectory also proved to be optimal, with the initial states reacquired with only small error, as shown in Table 2. The system Hamiltonian was again nearly constant at -1; its maximum deviation was to a value of -0.995. The optimality condition $\partial H/\partial u = 0$ was similarly satisfied. Again, 60 segments and six RK steps per segment were used, yielding discrete controls roughly 50° apart in mean longitude. The satisfaction of all the necessary conditions tested, to four significant figures, indicates the adequacy of this discretization.

Table 2 Results of posteriori backward integration of system equations for LEO-to-GEO transfer

State	Value at $t_f = 14.75$ days	Value at $t_i = 0$ days (from integration)	Exact value at $t_i = 0$ days
a, DU	6.63	1.0421	1.0423
P_1	0.0	-1.34×10^{-3}	0.0
P_2	0.0	-2.47×10^{-3}	0.0
l, rads	632.17	-1.26×10^{-3}	0.0
α, g	4.3718×10^{-4}	3.25×10^{-4}	3.25×10^{-4}

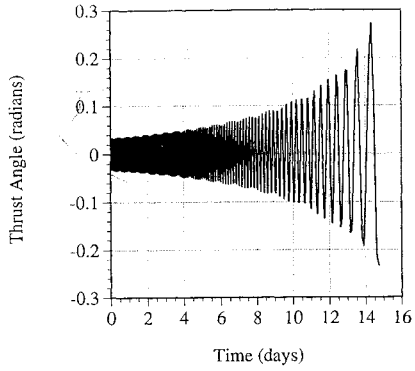

Fig. 7 LEO-to-GEO transfer optimal thrust angle history.

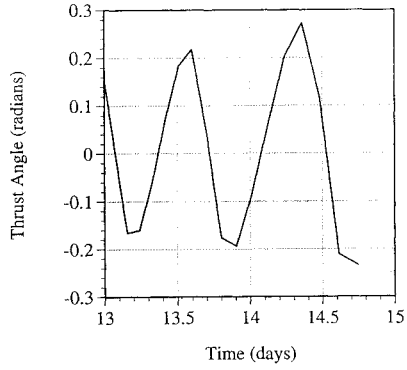

Fig. 8 LEO-to-GEO transfer optimal thrust angle history (near GEO).

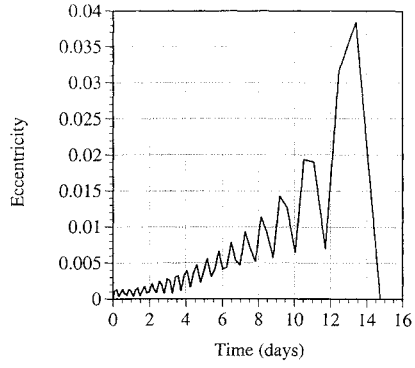

Fig. 9 LEO-to-GEO transfer eccentricity history.

The inclusion of Earth oblateness due to the spherical harmonic J_2 (through first order) increased the time of flight by 0.09 time units (72.6 s) from the 1579.16 time units of the unperturbed LEO-to-GEO case. Comparing the resulting control angle history of Fig. 10 with that for the case without consideration of oblateness (Fig. 7) reveals a significant difference. This is not unexpected since in low Earth orbit the J_2 contribution to Earth's potential is not negligible, as it is at geosynchronous altitude. Near LEO, with the Earth oblateness modeled through J_2, the thrust angle is three to four times as large as that for the case when J_2 is not included. When the spacecraft nears GEO altitudes, comparison of Figs. 7–10 shows that the optimal thrust angle is not as large as that required when J_2 is ignored.

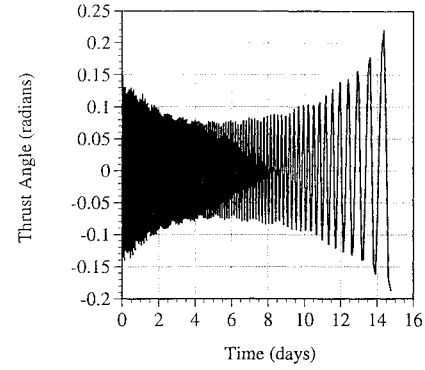

Fig. 10 LEO-to-GEO transfer optimal thrust angle history (with J_2 modeled).

The resulting time histories for the semimajor axis, thrust acceleration, and true longitude were indistinguishable by eye from the case in which J_2 is ignored and for the sake of brevity are not shown. In contrast to the orbit raising cases from GEO to outer orbits, modeling the Earth's oblateness yields nonnegligible changes in the LEO-to-GEO optimal trajectory. Although the transfer time is essentially the same, the optimal control history to minimize flight time is considerably different for the initial part of the trajectory (near LEO). This is entirely reasonable because of the strong dependence of the oblateness perturbation on orbit radius.

When lunar attraction is considered, either alone or in combination with the oblateness perturbation, the time of flight is increased insignificantly, i.e., by less than 1.1 min, from the corresponding case without consideration of the moon. The thrust angle time histories are also insignificantly changed. This is not surprising considering how much the lunar perturbation is reduced in magnitude for this case where the spacecraft spends most of its time in low Earth orbit.

Conclusion

This work is not intended to be a comprehensive discussion of low-thrust orbit raising. The two example cases presented (GEO to outer circular orbit and LEO to GEO) were intended to demonstrate that the RK parallel-shooting transcription and nonlinear programming method can be successfully applied to the optimization of many-revolution orbit raising trajectories about Earth. These problems would not have been tractable using the DCLNP method discussed earlier (i.e., that of Hargraves and Paris[4]) due to the large size of the resulting (NLP) problem. The size of the example problems, 1027 independent variables, coincides with the practical upper limit of the NLP solver NZSOL. With the initial guess determined as previously described, the optimization code produced for this work has proven to be robust, provided a sufficient number of major iterations for the NLP solver were allowed, and the solution does not have to be restarted.

For the admittedly very small number of example problems, including the lunar attraction for the spacecraft not unexpectedly produced negligible differences in both the state and control time histories and in total flight time. Including the Earth's oblateness effects through first order (J_2) produced negligible differences in both the state and control histories and in total flight time for the cases originating in GEO. However, for the LEO-to-GEO transfers, including the Earth's oblateness produced considerable differences in the optimal control history and the eccentricity history of the transfer trajectory.

The assumption that the spacecraft orbit (and the lunar orbit!) are always in the equatorial plane will be dispensed with in work to follow. The extension to noncoplanar trajectories is straightforward, and only two more differential equations need to be added to the governing equations. Including the time rates of change for the equinoctial elements Q_1 and Q_2 will allow three-dimensional trajectories. The effect of the Earth's oblateness will now involve more terms, since $i \neq 0$ deg, and similarly, the lunar perturbation will need to be resolved into the three-dimensional coordinate system at the spacecraft (radial, tangential, and normal). This will increase the size of the NLP problem, since two more states per segment

are introduced and an additional, out-of-plane thrust angle variable is necessary at all discrete points. The size of the NLP problem in this work is about the maximum the NLP code can handle. Thus, three-dimensional trajectories solved for all need to use fewer revolutions, i.e., need to use larger thrust accelerations due to the large increase in the number of NLP variables. Solution of such a problem will be considerably aided by the availability, in the near future, of NLP problem solvers that exploit the sparse Jacobian matrix structure that is a feature of the NLP problems resulting from the use of direct-transcription methods.[3,4]

References

[1]Enright, P. J., and Conway, B. A., "Discrete Approximations to Optimal Trajectories Using Direct Transcription and Nonlinear Programming," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 4, 1992, pp. 994–1002.

[2]Redding, D. C., and Breakwell, J. V., "Optimal Low-Thrust Transfers to Synchronous Orbit," *Journal of Guidance, Control, and Dynamics*, Vol. 7, No. 2, 1984, pp. 148–155.

[3]Enright, P. J., "Optimal Finite-Thrust Spacecraft Trajectories Using Direct Transcription and Nonlinear Programming," Ph.D. Dissertation, Univ. of Illinois, Urbana, IL, 1991.

[4]Hargraves, C. R., and Paris, S. W., "Direct Trajectory Optimization Using Nonlinear Programming and Collocation," *Journal of Guidance, Control, and Dynamics*, Vol. 10, No. 4, 1987, pp. 338–342.

[5]Dickmanns, E. D., and Well, K. H., "Approximate Solution of Optimal Control Problems Using Third-Order Hermite Polynomial Functions," *Proceedings for the Sixth Technical Conference on Optimization Techniques*, Springer-Verlag, New York, 1975.

[6]Russell, R. D., and Shampine, L. F., "A Collocation Method for Boundary Value Problems," *Numerical Mathematics*, Vol. 19, 1972, pp. 1–28.

[7]Enright, P. J., and Conway, B. A., "Optimal Finite-Thrust Spacecraft Trajectories Using Collocation and Nonlinear Programming," *Journal of Guidance, Control, and Dynamics*, Vol. 14, No. 5, 1991, pp. 981–985.

[8]Tang, S., "Optimization of Interplanetary Trajectories Using Direct Collocation and Nonlinear Programming," M.S. Thesis, Univ. of Illinois, Urbana, IL, 1992.

[9]Herman, A. L., "Optimal Spacecraft Attitude Control Using Nonlinear Programming," M.S. Thesis, Univ. of Illinois, Urbana, IL, 1989.

[10]Downey, J. R., and Conway, B. A., "Optimal Finite-Thrust Time-Limited Direct-Ascent Interception," AIAA Paper 92-4512, Aug. 1992.

[11]Arsenault, J. L., Ford, K. C., and Koskela, P. E., "Orbit Determination Using Analytic Partial Derivatives of Perturbed Motion," *AIAA Journal*, Vol. 8, 1970, pp. 4–12.

[12]Broucke, R. A., and Cefola, P. J., "On the Equinoctial Elements," *Celestial Mechanics*, Vol. 5, No. 3, 1972, pp. 303–310.

[13]Battin, R. H., *An Introduction to the Mathematics and Methods of Astrodynamics*, AIAA, New York, 1987.

[14]Nelson, R. L., and Gill, P. E., "Optimal Trajectories by Implicit Simulation and NPSOL: Software for Constrained Optimization," presentation at Wright Patterson Air Force Base, Dayton, OH, April 1993.

[15]Kelley, C. T., and Sachs, E .W., "Quasi-Newton Methods and Unconstrained Optimal Control Problems," *SIAM Journal of Control and Optimization*, Vol. 25, No. 6, 1987, pp. 1503–1516.

[16]Bryson, A. E., and Ho, Y. C., *Applied Optimal Control*, Hemisphere, New York, 1975.

[17]Gill, P. E., Murray, W., Saunders, M. A., and Wright, M. H., *User's Guide for NPSOL (Version 4.0): A FORTRAN Package for Nonlinear Programming*, Stanford Univ., Stanford, CA, Jan. 1986.

[18]Boltz, F. W., "Orbital Motion Under Continuous Tangential Thrust," *Journal of Guidance, Control, and Dynamics*, Vol. 15, No. 6, 1992, pp. 1503–1507.

[19]Prussing, J. E., and Conway, B. A., *Orbital Mechanics*, Oxford Univ. Press, New York, 1993.

[20]Danby, J. M. A., "The Solution of Kepler's Equation, III," *Celestial Mechanics*, Vol. 40, 1987, pp. 303–312.

[21]Turchi, P. J., Curran, F. M., Andrews, J. C., Beattie, J. R., and Gilland, J., "Electric Propulsion: The Future is Now," *Aerospace America*, Vol. 30, No. 7, July 1992, pp. 38–41.