JOURNAL OF GUIDANCE, CONTROL, AND DYNAMICS
Vol. 35, No. 1, January–February 2012

Practical Techniques for Low-Thrust Trajectory Optimization with Homotopic Approach

Fanghua Jiang,* Hexi Baoyin,† and Junfeng Li‡
Tsinghua University, 100084 Beijing, People's Republic of China

DOI: 10.2514/1.52476

This paper concerns the application of the homotopic approach, which solves the fuel-optimal problem of low-thrust trajectory by starting from the related and easier energy-optimal problem. To this end, some effective techniques are presented to reduce the computational time and increase the probability of finding the globally optimal solution. First, the optimal control problem is made homogeneous to the Lagrange multipliers by multiplying the performance index by a positive unknown factor. Hence, normalization is applicable to restrict the unknown multipliers on a unit hypersphere. Second, the switching function's first- and second-order derivatives with respect to time are derived to detect switching. The switching detection is embedded in the fourth-order Runge–Kutta algorithm with fixed step size to ensure integration accuracy for bang-bang control. Third, combined with the techniques of normalization and switching detection, the particle swarm optimization with well-chosen parameters considerably increases the probability of finding the approximate initial values of the globally optimal solution. Moreover, intermediate gravity assist, which brings complex inner constraints, is considered. To determine the approximate gravity assist date, analytical formulas are presented to evaluate the minimal maneuver impulse based on the results of Lambert problems. The first-order necessary conditions for gravity assist constraints are derived analytically. The optimal solution can be rapidly obtained by applying the techniques presented to solve the shooting function. The unknowns are far less than with direct methods, and the computational effort is also far lower. Two examples of fuel-optimal rendezvous problems from the Earth directly to Venus and from the Earth to Jupiter via Mars gravity assist are given to substantiate the perfect efficiency of these techniques.

Nomenclature

$\boldsymbol{a}$	=	acceleration vector in heliocentric ecliptic reference frame
g_0	=	standard acceleration of gravity at sea level
H	=	homotopy
H	=	Hamiltonian
h	=	fixed step size of integration
I_{sp}	=	thruster specific impulse
J	=	performance index
m	=	nondimensionalized mass
p_f	=	quadratic penalty factor
$\boldsymbol{r}$	=	position vector in heliocentric ecliptic reference frame
$r_{\min}$	=	minimal admissible periapsis radius of planetary gravity assist
r_p	=	periapsis radius of planetary gravity assist
$T_{\max}$	=	thrust magnitude
$\boldsymbol{U}$	=	general control vector
u	=	engine thrust ratio
$\boldsymbol{v}$	=	velocity vector in heliocentric ecliptic reference frame
$\boldsymbol{v}_\infty$	=	hyperbolic excess speed
$\boldsymbol{X}$	=	search-variable vector, each component with domain [0, 1]
$\boldsymbol{x}$	=	general state vector
$\boldsymbol{z}$	=	shooting unknown
$\boldsymbol{\alpha}$	=	unit vector of thrust direction
δ	=	variation operation
ε	=	homotopy parameter with domain [0, 1]
θ	=	tuning angle of hyperbolic excess speed for planetary gravity assist
$\boldsymbol{\kappa}$	=	numerical nonnegative multiplier associated with inequality constraints
$\boldsymbol{\lambda}$	=	Lagrange multiplier associated with state, i.e., costate
λ_0	=	positive factor acting on performance index
ρ	=	switching function
$\boldsymbol{\sigma}$	=	combination of inequality constraints
$\boldsymbol{\chi}$	=	numerical multiplier associated with equality constraints
$\boldsymbol{\Phi}$	=	combination of shooting functions
$\boldsymbol{\psi}$	=	combination of equality constraints

Subscripts

f	=	final time
m	=	intermediate event date
0	=	initial time

Superscripts

$-$	=	the date just before gravity assist
$+$	=	the date just after gravity assist

I. Introduction

ELECTRIC propulsion has been attracting much attention since its first use in the mission of Deep Space 1 [1] and has won significant appeal by the Japanese mission named HAYABUSA [2] in 2010. It is able to give higher specific impulse and thus is more efficient compared with traditional chemical propulsion. Combined with the technique of gravity assist (GA) applied in Mariner 10, Voyager 1, and others, electric propulsion is one of the viable options for future interplanetary missions. However, as for mission designers, electric propulsion is more challenging than traditional chemical propulsion mainly due to the fact that the former enables continuous and durable low thrust. Therefore, it is rather difficult to find the optimal trajectory for a mission with electric propulsion. To

Received 20 September 2010; revision received 24 June 2011; accepted for publication 26 June 2011.

*Assistant Researcher, School of Aerospace; jiangfh04@mails.thu.edu.cn.
†Associate Professor, School of Aerospace; baoyin@tsinghua.edu.cn.
‡Professor, School of Aerospace; lijunf@tsinghua.edu.cn.

overcome such an obstacle, lots of research has been conducted on low-thrust trajectory design that inevitably resorts to optimization.

In general, the methods under consideration can be categorized as direct methods [3,4] and indirect methods [5,6], the combinations of which are usually termed hybrid methods [7,8]. Direct methods convert the original optimal control problem into a parameter optimization problem through discretization and then resort to nonlinear programming to find the optimal solution. They usually require a large amount of computational effort. Indirect methods are always involved in solving a two-point boundary-value problem (TPBVP) derived from the calculus of variations and Pontryagin's maximum principle (PMP). The solution to the TPBVP is difficult to obtain because of the small convergence radius and the sensitivity of the initial guesses. Hybrid methods combine the advantages of both direct and indirect methods. The thrust structure is usually assumed a priori, and the optimal control is determined by the optimality condition involved in indirect methods. This paper concerns only the indirect method in a modern context.

To the authors' knowledge, there is less literature available on indirect methods than on direct methods. The PMP indicates that, generally, the fuel-optimal control is bang-bang control (exclude singular cases). If the switching dates are not known in advance, it is very difficult to solve such a problem, due to a number of difficulties, such as small convergence radius, discontinuous integrated functions, and the singular Jacobian matrix [9,10]. Accordingly, most of the literature on indirect methods resorted to a solution approaching the optimal one, including assigning a priori the switching structure [10–12], using a fuel consumption model that leads to continuous optimal control [13,14], and reducing the dimension of the problem considered [15]. Assigning the switching structure has the risk of missing the optimal solution, though it may be overcome by enumeration when the problem is inherently of a simple optimal switching structure. The continuous optimal control reduces resolution difficulty greatly. The PMP indicates that the velocity costate vector, termed as the primer vector [16], must lie along the opposite direction of the optimal thrust direction. Russell [17] applied primer vector theory to search the global solution through parallel computing.

Recently, the homotopic approach has been applied to solve bang-bang control problems [9,18–20] and singular optimal control problems [21,22] (opposite to bang-bang control, the thrust magnitude is neither zero nor maximal). The principle of a homotopic approach is to solve a difficult problem by starting from the solution of a somewhat related, but easier problem. A main advantage of the homotopic approach is that the control structure is not assumed a priori but determined automatically. For a fuel-optimal problem, Bertrand and Epenoy [9] and Haberkorn et al. [18] constructed the easier problem with the performance index of the integral of the quadratic of thrust magnitude, which is called the energy-optimal problem. It implies continuous optimal control and thus has a relatively large convergence radius. Moreover, perturbing the performance index by logarithmic terms even implies differentiable optimal controls. The fuel-optimal problem (the parameter is equal to zero) is connected with the energy-optimal problem (the parameter equals one) through a perturbation parameter within the domain [0, 1]. Once the energy-optimal problem has been solved, the fuel-optimal problem can be solved by continuously decreasing the perturbation parameter from one to zero and taking the obtained solution as an initial guess for the next iteration. Accordingly, the homotopic approach is also known as the continuation method. Haberkorn et al. [18] presented several ways to decrease the parameter and ensure convergence at the same time. Some ways are difficult to carry out. The Jacobian matrix of the shooting function can be approximately computed by finite differences or precisely by integrating the variational equations presented by Martinon and Gergaud [19].

The goal of this paper is to develop a suite of easy-to-use and effective techniques applied in the homotopic approach to reduce the computational time spent on finding out the global solution. So far, the initial costates are still guessed within an unbounded space. In fact, normalization to the initial costates can be made to benefit initial guesses. The probability of finding out the global solution is enlarged greatly through searching the initial values by the particle swarm optimization [23] (PSO). The PSO algorithm is generally applied to trajectory optimization with a multiple impulse [24]. The application to low-thrust instances is few. Pontani and Conway [25] summarized literature both on impulse and low-thrust applications and studied the applications to fuel-optimal finite thrust transfer and to time-optimal low-thrust transfer under several assumptions. The residuals of the constraints added to the objective function as penalty terms make it difficult to satisfy a high-accuracy requirement. This is not a problem for the application here, because of the fact that the PSO results are just used as approximate initial values. This paper will develop an a priori switching detection technique to replace Martinon's a posteriori technique. The first- and second-order derivatives of switching function (SF) with respect to time are derived and then applied in a fixed step integrator to predict the thrust trend. It is expected to be not only easier to implement but also more effective. The fuel-optimal problem with intermediate GA is considered. A pure mathematical method to analytically evaluate the minimal maneuver impulse based on the results of the Lambert problem will be presented. This is different from the model with the use of the physical B-plane angle [26]. The first-order necessary conditions (FONCs) of GA are derived and conducted analytically. Although other literature [10] derived the FONCs too, the results here are more complete and concise. The efficiency of these techniques will be substantiated by two examples on the rendezvous problems without and with intermediate GA.

The rest of this paper is organized as follows. There are five parts in Sec. II. Immediately after stating the fuel-optimal problem with fixed boundary conditions, the normalization method is presented. Subsequently, an integration rule of using the Runge–Kutta algorithm with fixed step and switching detection is constructed. Then, the PSO algorithm is introduced to globally search the approximate initial values. Finally, the solving strategy is summarized. In Sec. III, the method of evaluating the minimal maneuver impulse is presented, and then the FONCs of GA are derived analytically. Two examples are given in Sec. IV to support the efficiency of the techniques presented in Secs. II and III, respectively. Section V concludes this paper.

II. Techniques for Solving the Fuel-Optimal Problem

A. Fuel-Optimal Problem Statement

Let us consider the preliminary problem that the spacecraft is subject only to the central force of the sun's gravity and the thrust of its own electric propulsion system:

$$\dot{\boldsymbol{r}} = \boldsymbol{v}, \qquad \dot{\boldsymbol{v}} = -\frac{\mu}{r^3}\boldsymbol{r} + \frac{T_{\max}u}{m}\boldsymbol{\alpha}, \qquad \dot{m} = -\frac{T_{\max}u}{I_{\mathrm{sp}}g_0} \tag{1}$$

where $\boldsymbol{r}$ and $\boldsymbol{v}$ denote the position and velocity vectors in the heliocentric ecliptic reference frame (HERF), respectively, m is the instantaneous mass of spacecraft, $T_{\max}$ is the maximal thrust magnitude, I_{sp} is the thruster specific impulse, and g_0 is the standard acceleration of gravity at sea level, $9.80665\ \mathrm{m/s^2}$. The sun's gravitational constant is denoted by μ, $1.32712440018 \times 10^{11}\ \mathrm{km^3/s^2}$. The control variable consists of the unit vector of thrust direction $\boldsymbol{\alpha}$ and the engine thrust ratio $u \in [0, 1]$. For computing convenience, the quantities about length, time, and mass are nondimensionalized by the astronomical unit (AU, 149597870.66 km), year (yr, $365.25 \times 86,400$ s), and spacecraft initial mass m_0, respectively. Therefore, the new value of μ is $39.476926\ \mathrm{AU^3/yr^2}$. The quantities $T_{\max}$, I_{sp}, and g_0 are customarily in international units. In that case, their values should be transformed into those consistent with the nondimensionalized units. Generally, the optimal control problem is to maximize the final mass, or to minimize the fuel consumption (normalized by spacecraft initial mass) expressed by

$$J = \frac{T_{\max}}{I_{\mathrm{sp}}g_0}\int_{t_0}^{t_f} u\,\mathrm{d}t \tag{2}$$

where t_0 and t_f denote the initial and final times, respectively, both fixed. The initial mass, initial states, and final states are all fixed. That is to say, the following terminal constraints must be satisfied:

$$\boldsymbol{r}(t_0) = \boldsymbol{r}_0, \qquad \boldsymbol{v}(t_0) = \boldsymbol{v}_0, \qquad m(t_0) = 1 \tag{3}$$

$$\boldsymbol{r}(t_f) = \boldsymbol{r}_f, \qquad \boldsymbol{v}(t_f) = \boldsymbol{v}_f \tag{4}$$

The optimal control problem can be transformed into a TPBVP by the use of PMP. First, with reference to Eq. 4 of [9], by introducing the costate vector $\boldsymbol{\lambda} \triangleq (\boldsymbol{\lambda}_r; \boldsymbol{\lambda}_v; \lambda_m)$, which is known as the functional Lagrange multiplier, the Hamiltonian is built as

$$H = \boldsymbol{\lambda}_r \cdot \boldsymbol{v} + \boldsymbol{\lambda}_v \cdot \left(-\frac{\mu}{r^3}\boldsymbol{r} + \frac{T_{\max} u}{m}\boldsymbol{\alpha}\right) - \lambda_m \frac{T_{\max}}{I_{\mathrm{sp}} g_0} u + \frac{T_{\max}}{I_{\mathrm{sp}} g_0} u \tag{5}$$

Then, the optimal thrust direction and magnitude, which minimize the Hamiltonian, are determined by

$$\boldsymbol{\alpha} = -\frac{\boldsymbol{\lambda}_v}{\|\boldsymbol{\lambda}_v\|} \tag{6}$$

$$\begin{cases} u = 0 & \text{if } \rho > 0 \\ u = 1 & \text{if } \rho < 0 \\ u \in [0, 1] & \text{if } \rho = 0 \end{cases} \tag{7}$$

where the SF ρ holds the form

$$\rho = 1 - \frac{I_{\mathrm{sp}} g_0 \|\boldsymbol{\lambda}_v\|}{m} - \lambda_m \tag{8}$$

Generally, the SF ρ takes the value zero only at finite isolated points. So, the normalized optimal thrust magnitude u seldom takes the value of neither zero nor one on any interval. This attribute is called bang-bang. The costate differential equations that are termed as Euler–Lagrange equations are given as

$$\dot{\boldsymbol{\lambda}}_r = \frac{\mu}{r^3}\boldsymbol{\lambda}_v - \frac{3\mu \boldsymbol{r} \cdot \boldsymbol{\lambda}_v}{r^5}\boldsymbol{r}, \qquad \dot{\boldsymbol{\lambda}}_v = -\boldsymbol{\lambda}_r, \qquad \dot{\lambda}_m = -\frac{T_{\max} u}{m^2}\|\boldsymbol{\lambda}_v\| \tag{9}$$

Subject to the bang-bang control, the right-hand sides of differential Eqs. (1) and (9) are discontinuous. The transversality conditions tell that when the boundary states are fixed the corresponding boundary costates are free; when the former are free, the latter are zero. For a detailed derivation see [27]. So the final mass costate should be zero due to free final mass:

$$\lambda_m(t_f) = 0 \tag{10}$$

Given the initial costates $[\boldsymbol{\lambda}_r(t_0), \boldsymbol{\lambda}_v(t_0), \lambda_m(t_0)]$, and together with the initial states $[\boldsymbol{r}(t_0), \boldsymbol{v}(t_0), m(t_0)]$, the final states and costates can be obtained through integrating Eqs. (1) and (9). The final position and velocity vectors should satisfy Eq. (4), and the final mass costate should satisfy Eq. (10). Therefore, the optimal control problem yields a TPBVP consisting of a set of equations of the form

$$\boldsymbol{\Phi}(\boldsymbol{z}) = [\boldsymbol{r}(t_f) - \boldsymbol{r}_f, \boldsymbol{v}(t_f) - \boldsymbol{v}_f, \lambda_m(t_f)]^T = \boldsymbol{0} \tag{11}$$

where $\boldsymbol{\Phi}$ is called the shooting function, and $\boldsymbol{z} \triangleq [\boldsymbol{\lambda}_r(t_0); \boldsymbol{\lambda}_v(t_0); \lambda_m(t_0)]$ denotes the combination of unknown initial costates. However, because of nonsmoothness and discontinuous right-hand sides of ordinary differential equation (ODE)s, the shooting function is difficult to solve by single shooting methods such as Newton's or Powell's.

Bertrand and Epenoy [9] introduced smoothing techniques that involved the homotopic approach to overcome these difficulties. One manner to build the homotopy $H(\boldsymbol{z}, \varepsilon)$ is through perturbing the performance index as

$$J = \int_{t_0}^{t_f} [u - \varepsilon u(1-u)]\,\mathrm{d}t \tag{12}$$

where the parameter ε links the mass criterion ($\varepsilon = 0$) with the energy criterion ($\varepsilon = 1$). The energy-optimal problem $H(\boldsymbol{z}, 1)$ corresponds to continuous optimal control, which will be discussed in the next subsection. Compared with the fuel-optimal problem $H(\boldsymbol{z}, 0)$, the energy-optimal problem is easier to solve. After obtaining the solution of the energy problem, let the parameter ε decrease by following a certain series, and then take the solution obtained in the current iteration as an initial guess for that in the next one. Finally, the solution of the fuel-optimal problem can be obtained. However, there are still some deficiencies in this manner. First, the initial costates are guessed within an unbounded space. Second, when ε is close to zero, ordinary integration algorithms like Runge–Kutta with an adaptive step cannot ensure accuracy any longer because the right-hand sides of ODEs vary rapidly around switching points, though they are continuous in theory. Third, since only first-order conditions are considered, the convergent solutions are possibly local optima, though multistart techniques may overcome this deficiency more or less. Accordingly, some easy-to-use and effective techniques will be presented in the following three subsections to try overcoming these deficiencies.

B. Normalization of the Initial Costate Vector

Multiplying the homotopic performance index by a positive numerical factor λ_0 does not inherently change the optimal control problem. The new performance index is modified as

$$J = \lambda_0 \frac{T_{\max}}{I_{\mathrm{sp}} g_0} \int_{t_0}^{t_f} [u - \varepsilon u(1-u)]\,\mathrm{d}t \tag{13}$$

which is identical to physical fuel consumption (normalized by initial mass) when $\varepsilon = 0$. In theory, the numerical factor λ_0 is allowed to be arbitrarily positive. Here it is selected by satisfying the normalization condition that will be introduced next. For the new performance index, the Hamiltonian is built as

$$H = \boldsymbol{\lambda}_r \cdot \boldsymbol{v} + \boldsymbol{\lambda}_v \cdot \left(-\frac{\mu}{r^3}\boldsymbol{r} + \frac{T_{\max} u}{m}\boldsymbol{\alpha}\right) - \lambda_m \frac{T_{\max}}{I_{\mathrm{sp}} g_0} u + \lambda_0 \frac{T_{\max}}{I_{\mathrm{sp}} g_0}[u - \varepsilon u(1-u)] \tag{14}$$

The optimal control direction holds the same form as Eq. (6), while the magnitude becomes

$$\begin{cases} u = 0 & \text{if } \rho > \varepsilon \\ u = 1 & \text{if } \rho < -\varepsilon \\ u = \frac{1}{2} - \frac{\rho}{2\varepsilon} & \text{if } |\rho| \le \varepsilon \end{cases} \tag{15}$$

where the SF holds the form

$$\rho = 1 - \frac{I_{\mathrm{sp}} g_0 \|\boldsymbol{\lambda}_v\|}{\lambda_0 m} - \frac{\lambda_m}{\lambda_0} \tag{16}$$

Equations (15) indicate that optimal control is continuous only if $\varepsilon \neq 0$, though it is undifferentiatable at some isolated points. Anyway, the integration accuracy of using ordinary integrators such as Runge–Kutta with an adaptive step size is expected to be improved. The Euler–Lagrange equations remain the same as Eqs. (9). Under the same boundary constraints expressed by Eqs. (3) and (4), the shooting function remains the same as Eq. (11), too.

Let us regard the functional costates $[\boldsymbol{\lambda}_r(t); \boldsymbol{\lambda}_v(t); \lambda_m(t)]$ and numerical factor λ_0 together as Lagrange multipliers $\boldsymbol{\lambda} \triangleq [\lambda_0; \boldsymbol{\lambda}_r(t); \boldsymbol{\lambda}_v(t); \lambda_m(t)]$. Then we see that the optimal control problem, which is expressed by the performance index (13), the dynamic Eqs. (1), the Euler–Lagrange Eqs. (9), the optimal control conditions and Eqs. (6) and (15), the Hamiltonian Eq. (14), the SF Eq. (16), and the shooting function (11), is homogeneous to the Lagrange multipliers. That is to say, multiplying these multipliers by

a positive factor does not change the nature of this problem. Thereby, it is feasible to divide these multipliers by the Euclidean norm of their initial values. Define the new costates as

$$\boldsymbol{\lambda} \triangleq \frac{\boldsymbol{\lambda}}{\|\boldsymbol{\lambda}(t_0)\|} \tag{17}$$

Note that the numerical factor λ_0 does not vary with time. Obviously, the redefined multipliers, named as normalized multipliers, have the attribute

$$\|\boldsymbol{\lambda}(t_0)\| = 1 \tag{18}$$

Consequently, by adopting Eq. (18), the old shooting function with seven components is extended to a new one with eight components:

$$\boldsymbol{\Phi}(\boldsymbol{\lambda}(t_0)) = [\boldsymbol{r}(t_f) - \boldsymbol{r}_f, \boldsymbol{v}(t_f) - \boldsymbol{v}_f, \lambda_m(t_f), \|\boldsymbol{\lambda}(t_0)\| - 1]^T = \boldsymbol{0} \tag{19}$$

where the normalized multipliers with eight components include the numerical one acting on the performance index. Most importantly, through this normalization, the solution variables including seven initial costate values are restricted on a unit 8-D hypersphere. At the same time, all the formulas necessary to solve the shooting function remain in the same forms as before. Note that the condition $\lambda_0 > 0$ should be considered. Moreover, Eqs. (9) indicate that the derivative of λ_m with respect to time is nonpositive, and Eq. (10) says the final value of λ_m is zero, so it yields $\lambda_m(t_0) \geq 0$. The two conditions reduce the solution space further.

C. Integration with Fixed Step and Switching Detection

It has been mentioned that when the parameter ε of the homotopic problem is close to zero, Runge–Kutta methods with adaptive step size are not able to ensure the integration accuracy because the right-hand sides of ODEs vary rapidly around switching points. Here the fourth-order Runge–Kutta algorithm with fixed step size (RK4) combined with switching detection is chosen to try overcoming this problem. For an initial value problem

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(t, \boldsymbol{x}), \qquad t_0 \leq t \leq t_f, \qquad \boldsymbol{x}(t_0) = \boldsymbol{x}_0 \tag{20}$$

the RK4 iteration algorithm with step size h is well known:

$$\begin{aligned}
&\boldsymbol{x}_{k+1} = \boldsymbol{x}_k + \frac{h}{6}(\boldsymbol{k}_1 + 2\boldsymbol{k}_2 + 2\boldsymbol{k}_3 + \boldsymbol{k}_4)\\
&\boldsymbol{k}_1 = \boldsymbol{f}(t_k, \boldsymbol{x}_k) \qquad \boldsymbol{k}_2 = \boldsymbol{f}(t_k + h/2, \boldsymbol{x}_k + h/2\boldsymbol{k}_1)\\
&\boldsymbol{k}_3 = \boldsymbol{f}(t_k + h/2, \boldsymbol{x}_k + h/2\boldsymbol{k}_2)\\
&\boldsymbol{k}_4 = \boldsymbol{f}(t_k + h, \boldsymbol{x}_k + h\boldsymbol{k}_3)
\end{aligned} \tag{21}$$

For briefness, denote the one step iteration result $\boldsymbol{x}_{k+1}$ that starts from time t_k and state $\boldsymbol{x}_k$ with step h upon right function $\boldsymbol{f}(t, \boldsymbol{x})$ by

$$\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}, t_k, \boldsymbol{x}_k, h) \tag{22}$$

Hereafter, use $\boldsymbol{x}$ to denote the combined vector $[\boldsymbol{r}; \boldsymbol{v}; m; \boldsymbol{\lambda}_r; \boldsymbol{\lambda}_v; \lambda_m]$ with 14 components. Since the optimal control is entirely decided by the vector $\boldsymbol{x}$, the vector function $\boldsymbol{f}(t, \boldsymbol{x})$ is appropriate to denote the combination of the right-hand sides of the dynamic Eq. (1) and the Euler–Lagrange Eqs. (9). So, one has the form

$$\boldsymbol{f} = \begin{cases}
\boldsymbol{v}\\
-\frac{\mu}{r^3}\boldsymbol{r} - \frac{\boldsymbol{\lambda}_v}{\|\boldsymbol{\lambda}_v\|}\frac{T_{\max}}{m}u\\
-\frac{T_{\max}}{I_{\mathrm{sp}}g_0}u\\
\frac{\mu}{r^3}\boldsymbol{\lambda}_v - \frac{3\mu\boldsymbol{r}\cdot\boldsymbol{\lambda}_v}{r^5}\boldsymbol{r}\\
-\boldsymbol{\lambda}_r\\
-\|\boldsymbol{\lambda}_v\|\frac{T_{\max}}{m^2}u
\end{cases} \tag{23}$$

where the normalized thrust magnitude u rests on Eq. (15). Three cases concerning the value of u need to be distinguished: 0, $1/2 - \rho/(2\varepsilon)$, and 1, corresponding to the right-hand side functions denoted by $\boldsymbol{f}_N$, $\boldsymbol{f}_M$, and $\boldsymbol{f}_F$, respectively. As shown, the value of SF decides the thrust magnitude, so it is critical and thus must be carefully treated during integration. Luckily, the first and second derivatives of SF with respect to time both exist. Their forms are derived as

$$\dot{\rho} = \frac{I_{\mathrm{sp}}g_0}{\lambda_0}\frac{\boldsymbol{\lambda}_v \cdot \boldsymbol{\lambda}_r}{m\|\boldsymbol{\lambda}_v\|} \tag{24}$$

$$\ddot{\rho} = \frac{I_{\mathrm{sp}}g_0}{\lambda_0}\left[-\frac{\boldsymbol{\lambda}_r \cdot \boldsymbol{\lambda}_r}{m\|\boldsymbol{\lambda}_v\|} + \frac{\mu}{r^3}\frac{\|\boldsymbol{\lambda}_v\|}{m} - \frac{3\mu}{r^5}\frac{\|\boldsymbol{\lambda}_v \cdot \boldsymbol{r}\|^2}{m\|\boldsymbol{\lambda}_v\|} + \frac{T_{\max}}{I_{\mathrm{sp}}g_0}\frac{u}{m^2}\frac{\boldsymbol{\lambda}_v \cdot \boldsymbol{\lambda}_r}{\|\boldsymbol{\lambda}_v\|} + \frac{(\boldsymbol{\lambda}_v \cdot \boldsymbol{\lambda}_r)^2}{m\|\boldsymbol{\lambda}_v\|^3}\right] \tag{25}$$

If the fixed step size h is small enough, the value of SF at the $(k+1)$th iteration can be approximately examined by the value at the kth iteration through

$$\rho_{k+1} = \rho_k + \dot{\rho}_k h + \frac{1}{2}\ddot{\rho}_k h^2 \tag{26}$$

Suppose that the kth iteration result $\boldsymbol{x}_k$ is already obtained accurately enough, and the values of ρ_k, $\dot{\rho}_k$, and $\ddot{\rho}_k$ are also computed accurately. The value of SF at the $(k+1)$th iteration, denoted by $\tilde{\rho}_{k+1}$ is evaluated approximately by Eq. (26). Dividing the range of SF into three intervals: $(-\infty, -\varepsilon)$, $[-\varepsilon, \varepsilon]$, and (ε, ∞), the iteration from $\boldsymbol{x}_k$ to $\boldsymbol{x}_{k+1}$ is carried out by first examining which intervals the values of ρ_k and $\tilde{\rho}_{k+1}$ respectively fall into. Thus, there are three cases with respect to ρ_k, for each of which there are also three cases with respect to $\tilde{\rho}_{k+1}$.

1) $\rho_k < -\varepsilon$. The thrust begins fully.

a) $\tilde{\rho}_{k+1} \leq -\varepsilon$. The thrust is always full. Thus, the iteration can be accomplished by $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_F, t_k, \boldsymbol{x}_k, h)$.

b) $-\varepsilon < \tilde{\rho}_{k+1} \leq \varepsilon$. The thrust is first full and then moderate. Thus, the iteration should be accomplished through two substeps. Let us assume that the first substep size ish_1, during which the value of SF subject to Eq. (26) varies from ρ_k to $-\varepsilon$. Therefore, by solving the equation $\rho_k + \dot{\rho}_k h_1 + \ddot{\rho}_k h_1^2/2 = -\varepsilon$ the substep size h_1 is obtained as Eq. (27). Note that its conjugate solution is excluded because it is either less than zero or larger than h. So the two substep iterations are accomplished by the first $\boldsymbol{x}_m = \mathrm{RK4}(\boldsymbol{f}_F, t_k, \boldsymbol{x}_k, h_1)$, and then $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_M, t_k + h_1, \boldsymbol{x}_m, h - h_1)$, where $\boldsymbol{x}_m$ denotes the intermediate result.

c) $\tilde{\rho}_{k+1} > \varepsilon$. The thrust is first full, then moderate, and finally null. Three substeps are necessary. Denote the duration among which the SF varies from ρ_k to $-\varepsilon$ and ε, respectively, by h_1 and h_2. The first duration h_1 is the same as Eq. (27), and similarly the second duration h_2 is obtained as Eq. (28), whose conjugate is also excluded. So the three substep iterations are accomplished by the first $\boldsymbol{x}_{m1} = \mathrm{RK4}(\boldsymbol{f}_F, t_k, \boldsymbol{x}_k, h_1)$, then $\boldsymbol{x}_{m2} = \mathrm{RK4}(\boldsymbol{f}_M, t_k + h_1, \boldsymbol{x}_{m1}, h_2 - h_1)$, and finally $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_N, t_k + h_2, \boldsymbol{x}_{m2}, h - h_2)$, where $\boldsymbol{x}_{m1}$ and $\boldsymbol{x}_{m2}$ denote the intermediate results.

2) $-\varepsilon \leq \rho_k \leq \varepsilon$. The thrust begins moderately.

a) $-\varepsilon \leq \tilde{\rho}_{k+1} \leq \varepsilon$. The thrust is always moderate. Thus, the iteration is given by $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_M, t_k, \boldsymbol{x}_k, h)$.

b) $\tilde{\rho}_{k+1} < -\varepsilon$. The thrust is first moderate and then full. Thus, the two substep iterations are accomplished by the first $\boldsymbol{x}_m = \mathrm{RK4}(\boldsymbol{f}_M, t_k, \boldsymbol{x}_k, h_1)$, and then $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_F, t_k + h_1, \boldsymbol{x}_m, h - h_1)$, where h_1 is given by Eq. (29).

c) $\tilde{\rho}_{k+1} > \varepsilon$. The thrust is first moderate and then null. Thus, the two substep iterations are accomplished by the first $\boldsymbol{x}_m = \mathrm{RK4}(\boldsymbol{f}_M, t_k, \boldsymbol{x}_k, h_1)$, and then $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_N, t_k + h_1, \boldsymbol{x}_m, h - h_1)$, where h_1 holds the same form as Eq. (28).

3) $\rho_k > \varepsilon$. The thrust begins with zero.

a) $\tilde{\rho}_{k+1} > \varepsilon$. The thrust is always null. Thus, the iteration is accomplished by $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_N, t_k, \boldsymbol{x}_k, h)$.

b) $-\varepsilon \leq \tilde{\rho}_{k+1} \leq \varepsilon$. The thrust is first null and then moderate. Thus, the two substep iterations are accomplished by the first

$\boldsymbol{x}_m = \mathrm{RK4}(\boldsymbol{f}_N, t_k, \boldsymbol{x}_k, h_1)$, and then $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_M, t_k + h_1, \boldsymbol{x}_m, h - h_1)$, where h_1 holds the form as Eq. (30).

c) $\tilde{\rho}_{k+1} < -\varepsilon$. The thrust is first null, then moderate, and finally full. Thus, the three substep iterations are accomplished by the first $\boldsymbol{x}_{m1} = \mathrm{RK4}(\boldsymbol{f}_N, t_k, \boldsymbol{x}_k, h_1)$, then $\boldsymbol{x}_{m2} = \mathrm{RK4}(\boldsymbol{f}_M, t_k + h_1, \boldsymbol{x}_{m1}, h_2 - h_1)$, and finally $\boldsymbol{x}_{k+1} = \mathrm{RK4}(\boldsymbol{f}_F, t_k + h_2, \boldsymbol{x}_{m2}, h - h_2)$, where h_1 and h_2 hold the same forms as Eqs. (30) and (29), respectively:

$$h_i = \frac{2(\rho_k + \varepsilon)}{-\dot{\rho}_k - \sqrt{\dot{\rho}_k^2 - 2\ddot{\rho}_k(\rho_k + \varepsilon)}} \tag{27}$$

$$h_i = \frac{2(\rho_k - \varepsilon)}{-\dot{\rho}_k - \sqrt{\dot{\rho}_k^2 - 2\ddot{\rho}_k(\rho_k - \varepsilon)}} \tag{28}$$

$$h_i = \frac{2(\rho_k + \varepsilon)}{-\dot{\rho}_k + \sqrt{\dot{\rho}_k^2 - 2\ddot{\rho}_k(\rho_k + \varepsilon)}} \tag{29}$$

$$h_i = \frac{2(\rho_k - \varepsilon)}{-\dot{\rho}_k + \sqrt{\dot{\rho}_k^2 - 2\ddot{\rho}_k(\rho_k - \varepsilon)}} \tag{30}$$

Although RK4 is chosen to be the integrator, other fixed step integrators are also expected to be feasible. The reason for choosing fixed step integrators is that they are easy to carry out. In fact, adaptive step integrators are expected to be more effective, as long as the switching detection technique can be well embedded in them. In principle, the prediction is more accurate if higher-order derivatives of SF are used. Unfortunately, as expressed by Eq. (25), the second-order derivative explicitly contains bang-bang control, thus resulting in the nonexistences of third- and higher-order derivatives. With a fixed step integrator, where each step has the potential to be subdivided into three substeps at most, it is possible that a poor fixed step choice will encompass more than three separate control modes. Since the fixed step is set to be very small, such as 0.0001 year in the example of an Earth–Venus rendezvous, this encompassment is expected to not happen; otherwise, it is really a problem.

D. Global Search Using Particle Swarm Optimization

The PSO which was first presented by Kennedy and Eberhart [23] in mid-1990s is an evolutionary search algorithm that emulates the collective behavior of bird flocks to solve complex nonlinear optimization problems. This algorithm does not require any gradient information of the objective function to be optimized and uses only some basic mathematical operators so that it is conceptually rather simple. In this subsection, this algorithm is used to globally search the initial unknowns of our energy-optimal problem. Suppose an N-dimensional unconstrained function $F(\boldsymbol{X})$ is to be globally minimized, where $\boldsymbol{X}$ is the search-variable vector with N components, and it is also the position of a so-called particle. In our application, each component of the search-variable vector $\boldsymbol{X}$ is normalized to be in the domain [0, 1] before it is operated by the PSO algorithm. For detail on this algorithm see [28].

The relevant parameters are chosen as follows. When the iteration number n increases from 1 to $n_{\max}$, the inertia factor ω is decreased linearly from $\omega_{\max} = 0.9$ to $\omega_{\min} = 0.4$, the "self-confidence" c_1 is decreased linearly from $c_{1\max} = 2.5$ to $c_{1\min} = 0.5$, and the "swarm confidence" c_2 is increased linearly from $c_{2\min} = 0.5$ to $c_{2\max} = 2.5$. The maximal velocity $V_{\max}$ is set to be 0.8. Generally, the maximal iteration number $n_{\max}$ is set to be 1000, and the number of particles is set to be $20 \times N$; i.e., the swarm size is 20.

For the energy-optimal problem, construct the objective function to be minimized as

$$F(\boldsymbol{X}) = \frac{T_{\max}}{I_{\mathrm{sp}} g_0} \int_{t_0}^{t_f} u^2 \, \mathrm{d}t + p_f \|\boldsymbol{\Phi}(\lambda(t_0))\|^2 \tag{31}$$

which consists of two parts: the first one considers the performance index (13) with $\varepsilon = 1$ and ignores the effect of λ_0; the second one considers the residual of the shooting function as a quadratic penalty with the penalty factor p_f needing to be experimented with, noting that the last component $\|\boldsymbol{\lambda}(t_0)\| - 1$ is excluded from the shooting function because $\boldsymbol{\lambda}(t_0)$ is selected previously to satisfy the normalization condition. To transform the unknown normalized multipliers $\boldsymbol{\lambda}(t_0)$ to the search variables each with domain [0, 1], first, define seven angle variables β_i $(i = 1, 2, \ldots, 7)$ as

$$\beta_{1,2,3} = \frac{\pi}{2}\boldsymbol{X}_{1,2,3} \in \left[0, \frac{\pi}{2}\right] \qquad \beta_{4,5} = \pi\left(\boldsymbol{X}_{4,5} - \frac{1}{2}\right)\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$$
$$\beta_{6,7} = 2\pi\boldsymbol{X}_{6,7}[0, 2\pi] \tag{32}$$

where $\boldsymbol{X}_i$ $(i = 1, 2, \ldots, 7)$ are the components of the search-variable vector. Then, relate $\boldsymbol{\lambda}(t_0)$ to these angle variables by

$$\begin{aligned}
&\lambda_0 = \sin\beta_1 \\
&\boldsymbol{\lambda}_{\boldsymbol{r}}(t_0) = \cos\beta_1 \cos\beta_2 \cos\beta_3[\cos\beta_4 \cos\beta_6, \cos\beta_4 \sin\beta_6, \sin\beta_4]^T \\
&\boldsymbol{\lambda}_{\boldsymbol{v}}(t_0) = \cos\beta_1 \cos\beta_2 \sin\beta_3[\cos\beta_5 \cos\beta_7, \cos\beta_5 \sin\beta_7, \sin\beta_5]^T \\
&\lambda_m(t_0) = \cos\beta_1 \sin\beta_2
\end{aligned} \tag{33}$$

which automatically satisfy $\|\boldsymbol{\lambda}(t_0)\| - 1 = 0$, $\lambda_0 \geq 0$, and $\lambda_m(t_0) \geq 0$. Although the ill-conditioned $\lambda_0 = 0$ is not excluded, it seldom happens among computation. To compute the objective function of integral form, besides the 14-D dynamic and costate equations, an additional dimension of the right-hand side $c_2 u^2$ should be integrated. Note that, since the energy-optimal problem is of continuous optimal thrust, Runge–Kutta methods with adaptive step size, expected to be of less computational time than those with fixed step size, are competent for numerical integration. Here the Runge–Kutta–Fehlberg method of orders 7 and 8 [RKF7(8)] [29] is chosen.

E. Nonlinear Equation Solver and Solving Strategy Summary

To solve the shooting function of TPBVP, MinPack-1 [30], a package of FORTRAN subprograms for the numerical solution of systems of nonlinear equations and nonlinear least-squares problems, is used. Its section about solving nonlinear equations is translated into C++ programming language to accord with our other subprograms. In MinPack-1, a modification of Powell's hybrid algorithm [31] that is a combination of Newton's method and the method of the gradient is implemented to solve nonlinear equations. The algorithm used in MinPack-1 is similar in nature to the algorithm implemented by the default solver of MATLAB's *fsolve*. For the problems considered, the Jacobian is simply computed by a forward-difference approximation. Another way that is rather complicated but of improved accuracy is to integrate the variational equations [19].

The solving strategy of the fuel-optimal problem is summarized as follows.

1) Guess the penalty factor, then use the PSO algorithm to search the approximate solutions of the normalized multipliers $[\lambda_0; \boldsymbol{\lambda}_{\boldsymbol{r}}(t_0); \boldsymbol{\lambda}_{\boldsymbol{v}}(t_0); \lambda_m(t_0)]$ to the energy-optimal problem ($\varepsilon = 1$). The integrator is RKF7(8).

2) Use the approximate solutions of step 1 as initial values to solve $[\lambda_0; \boldsymbol{\lambda}_{\boldsymbol{r}}(t_0); \boldsymbol{\lambda}_{\boldsymbol{v}}(t_0); \lambda_m(t_0)]$ of the energy-optimal problem by making use of the MinPack-1 package. The integrator is RKF7(8). If it does not converge, go to step 1.

3) Solve successively $[\boldsymbol{\lambda}_{\boldsymbol{r}}(t_0); \boldsymbol{\lambda}_{\boldsymbol{v}}(t_0); \lambda_m(t_0)]$ corresponding to a series of decreased ε_n, while $\varepsilon_{n+1} = \varepsilon_n - d$, $\varepsilon_0 = 1.0$, $\varepsilon_n \geq 0$, and d is the decreasing step. The multiplier λ_0 is fixed with the value obtained from step 2. The preceding solution is used as the initial guess for the current iteration. The integrator is RK4. If no convergence happens for ε_{n+1}, shorten step d by multiplying a factor less than one, and solve again the solution corresponding to updated ε_{n+1}.

4) Output the convergent shooting unknowns as well as the optimal thrust.

The initial step d is set to be, for example, 0.1, and its shortening factor 0.8. If it does not converge for ε_1, reset d with a smaller value and try step 3 again. The convergence condition is that the Euclidean

norm of the residual of the shooting function is less than the admissible tolerance. In MinPack-1, the input parameter "factor" used to determine the initial step bound in the subroutine "hybrid" is recommended with the value 100; since our unknowns are normalized, 0.01 seems preferable. Now that nondimensionalization has been carried out, the relative tolerance and absolute tolerances of RKF7(8) are all set to be 10^{-10}. The fixed step size of RK4 is conservatively estimated by observing the adaptive step size of RKF7(8) used to propagate the dynamic equation under full thrust.

III. Intermediate Gravity Assist Consideration

The optimal control problem with fixed boundary conditions may be the simplest case of constraint consideration, whereas the cases of complex inner constraints such as intermediate GA are frequently encountered in the mission design of deep space exploration. It will be shown that the techniques presented in the preceding section are available to solve the fuel-optimal solution of these complex cases involving both equality and inequality constraints.

A. Maneuver Impulse Evaluation

The GA impulse model [32] is used. The time spent inside planetary influence is neglected. The spacecraft positions just before and after GA are both required to be equal to the planet position. The spacecraft velocity relative to the planet when the spacecraft is far from the planet is called the hyperbolic excess velocity and is denoted by $\boldsymbol{v}_\infty$. Its magnitude remains invariant just before and after GA. The rotation angle θ that the $\boldsymbol{v}\infty$ is turned during the GA depends on the periapsis radius r_p of the spacecraft relative to the planet and the magnitude of hyperbolic excess velocity v_∞. These relationships are expressed by

$$\boldsymbol{r}(t_m) - \boldsymbol{r}_a(t_m) = 0 \tag{34}$$

$$\boldsymbol{v}_\infty^{\pm} \triangleq \boldsymbol{v}(t_m^{\pm}) - \boldsymbol{v}_a(t_m) \tag{35}$$

$$\|\boldsymbol{v}_\infty^-\| = \|\boldsymbol{v}_\infty^+\| \triangleq v_\infty \tag{36}$$

$$\boldsymbol{v}_\infty^- \cdot \boldsymbol{v}_\infty^+ = v_\infty^2 \cos\theta \tag{37}$$

$$\sin\frac{\theta}{2} = \frac{\mu_a/r_p}{v_\infty^2 + \mu_a/r_p} \tag{38}$$

$$r_p \geq r_{\min} \tag{39}$$

where superscripts $-$ and $+$ denote quantities, respectively, just before and after GA, t_m is the date that GA happens, μ_a denotes the gravitational constant of the planet, subscript a denotes quantities about the planet, and $r_{\min}$ is the minimal admissible periapsis radius. The time is considered to vary continuously so that $t_m^- = t_m^+ = t_m$, whereas the spacecraft velocity suffers a sudden change due to the GA effect.

In preliminary mission design, the patched-conic approximation is usually used, and a series of Lambert problems are solved. The sum of the magnitudes of velocity differences between each leg's arrival velocity, denoted by $\boldsymbol{v}_{\text{in}}$, and those immediately following the leg's departure velocity, denoted by $\boldsymbol{v}_{\text{out}}$, is regarded as the maneuver impulse to be minimized. When the GA effect is taken into account, the value $\|\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}}\|$ is not necessarily offered entirely by the thruster. Thus, the problem coming out is how to analytically evaluate the net maneuver impulse offered by the thruster. Let us denote the GA impulse by $\Delta\boldsymbol{v}_G$ and the net maneuver impulse by $\Delta\boldsymbol{v}_T$. Then one has

$$\boldsymbol{v}_{\text{out}} = \boldsymbol{v}_{\text{in}} + \Delta\boldsymbol{v}_G + \Delta\boldsymbol{v}_T \tag{40}$$

where $\Delta\boldsymbol{v}_G$ is determined to cause $\|\Delta\boldsymbol{v}_T\|$ to be minimal.

Suppose that the GA impulse is contributed immediately after the maneuver impulse (the case that $\Delta\boldsymbol{v}_G$ is contributed just before $\Delta\boldsymbol{v}_T$ can be analyzed analogously). Therefore, the hyperbolic excess velocity after GA holds the form

$$\boldsymbol{v}_\infty^+ = \boldsymbol{v}_{\text{out}} - \boldsymbol{v}_a \tag{41}$$

where $\boldsymbol{v}_a$ is the velocity of the planet at the moment GA occurs. Using the rotation angle θ, assume the $\boldsymbol{v}_\infty^-$ with the form

$$\boldsymbol{v}_\infty^- = v_\infty(\sin\theta\cos\varphi\boldsymbol{k} + \sin\theta\sin\varphi\boldsymbol{j} + \cos\theta\boldsymbol{i}) \tag{42}$$

where $v_\infty = \|\boldsymbol{v}_\infty^+\|$, $\boldsymbol{i} = \boldsymbol{v}_\infty^+/v_\infty$, $\boldsymbol{k}\perp\boldsymbol{i}$, $\boldsymbol{j} = \boldsymbol{k}\times\boldsymbol{i}$, $\theta \in [0, \theta_{\max}]$, and $\varphi \in [0, 2\pi]$. The expression for $\theta_{\max}$ can be derived from Eq. (38) by replacing r_p with $r_{\min}$:

$$\theta_{\max} = 2\sin^{-1}\left[\frac{1}{1 + v_\infty^2 r_{\min}/\mu_a}\right] \tag{43}$$

The unit vector $\boldsymbol{k}$, perpendicular to the unit vector $\boldsymbol{i}$, is not unique in theory. Nevertheless, its one reasonable resolution is given by

$$\boldsymbol{k} = \begin{cases} [0, -i_3, i_2]^T, & |i_1| \leq \min(|i_2|, |i_3|) \\ [i_3, 0, -i_1]^T, & |i_2| \leq \min(|i_1|, |i_3|) \\ [-i_2, i_1, 0]^T, & |i_3| \leq \min(|i_1|, |i_2|) \end{cases} \tag{44}$$

where i_1, i_2, and i_3 denote the X-, Y-, and Z-components of the unit vector $\boldsymbol{i}$ projected onto the HERF, respectively. Then, the GA impulse is expressed as

$$\Delta\boldsymbol{v}_G = \boldsymbol{v}_\infty^+ - \boldsymbol{v}_\infty^- = -v_\infty[\sin\theta\cos\varphi\boldsymbol{k} + \sin\theta\sin\varphi\boldsymbol{j} + (\cos\theta - 1)\boldsymbol{i}] \tag{45}$$

Let us project the velocity difference $\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}}$ onto the frame with the base vectors $\{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$ and divide each component by $-v_\infty$. It yields

$$\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}} = -v_\infty(\upsilon_1\boldsymbol{i} + \upsilon_2\boldsymbol{j} + \upsilon_3\boldsymbol{k}) \tag{46}$$

where

$$\upsilon_1 = -\frac{(\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}})\cdot\boldsymbol{i}}{v_\infty}, \qquad \upsilon_2 = -\frac{(\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}})\cdot\boldsymbol{j}}{v_\infty}$$
$$\upsilon_3 = -\frac{(\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}})\cdot\boldsymbol{k}}{v_\infty} \tag{47}$$

Substitute Eqs. (45) and (46) into Eq. (40). Minimizing $\|\Delta\boldsymbol{v}_T\|$, i.e., $\|\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}} - \Delta\boldsymbol{v}_G\|$, is equivalent to minimizing the function

$$\frac{\|\boldsymbol{v}_{\text{out}} - \boldsymbol{v}_{\text{in}} - \Delta\boldsymbol{v}_G\|^2}{v_\infty^2} = (\sin\theta\cos\varphi - \upsilon_3)^2 + (\sin\theta\sin\varphi - \upsilon_2)^2 + (\cos\theta - 1 - \upsilon_1)^2 \tag{48}$$

where $\theta \in [0, \theta_{\max}]$ and $\varphi \in [0, 2\pi]$ are the unknowns to be determined. It is not difficult to obtain that, to cause Eq. (48) to be minimal, these must hold:

$$\varphi = \arctan 2(\upsilon_2, \upsilon_3) \tag{49}$$

$$\theta = \min(\theta_m, \gamma) \tag{50}$$

where

$$\gamma = \cos^{-1}\frac{1 + \upsilon_1}{\sqrt{(1 + \upsilon_1)^2 + \upsilon_2^2 + \upsilon_3^2}} \tag{51}$$

and the function $\arctan 2(y, x)$ is the four quadrant inverse tangent function.

In conclusion, for each pair of arrival velocity $\boldsymbol{v}_{\text{in}}$ and departure velocities $\boldsymbol{v}_{\text{out}}$, the GA periapsis radius and the corresponding GA impulse that result in minimal maneuver impulse can be analytically evaluated by Eqs. (38) and (45), respectively; the expressions of other relevant quantities are all presented previously.

B. First-Order Necessary Conditions of Gravity Assist

First, let us consider a general controlled system of an n-D dynamic equation:

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{U}, t) \tag{52}$$

where $\boldsymbol{U}$ is the general control vector. The p-D equality constraint that combines both boundary constraints and inner constraints (termed an event) is given by

$$\boldsymbol{\psi}\left(\boldsymbol{x}(t_0), \boldsymbol{x}(t_f), \boldsymbol{x}(t_m^-), \boldsymbol{x}(t_m^+), t_0, t_f, t_m^-, t_m^+\right) = \boldsymbol{0} \tag{53}$$

and the q-D inequality constraint is

$$\boldsymbol{\sigma}\left(\boldsymbol{x}(t_0), \boldsymbol{x}(t_f), \boldsymbol{x}(t_m^-), \boldsymbol{x}(t_m^+), t_0, t_f, t_m^-, t_m^+\right) \le \boldsymbol{0} \tag{54}$$

where the relevant quantities are all of certain degrees of freedom. All vectors and vector functions are arranged in columns. The performance index to be minimized is defined as

$$J = \phi\left(\boldsymbol{x}(t_0), \boldsymbol{x}(t_f), \boldsymbol{x}(t_m^-), \boldsymbol{x}(t_m^+), t_0, t_f, t_m^-, t_m^+\right) + \int_{t_0}^{t_f} L(\boldsymbol{x}, \boldsymbol{U}, t)\,\mathrm{d}t \tag{55}$$

The Lagrange multiplier method is used to transform the constrained optimization problem into an unconstrained one. Furthermore, the performance index is multiplied by a positive factor λ_0. It is augmented as

$$J_m = \lambda_0\phi + \int_{t_0}^{t_m^-}(H - \boldsymbol{\lambda}\cdot\dot{\boldsymbol{x}})\,\mathrm{d}t + \int_{t_m^+}^{t_f}(H - \boldsymbol{\lambda}\cdot\dot{\boldsymbol{x}})\,\mathrm{d}t + \boldsymbol{\chi}\cdot\boldsymbol{\psi} + \boldsymbol{\kappa}\cdot\boldsymbol{\sigma} \tag{56}$$

where $\boldsymbol{\lambda}$ is an n-D functional multiplier and $\boldsymbol{\chi}$ and $\boldsymbol{\kappa}$ are, respectively, p- and q-D numerical multipliers. The Hamiltonian is of the form

$$H(\boldsymbol{x}, \boldsymbol{U}, \boldsymbol{\lambda}, t) = \lambda_0 L(\boldsymbol{x}, \boldsymbol{U}, t) + \boldsymbol{\lambda}\cdot\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{U}, t) \tag{57}$$

The FONCs for a minimum of J_m are derived by setting the first-order variation δJ_m equal to zero, among which the optimal control is decided by $\partial H/\partial \boldsymbol{U} = 0$ for the unconstrained case, and is decided through applying the minimum principle (equivalent to PMP) for the constrained case. Through the derivation (see [27] for details on deriving the first-order variation), the Euler–Lagrange equation $\dot{\boldsymbol{\lambda}} = -\partial H/\partial \boldsymbol{x}$ (for $\delta\boldsymbol{x}$), the transversality conditions, and the stationarity conditions are deduced. Besides, the complementary slackness conditions

$$\boldsymbol{\kappa}_j\boldsymbol{\sigma}_j = 0, \qquad j = 1, 2, \cdots, q \tag{58}$$

and the nonnegativity conditions

$$\lambda_0 > 0, \qquad \boldsymbol{\kappa}_j \ge 0, \qquad j = 1, 2, \cdots, q \tag{59}$$

should be satisfied at the same time. Most importantly, it is reasonable to apply the normalization condition

$$\sqrt{\lambda_0^2 + \boldsymbol{\chi}\cdot\boldsymbol{\chi} + \boldsymbol{\kappa}\cdot\boldsymbol{\kappa} + \ldots} = 1 \tag{60}$$

where $\ldots$ denotes the square of the Euclidean norm of other multipliers such as the initial costate $\boldsymbol{\lambda}(t_0)$. Note that if the performance index is to be maximized the inequality constraint must be written in the form $\boldsymbol{\sigma} \ge 0$ so that the complementary slackness conditions and the nonnegativity conditions, respectively, hold the same forms as Eqs. (58) and (59). The number of unknowns is always equal to the number of equations. For the instances of multiple inner events, the derivation is analogous.

Now let us derive the FONCs of intermediate GA for the fuel-optimal problem. As listed in the preceding subsection, for an intermediate GA, there is the 4-D equality constraint

$$\boldsymbol{\psi} \triangleq \begin{pmatrix} \boldsymbol{r}(t_m) - \boldsymbol{r}_a(t_m) \\ v_\infty^- - v_\infty^+ \end{pmatrix} = \boldsymbol{0} \tag{61}$$

and the 1-D inequality constraint

$$\boldsymbol{\sigma} \triangleq (1 - r_p/r_{\min}) \le \boldsymbol{0} \tag{62}$$

Substitute Eqs. (61) and (62) into the augmented performance index, and derive the first-order variation

$$\begin{aligned}\delta J_m = \cdots &+ [\boldsymbol{\lambda}_r(t_m^+) - \boldsymbol{\lambda}_r(t_m^-) + \boldsymbol{\chi}_{1\sim3}]\cdot\delta\boldsymbol{r}(t_m) + \left[-\boldsymbol{\lambda}_v(t_m^-) + \boldsymbol{\chi}_4\frac{\partial v_\infty^-}{\partial \boldsymbol{v}(t_m^-)} - \kappa\frac{\partial r_p}{r_{\min}\partial \boldsymbol{v}(t_m^-)}\right]\cdot\delta\boldsymbol{v}(t_m^-) + \left[\boldsymbol{\lambda}_v(t_m^+) - \boldsymbol{\chi}_4\frac{\partial v_\infty^+}{\partial \boldsymbol{v}(t_m^+)} - \kappa\frac{\partial r_p}{r_{\min}\partial \boldsymbol{v}(t_m^+)}\right]\cdot\delta\boldsymbol{v}(t_m^+) \\ &+ \left[H(t_m^-) - H(t_m^+) - \boldsymbol{\chi}_{1\sim3}\cdot\boldsymbol{v}_a(t_m) + \boldsymbol{\chi}_4\frac{\partial v_\infty^-}{\partial t_m} - \boldsymbol{\chi}_4\frac{\partial v_\infty^+}{\partial t_m} - \kappa\frac{\partial r_p}{r_{\min}\partial t_m}\right]\delta t_m\end{aligned} \tag{63}$$

where $\boldsymbol{x}$ denotes $[\boldsymbol{r}; \boldsymbol{v}]$, and the relations $t_m^- = t_m^+ = t_m$ and $\boldsymbol{r}(t_m^-) = \boldsymbol{r}(t_m^+) = \boldsymbol{r}(t_m)$ are taken into account. The relations $\phi = 0$ and $\partial r_p/\partial \boldsymbol{r}(t_m) = 0$ are used, and only the terms associated with the intermediate state and date are listed. Keep in mind that the intermediate date must be later than the initial time and earlier than the final time.

To derive out the derivatives of v_∞^-, v_∞^+, and r_p, with respect to the basic variables $\boldsymbol{v}(t_m^-)$, $\boldsymbol{v}(t_m^+)$, and t_m, first let us define two unit vectors:

$$\hat{\boldsymbol{\imath}}^{\pm} \triangleq \frac{\boldsymbol{v}_\infty^{\pm}}{v_\infty^{\pm}} \tag{64}$$

From Eq. (35), the variations of $\boldsymbol{v}_\infty^-$ and $\boldsymbol{v}_\infty^+$ with respect to the basic variables are derived as

$$\delta\boldsymbol{v}_\infty^{\pm} = \delta\boldsymbol{v}(t_m^{\pm}) - \boldsymbol{a}_a(t_m)\delta t_m \tag{65}$$

where $\boldsymbol{a}_a(t_m)$ denotes the acceleration vector of the planet at the moment of GA. Subject to the center gravitation of the sun, the acceleration of the planet is expressed by

$$\boldsymbol{a}_a(t_m) = -\frac{\mu}{r_a^3(t_m)}\boldsymbol{r}_a(t_m) \tag{66}$$

Taking into account the definitions of the unit vectors $\hat{\boldsymbol{\imath}}^{\pm}$, the variation of the magnitude of the hyperbolic access velocity is derived as

$$\delta v_\infty^{\pm} = \hat{\boldsymbol{\imath}}^{\pm}\cdot(\delta\boldsymbol{v}(t_m^{\pm}) - \boldsymbol{a}_a(t_m)\delta t_m) \tag{67}$$

Then the derivatives of v_∞^- and v_∞^+ with respect to $\boldsymbol{v}(t_m^-)$, $\boldsymbol{v}(t_m^+)$, and t_m come out

$$\frac{\partial v_\infty^{\pm}}{\partial \boldsymbol{v}(t_m^{\pm})} = \hat{\boldsymbol{\imath}}^{\pm}, \qquad \frac{\partial v_\infty^{\pm}}{\partial t_m} = -\hat{\boldsymbol{\imath}}^{\pm}\cdot\boldsymbol{a}_a(t_m) \tag{68}$$

The derivatives of r_p with respect to the basic variables seem rather complicated. Nevertheless, first the variation of $\hat{\boldsymbol{\imath}}^{\pm}$ is derived:

$$\delta\hat{\boldsymbol{\imath}}^{\pm} = \frac{1}{v_\infty^{\pm}}(\boldsymbol{I} - \hat{\boldsymbol{\imath}}^{\pm}\hat{\boldsymbol{\imath}}^{\pm})\cdot(\delta\boldsymbol{v}(t_m^{\pm}) - \boldsymbol{a}_a(t_m)\delta t_m) \tag{69}$$

where $\boldsymbol{I}$ is a unit 3×3 matrix, and $\hat{\boldsymbol{\imath}}^{\pm}\hat{\boldsymbol{\imath}}^{\pm}$ is a dyadic product. Rewrite Eqs. (37) and (38) as the forms

$$\cos\theta = \hat{\boldsymbol{\imath}}^{-}\cdot\hat{\boldsymbol{\imath}}^{+} \tag{70}$$

$$r_p = \frac{\mu_a}{v_\infty^- v_\infty^+}\left(\frac{1}{\sin\theta/2} - 1\right) \tag{71}$$

Then one has

$$\delta\theta = -\frac{1}{\sin\theta}(\hat{\boldsymbol{\imath}}^- \cdot \delta\hat{\boldsymbol{\imath}}^+ + \hat{\boldsymbol{\imath}}^+ \cdot \delta\hat{\boldsymbol{\imath}}^-) \tag{72}$$

After some complicated derivation and assembly, finally we have

$$\delta r_p = \boldsymbol{A} \cdot \delta\boldsymbol{v}(t_m^-) + \boldsymbol{B} \cdot \delta\boldsymbol{v}(t_m^+) + C\delta t_m \tag{73}$$

where

$$\begin{cases} \boldsymbol{A} = \frac{r_p}{v_\infty^-}\left[\frac{1}{4\sin^2\theta/2(1-\sin\theta/2)}(\hat{\boldsymbol{\imath}}^+ - \cos\theta\boldsymbol{u}^-) - \hat{\boldsymbol{\imath}}^-\right] \\ \boldsymbol{B} = \frac{r_p}{v_\infty^+}\left[\frac{1}{4\sin^2\theta/2(1-\sin\theta/2)}(\hat{\boldsymbol{\imath}}^- - \cos\theta\boldsymbol{u}^+) - \hat{\boldsymbol{\imath}}^+\right] \\ C = r_p\left\{-\frac{1}{4\sin^2\theta/2(1-\sin\theta/2)}\left[\frac{1}{v_\infty^+}(\hat{\boldsymbol{\imath}}^- - \cos\theta\hat{\boldsymbol{\imath}}^+) + \frac{1}{v_\infty^-}(\hat{\boldsymbol{\imath}}^+ - \cos\theta\boldsymbol{u}^-)\right] + \frac{\hat{\boldsymbol{\imath}}^+}{v_\infty^+} + \frac{\hat{\boldsymbol{\imath}}^-}{v_\infty^-}\right\} \cdot \boldsymbol{a}_a(t_m) \end{cases} \tag{74}$$

Then the derivatives of r_p with respect to $\boldsymbol{v}(t_m^-)$, $\boldsymbol{v}(t_m^+)$, and t_m come out

$$\frac{\partial r_p}{\partial \boldsymbol{v}(t_m^-)} = \boldsymbol{A}, \qquad \frac{\partial r_p}{\partial \boldsymbol{v}(t_m^+)} = \boldsymbol{B}, \qquad \frac{\partial r_p}{\partial t_m} = C \tag{75}$$

Substitute Eqs. (68) and (75) into Eq. (63). To cause δJ_m to identically vanish, the relation of the position costate just before and after GA is deduced:

$$\boldsymbol{\lambda}_r(t_m^+) = \boldsymbol{\lambda}_r(t_m^-) - \boldsymbol{\chi}_{1\sim3} \tag{76}$$

the velocity costate just after GA is decided by

$$\boldsymbol{\lambda}_v(t_m^+) = \chi_4\hat{\boldsymbol{\imath}}^+ + \frac{1}{r_{\min}}\kappa\boldsymbol{B} \tag{77}$$

the velocity costate just before GA should satisfy

$$\boldsymbol{\lambda}_v(t_m^-) - \chi_4\hat{\boldsymbol{\imath}}^- + \frac{1}{r_{\min}}\kappa\boldsymbol{A} = 0 \tag{78}$$

and the stationarity condition related to the GA date is

$$H(t_m^-) - H(t_m^+) - \boldsymbol{\chi}_{1\sim3} \cdot \boldsymbol{v}_a(t_m) + \chi_4(\hat{\boldsymbol{\imath}}^+ - \hat{\boldsymbol{\imath}}^-) \cdot \boldsymbol{a}_a(t_m) - \frac{1}{r_{\min}}\kappa C = 0 \tag{79}$$

For an intermediate GA, there are nine unknowns, consisting of the 4-D numerical multiplier $\boldsymbol{\chi}$, the 1-D nonnegative numerical multiplier κ, the 3-D velocity vector $\boldsymbol{v}(t_m^+)$ just after GA, and the GA date t_m. Meanwhile, there are also nine equations, consisting of the 4-D Eq. (61), the 1-D Eq. (58) with the application of inequality (62), the 3-D Eq. (78), and the stationarity condition (79). Regard the velocity increment $\Delta\boldsymbol{v}_G$ caused by GA rather than the velocity just after GA, as the unknowns are more convenient to solve for the shooting function. Note that Eqs. (76) and (77) are not used as parts of the shooting function but used to update the position and velocity of the costates just after GA. The variations of the mass as well as its costate are not considered because they are continuous just before and after GA. For deep space exploration, there is another type of mission called flyby which only requires a spacecraft position equal to that of the objects such as an asteroid, planet, comet, etc. Consequently, as a special case, the FONCs of flyby are simpler than that of GA: the last component of equality constraint (61) and inequality constraint (62) both vanish, the velocity vector and its costate just before and after flyby are both continuous so that Eqs. (77) and (78) vanish, and the terms of $\boldsymbol{\chi}_4$ and κ in the stationarity condition (79) also vanish.

IV. Examples and Discussion

Two examples of fuel-optimal problems of fixed boundary conditions without and with intermediate GA are given to substantiate the techniques and theories presented in Secs. II and III, respectively. All computations are executed on a desktop personal computer with a CPU of 3.0 GHz and memory 2.0 GB, and Microsoft Visual C++ 6.0 is used. The heliocentric position, velocity, and orbit elements of planets are computed online by the Jet Propulsion Laboratory Horizons system.[§]

A. Rendezvous from the Earth to Venus

The rendezvous problem from the Earth to Venus is considered, in which the spacecraft starts with the Earth heliocentric position and velocity and arrives at Venus with the same heliocentric position and velocity of Venus. This example is exactly the same as that in [9]. The parameters are listed in Table 1. It should be noted that among computation the quantities of length, time, and mass are always nondimensionalized by astronomical unit, year, and spacecraft initial mass, respectively.

The techniques and solving strategy presented in Sec. II are applied to solve this fuel-optimal rendezvous problem. The default settings for the PSO algorithm are the maximal iteration number $n_{\max} = 100$ and the swarm size $S = 10$. The fixed step size of RK4 is 10^{-4} year. Now that the suitable value of the penalty factor p_f is not known in advance, as a test, it is generated randomly with the form 10^n, where n is a uniform random distribution with the domain $[-5, 5]$ in default. Furthermore, to compare with the effect of the PSO search, random guesses of the initial value under the normalization condition are considered. And then the effects of random guesses whether with initial costate normalization are compared with each other. Twelve cases are computed, each with 1000 starting points. As listed in Table 2, $\varepsilon = 0$ corresponds to directly solving the fuel-optimal problem, $\varepsilon = 1$ corresponds to solving the energy-optimal problem, and $\varepsilon = 0.1$ is a compromise between the former two. When calling MinPack's subroutine hybird, the parameter maxfev that denotes the maximal number of calls to evaluate the shooting function is set to be 1000. For the purpose of comparison, the results of [9] of considering quadratic penalty to the performance index are listed, noting that the number of starting points are 100. Through lots of simulations, four local solutions s, s_1, s_2, and s_3 are found, as listed in Table 3, where the former three solutions are identical to the results of [9], but the last one is a new discovery. Since no better solution is obtained, solution s is regarded as the globally optimal one, as in [9]. The distribution of the convergent solutions of case 6 with respect to the penalty factor is plotted in Fig. 1. It shows that the penalty factor with the domain $[10^{-2}, 10^5]$ almost always leads to the globally optimal solution s. So case 10 with the penalty factor in the domain $[10^{-2}, 10^5]$ is intentionally simulated, and as

[§]Data available online at http://ssd.jpl.nasa.gov/?horizons [retrieved 5 Sept. 2010]

expected it shows that the probability of finding the global solution is up to 0.97.

When solving the energy-optimal problem by random guesses without costate normalization, the bound of guess space should be prescribed so that the sample starting points are taken randomly within this space. The difficulty is that the bound is unknown in advance. Prescribing the bound to be too much larger than the real one seriously decreases the probability of finding convergent solutions. The knowledge of the normalized results as listed in Table 2 dividing each solution by the corresponding value of λ_0 results in the unnormalized solution (corresponds to $\lambda_0 = 1$). The bound for the unnormalized position and velocity costate vectors is about $[-0.67, 1.35]$ and for the mass costate is about $[0, 0.6]$. Thus, case 11 is computed by making use of the bound conclusion arising from the normalized results. And case 12 corresponds to intentionally prescribing the bound with $[-100.0, 100.0]$ (the input parameter factor of the subroutine hybrid is then set to be 1.0), which is expected to be large enough. But without the knowledge of a normalized solution, this cannot be ensured at all.

Table 2 shows that when we directly solve the fuel-optimal problem with the normalized initial values through either random guesses or a PSO search, no convergent solution is obtained within 1000 samples, to say nothing of the 100 samples in [9]. An additional attempt of 10,000 samples by random guesses gives three s, one s_1, two s_2, and zero s_3, while about 12 h are spent. If starting from the energy-optimal problem, even by random guesses, the probability of finding the global solution is 0.056. This fairly shows the advantage of the homotopic approach. By guessing the normalized initial values for the problem with $\varepsilon = 0.1$, our probability of finding the global solution is 0.03, bigger than [9]'s result 0.01. This is because with the help of normalization it is possible to guess them within a bounded space. As shown by case 11, despite benefiting from using the normalized solution to determine the bound of the guess space, the probabilities of finding the global solution and local solutions without normalization are all a little smaller than those with normalization. The larger the prescribed bound is, the smaller the probabilities are, as shown by case 12. For the PSO search, it seems that the settings of $n_{\max} = 100$ and $S = 10$ are adequate. The average time of obtaining the global energy-optimal solution by the PSO search is 3898 s/970 $\approx$ 4.0 s, and by random guesses is 338 s/56 $\approx$ 6.0 s, indicating that the former is more effective than the latter. The total time of starting from the PSO search to obtaining the fuel-

Table 1 Parameters for the rendezvous example of the Earth to Venus

Parameter	Value	Units
Initial date	7 Oct. 2005 0:0:0.0	Coordinate time
Flight time	1000.0	Day
Initial position	$[9.708322 \times 10^{-1}, 2.375844 \times 10^{-1}, -1.671055 \times 10^{-6}]$	AU
Initial velocity	$[-1.598191, 6.081958, 9.443368 \times 10^{-5}]$	AU/yr
Final position	$[-3.277178 \times 10^{-1}, 6.389172 \times 10^{-1}, 2.765929 \times 10^{-2}]$	AU
Final velocity	$[-6.598211, -3.412933, 3.340902 \times 10^{-1}]$	AU/yr
I_{sp}	3800.0	s
$T_{\max}$	0.33	N
m_0	1500.0	kg

Table 2 Number of convergence cases for 1000 samples

	Solution					
Method	s	s_1	s_2	s_3	Unconverged solution	Time spent, s
Random guesses with normalization						
$\varepsilon = 0$	0	0	0	0	1000	4657
$\varepsilon = 0.1$	30	3	4	1	962	4524
$\varepsilon = 1$	56	48	105	25	766	338
PSO search with normalization						
$\varepsilon = 0$	0	0	0	0	1000	6639
$\varepsilon = 0.1$	153	5	20	0	822	8084
$\varepsilon = 1$	808	3	184	0	5	3700
$\varepsilon = 1$ and $n_{\max} = 50$	776	10	201	2	11	1927
$\varepsilon = 1$ and $n_{\max} = 500$	802	0	183	1	14	16,640
$\varepsilon = 1$ and $S = 20$	806	3	177	0	14	7275
$\varepsilon = 1$ and $p_f \in [10^{-2}, 10^5]$	970	5	20	3	2	3898
Random guesses without normalization						
$\varepsilon = 1$ and the bound $[-0.67, 1.35]$	50	37	54	22	837	341
$\varepsilon = 1$ and the bound $[-100.0, 100.0]$	2	0	1	0	997	230
Results of [9] (100 samples)						
$\varepsilon = 0$	0	0	0	——	100	——
$\varepsilon = 0.1$	1	0	1	——	98	——

Table 3 Four convergent solutions

Solution	λ_0	$[\boldsymbol{\lambda}_r(t_0)\boldsymbol{\lambda}_v(t_0)\lambda_m(t_0)]$	Final mass, kg
s	0.9728	[0.58771, 0.15075, 0.24139, −0.023469, 0.093287, −0.019684, 0.13757]	1290.578
s_1	0.6435	[0.60799, 0.21109, 0.28576, −0.013971, 0.10008, −0.023063, 0.13182]	1259.696
s_2	0.6547	[−0.43839, −0.15518, 0.083403, 0.0073929, −0.071066, −0.0047074, 0.23968]	1036.332
s_3	0.4573	[0.61601, 0.22008, 0.1067, −0.015108, 0.10662, −0.0086826, 0.22927]	1006.557

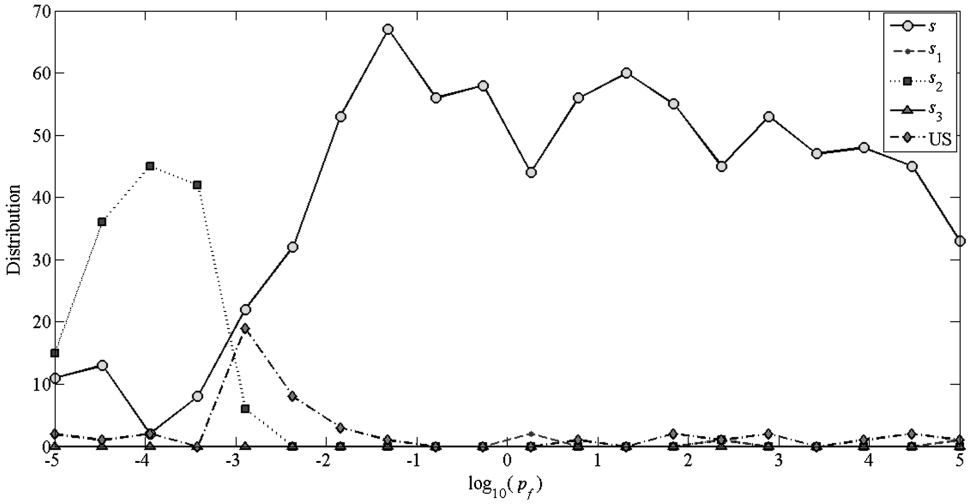

Fig. 1 Distribution of the number of convergence cases for the energy-optimal problem.

optimal solution is about 28 s, and the decreasing route of ε is [1.0, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3, 0.2, 0.1, 0.02, 0.0], just needing a few steps. The convergence process of optimal control is depicted in Fig. 2. It is different from [9] of Fig. 4 because the logarithmic barrier rather than the quadratic penalty is used there to modify the performance index. The bang-bang controls of four convergent solutions are sketched in Fig. 3, noting that to distinguish well the thrusts of three local solutions are scaled differently. It shows that the control of the global solution holds the regularity that the maximal thrust burns only around apogees and perigees, as shown in Fig. 4, but the controls of local solutions seem irregular.

B. Rendezvous from the Earth to Jupiter via Mars Gravity Assist

The rendezvous problem from the Earth to Jupiter via Mars GA is considered, of which the spacecraft starts with the Earth heliocentric position and velocity, and arrives at Jupiter with the same heliocentric position and velocity of Jupiter. This example is exactly the same as example A of [33]. The parameters are listed in Table 4. The motion of Mars is assumed to be Keplerian around the given epoch. Additional examinations tell that the position error of Mars between the Keplerian model and the Jet Propulsion Laboratory Horizons system within ten days around the given epoch is less than 100 km, and the velocity error is less than 1 m/s.

Since the initial and final states and flight time are fixed, two Lambert problems from the Earth to Mars and from Mars to Jupiter are solved to evaluate the maneuver impulse. The Mars GA date is the only variable, and the PSO search is used to find the minimal maneuver impulse. The GA impulse is analytically evaluated. It is easy to get that the GA data are 29 March 2024 (2.3658 years relative to the launch date), the GA hyperbolic excess velocity is 6.76 km/s, the flyby altitude is exactly on the lower bound 500 km, the GA impulse is 2.62 km/s, and the minimal maneuver impulse is 15.7 km/s, resulting in a final mass 15,317 kg.

For the related energy-optimal problem, there are 17 unknowns: $z \triangleq [\lambda_0; \boldsymbol{\lambda}_r(t_0); \boldsymbol{\lambda}_v(t_0); \lambda_m(t_0); \boldsymbol{\chi}; \kappa; \Delta\boldsymbol{v}_G; t_m]$, and the shooting function $\boldsymbol{\Phi}$ consists of the 7-D terminal condition (11), the 4-D equality constraint (61), the 1-D complementary slackness condition (58) with the application of inequality (62), the 3-D transversality condition (78), the 1-D stationary condition (79), and moreover, the available normalization condition (60)

$$\sqrt{\lambda_0^2 + \boldsymbol{\lambda}_r(t_0)\cdot\boldsymbol{\lambda}_r(t_0) + \boldsymbol{\lambda}_v(t_0)\cdot\boldsymbol{\lambda}_v(t_0) + \lambda_m^2(t_0) + \boldsymbol{\chi}\cdot\boldsymbol{\chi} + \kappa^2} = 1$$

Hence, the first 13 components of z are on a unit hypersphere. They can be transformed to the search variables, each with domain [0, 1] applied in the PSO algorithm, by introducing 12 angel variables. The implementation is similar to Eqs. (32) and (33) and thus is passed over. According to [32]'s conclusion, the maximal GA

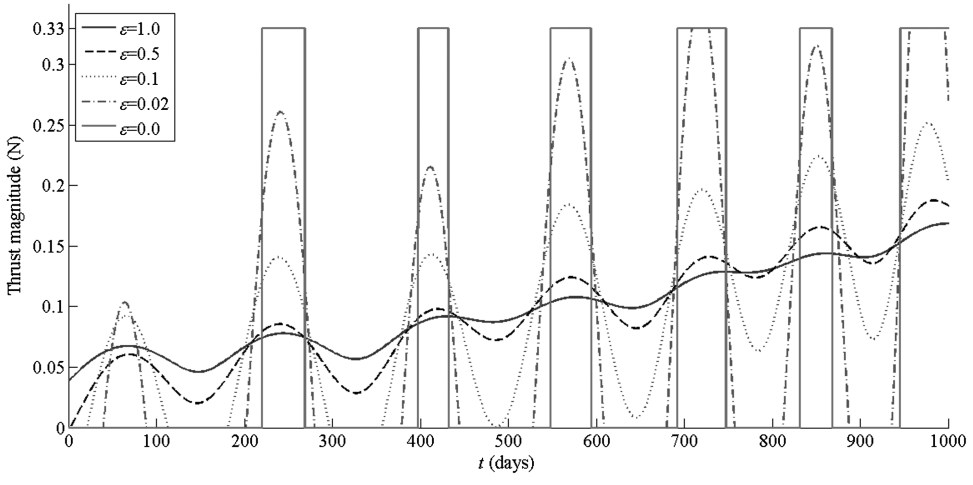

Fig. 2 Optimal thrust profile for decreased ε.

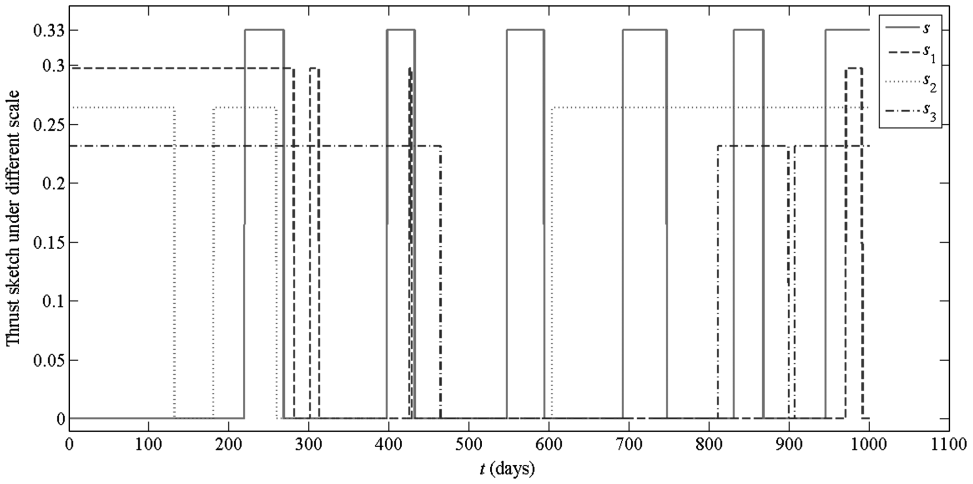

Fig. 3 Thrust sketches of four local solutions under different scales.

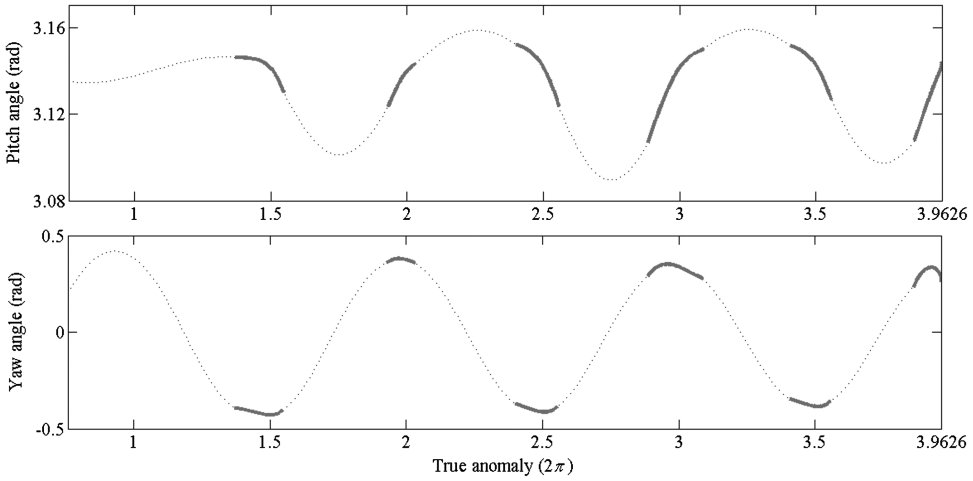

Fig. 4 Pitch angle and yaw angle of optimal control with respect to true anomaly.

impulse is the local circular speed at the minimal admissible periapsis radius. Therefore, transform the $\Delta \boldsymbol{v}_G$ to the search variables by

$$\Delta \boldsymbol{v}_G = \sqrt{\frac{\mu_a}{r_{\min}}} X_{13}[\cos\varphi\cos\delta, \cos\varphi\sin\delta, \sin\varphi]^T$$

where $\varphi = \pi(\boldsymbol{X}_{14} - 0.5) \in [-0.5\pi, 0.5\pi]$ and $\delta = 2\pi \boldsymbol{X}_{15} \in [0, 2\pi]$. The GA date t_m, relative to the launch date in years, can be searched around the impulse result obtained previously with a certain search radius, for example, 0.5 year. When applying the PSO search, set the maximal iteration number $n_{\max} = 500$, the swarm size $S = 10$, and the penalty factor $p_f = 0.1$. The fixed step size of RK4 is 5×10^{-4} yr.

Similar to example A, the PSO search is compared with random guesses to solve the energy-optimal problem. The parameter "maxfev" of MinPack's subroutine "hybrid" is set to be 2000. It

Table 4 Parameters for the example of Earth-to-Jupiter rendezvous by Mars GA

Parameter	Value	Units
Initial time	16 Nov. 2021 0:0:0.0	Coordinate time
Flight time	2201.0	Day
Initial position	$[5.876420 \times 10^{-1}, 7.954627 \times 10^{-1}, -3.845203 \times 10^{-5}]$	AU
Initial velocity	$[-5.155764, 3.707833, -3.191945 \times 10^{-4}]$	AU/yr
Final position	$[-5.204974, 1.495369, 1.102444 \times 10^{-1}]$	AU
Final velocity	$[-7.936872, -2.523063, 2.823220 \times 10^{-2}]$	AU/yr
I_{sp}	6000.0	s
$T_{\max}$	2.26	N
m_0	20,000.0	kg
Mars Gm (μ_a)	42,828.3	km^3/s^2
Mars $r_{\min}$ (Mars radius 3389.9)	3889.9	km
Mars orbital elements [a, e, i, Ω, ω, f] in the J2000 HERF and the epoch	[1.52363312 AU, 0.09327933, 1.84785414 deg, 49.48935357 deg, 286.67090811 deg, 328.88755274 deg], 20 March 2024 0:0:0.0	AU, deg, and coordinate time

Table 5 Earth–Mars–Jupiter rendezvous trajectory

Characteristics	GA date	GA v_∞, km/s	GA altitude, km	Final mass, kg
GALLOP	20 March 2024	3.61	500	16,026
Our results	18 March 2024	3.58	500	16,022

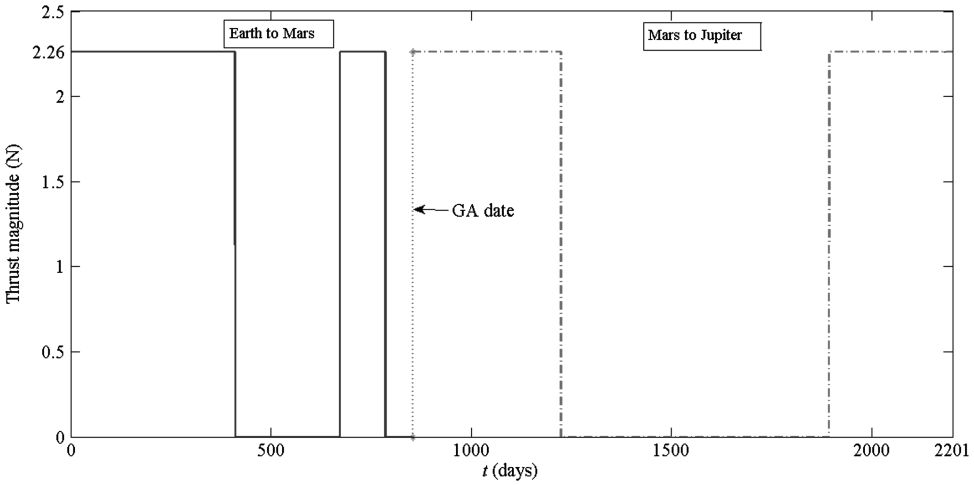

Fig. 5 Optimal thrust profile.

shows that 23 of 100 samples converge to the unique reasonable solution when searched by PSO within 3.1 h, while 6 of 1000 samples converge to the unique reasonable solution when searched by random guesses within 3.6 h. This indicates that the PSO search is more effective than random guesses. Once the solutions to the energy-optimal problem is obtained, the homotopic approach is used to solve the fuel-optimal problem. It takes about 3 min to get the optimal solution $z = [0.69526, -0.18265, -0.25352, -0.027369, 0.038339, -0.028608, -0.01153, 0.15188, 0.021723, -0.050418, 0.065285, -0.021939, 0.017445, 0.40276, 0.56979, -0.011508, 2.3373]$. The GA impulse is 3.31 km/s. Besides the same launch date, launch v_∞, arrival date, arrival v_∞, and flight time, the main different characteristics between our results and [33]'s results obtained by the optimization program called the gravity-assist, low-thrust, local optimization program (GALLOP) are listed in Table 5. GALLOP results were obtained through a parameter optimization technique, a typical direct method, in which low-thrust trajectories were modeled as a series of impulses patched together by conic arcs. The results differ from GALLOP results very slightly, such as an admissible error of 4 kg final mass (1% of the fuel consumption). The results should be globally optimal because they are based on the maximal principle and no other local solution is found. The coincidence between our results and GALLOP results approves mutually. The optimal bang-bang control is depicted in Fig. 5, which has four burn arcs. The trajectory projected onto the ecliptic plan is shown in Fig. 6, where dot lines denote coast arcs, solid lines denote burn arcs, and dash-dot lines denote Earth, Mars, and Jupiter orbits. The obtained trajectory seems also the same as that obtained by [33].

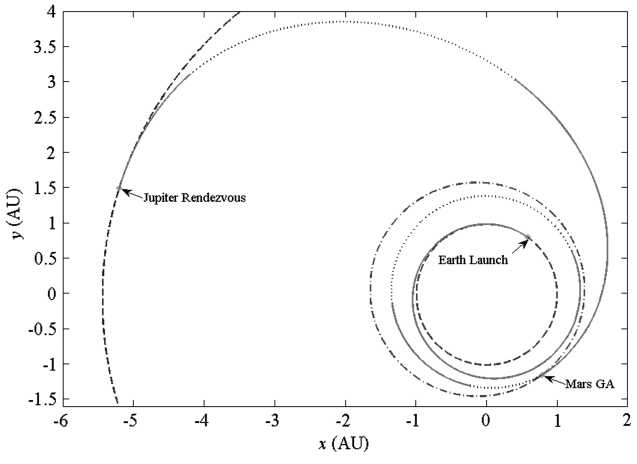

Fig. 6 Earth–Mars–Jupiter low-thrust trajectory.

C. Discussion

The scope of this paper bears stating. Both the derivation and examples only consider fixed boundary states and times. For the intermediate GA, the intermediate date is free, but only one GA is simulated. In principle, the presented techniques, especially the normalization method and switching detection integration algorithm, are expected to be extended to solve the cases of constrained or free boundary states and times with or without multiple GAs. It is substantiated by additional examples beyond this paper. In that case, the new stationary conditions and transversality conditions should be derived. For the problem of long mission duration and many inner constraints, the numeric difficulty increases seriously. The efficiency of searching the initial values by PSO is not remarkably greater than that of random guesses with multiple starting. It takes much time to obtain a convergent solution. Perhaps embedding the switching detection technique in adaptive step integrators rather than in fixed step integrators is worth attempting to reduce the computational time.

V. Conclusions

A suite of practical techniques are presented to solve the fuel-optimal problem of low-thrust trajectory by the homotopic approach combined with the single shooting method. For a general optimal control problem with various equality and inequality constraints, the numerical Lagrange multipliers can always be normalized to be on a unit hypersphere. Consequently, the probability of finding convergent solutions increases several tens of times. Searching the initial unknowns by PSO, the computational time that is taken to find the optimal solution is considerably less than that of random guesses. Embedded in the Runge–Kutta algorithm with a fixed step size, the switching detection technique of using the switching function's first- and second-order derivatives makes the propagation of good

accuracy for bang-bang control. Unlike direct methods, this kind of indirect method involves only a few unknowns. Importantly, when the PSO is used to search the approximate initial values, it seems easy to find the global solution of the optimal control problem with relatively simple constraints. This is substantiated by an example of a rendezvous from the Earth to Venus. Besides the simple problem of fixed boundary conditions, a complex problem with intermediate GA is also considered. The purely analytical first-order necessary conditions for GA are derived. It shows that each GA brings only nine new unknowns, consisting of four numerical multipliers related to the equality constraints, one nonnegative numerical multiplier related to the inequality constraint, three components of GA impulse, and the GA date. Normalization to the initial multipliers is still available. The practical techniques are still competent to solve the complex problem, though they are less effective for solving a simple problem. The example of a rendezvous from the Earth to Jupiter via Mars GA shows that the PSO search is still more effective than random guesses to get the initial values. The final mass and optimal trajectory are the same as those obtained by the optimization program GALLOP, which is a typical kind of direct method. A small number of unknowns are involved in our method: for instance, 17 unknowns in the example of intermediate Mars GA.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (No. 10832004) and the China Postdoctoral Science Foundation (No. 20100470301).

References

[1] Rayman, M. D., Varghese, P., Lehman, D. H., and Livesay, L. L., "Results from the Deep Space 1 Technology Validation Mission," *Acta Astronautica*, Vol. 47, Nos. 2–9, 2000, pp. 475–487. doi:10.1016/S0094-5765(00)00087-4

[2] Kuninaka, H., Nishivama, K., Funakai, I., Tetsuya, Shimizu, T., and Kawaguchi, J., "Asteroid Rendezvous of HAYABUSA Explorer Using Microwave Discharge Ion Engines," 29th International Electric Propulsion Conference, IEPC Paper 2005-10, Princeton Univ., Princeton, NJ, 2005.

[3] Hargraves, C., and Paris, S., "Direct Trajectory Optimization Using Nonlinear Programming and Collocation," *Journal of Guidance, Control, and Dynamics*, Vol. 10, No. 4, 1987, pp. 338–342. doi:10.2514/3.20223

[4] 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. doi:10.2514/3.20934

[5] Kechichian, J. A., "Optimal Low-Earth-Orbit-Geostationary-Earth-Orbit Intermediate Acceleration Orbit Transfer," *Journal of Guidance, Control, and Dynamics*, Vol. 20, No. 4, 1997, pp. 803–811. doi:10.2514/2.4116

[6] Ranieri, C. L., and Ocampo, C. A., "Indirect Optimization of Three-Dimensional Finite-Burning Interplanetary Transfers Including Spiral Dynamics," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 2, 2009, pp. 444–454.

[7] Kluever, C. A., and Pierson, B. L., "Optimal Low-Thrust Three-Dimensional Earth-Moon Trajectories," *Journal of Guidance, Control, and Dynamics*, Vol. 18, No. 4, 1995, pp. 830–837. doi:10.2514/3.21466

[8] Gao, Y., and Kluever, C. A., "Low-Thrust Interplanetary Orbit Transfers Using Hybrid Trajectory Optimization Method with Multiple Shooting," AIAA/AAS Astrodynamics Specialist Conference and Exhibit, AIAA Paper 2004-5088, Providence, RI, 16–19 Aug. 2004.

[9] Bertrand, R., and Epenoy, R., "New Smoothing Techniques for Solving Bang-Bang Optimal Control Problems: Numerical Results and Statistical Interpretation," *Optimal Control Applications and Methods*, Vol. 23, No. 4, 2002, pp. 171–197. doi:10.1002/oca.709

[10] Casalino, L., Colasurdo, G., and Pastrone, D., "Optimal Low-Thrust Escape Trajectories Using Gravity Assist," *Journal of Guidance, Control, and Dynamics*, Vol. 22, No. 5, 1999, pp. 637–642. doi:10.2514/2.4451

[11] Casalino, L., and Sentinella, M. R., "Genetic Algorithm and Indirect Method Coupling for Low-Thrust Trajectory Optimization," 42nd AIAA/ASME/SAE/ASEE Joint Propulsion Conference and Exhibit, AIAA Paper 2006-4468, Sacramento, CA, 9–12 July 2006.

[12] Mantia, M. L., and Casalino, L., "Indirect Optimization of Low-Thrust Capture Trajectories," *Journal of Guidance, Control, and Dynamics*, Vol. 29, No. 4, 2006, pp. 1011–1014. doi:10.2514/1.18986

[13] Nah, R. S., Vadali, S. R., and Braden, E., "Fuel-Optimal, Low-Thrust, Three-Dimensional Earth–Mars Trajectories," *Journal of Guidance, Control, and Dynamics*, Vol. 24, No. 6, 2001, pp. 1100–1107. doi:10.2514/2.4844

[14] Vadali, S. R., Nah, R., Braden, E., and Johnson, I. L., Jr., "Fuel-Optimal Planar Earth–Mars Trajectories Using Low-Thrust Exhaust-Modulated Propulsion," *Journal of Guidance, Control, and Dynamics*, Vol. 23, No. 3, 2000, pp. 476–482. doi:10.2514/2.4553

[15] Mengali, G., and Quarta, A. A., "Trajectory Design with Hybrid Low-Thrust Propulsion System," *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 2, 2007, pp. 419–426. doi:10.2514/1.22433

[16] Lawden, D. F., *Optimal Trajectories for Space Navigation*, Butterworths, London, 1963, pp. 5, 59.

[17] Russell, R. P., "Primer Vector Theory Applied to Global Low-Thrust Trade Studies," *Journal of Guidance, Control, and Dynamics*, Vol. 30, No. 2, 2007, pp. 460–472. doi:10.2514/1.22984

[18] Haberkorn, T., Martinon, P., and Gergaud, J., "Low-Thrust Minimum-Fuel Orbital Transfer: A Homotopic Approach," *Journal of Guidance, Control, and Dynamics*, Vol. 27, No. 6, 2004, pp. 1046–1060. doi:10.2514/1.4022

[19] Martinon, P., and Gergaud, J., "Using Switching Detection and Variational Equations for the Shooting Method," *Optimal Control Applications and Methods*, Vol. 28, No. 2, 2007, pp. 95–116. doi:10.1002/oca.794

[20] Thevenet, J., and Epenoy, R., "Minimum-Fuel Deployment for Spacecraft Formations via Optimal Control," *Journal of Guidance, Control, and Dynamics*, Vol. 31, No. 1, 2008, pp. 101–113. doi:10.2514/1.30364

[21] Bonnans, F., Martinon, P., and Trélat, E., "Singular Arcs in the Generalized Goddard's Problem," *Journal of Optimization Theory and Applications*, Vol. 139, No. 2, 2008, pp. 439–461. doi:10.1007/s10957-008-9387-1

[22] Martinon, P., Bonnas, F., Laurent-Varin, J., and Trélat, E., "Numerical Study of Optimal Trajectories with Singular Arcs for an Ariane 5 Launcher," *Journal of Guidance, Control, and Dynamics*, Vol. 32, No. 1, 2009, pp. 51–55. doi:10.2514/1.37387

[23] Kennedy, J., and Eberhart, R., "Particle Swarm Optimization," *Proceedings of IEEE International Conference on Neural Networks*, Vol. 4, Perth, Australia, 1995, pp. 1942–1948.

[24] R. Sentinella, M., and Casalino, L., "Cooperative Evolutionary Algorithm for Space Trajectory Optimization," *Celestial Mechanics and Dynamical Astronomy*, Vol. 105, Nos. 1–3, 2009, pp. 211–227. doi:10.1007/s10569-009-9223-4

[25] Pontani, M., and Conway, B. A., "Particle Swarm Optimization Applied to Space Trajectories," *Journal of Guidance, Control, and Dynamics*, Vol. 33, No. 5, 2010, pp. 1429–1441. doi:10.2514/1.48475

[26] McConaghy, T. T., Debban, T. J., Petropoulos, A. E., and Longuski, J., "Design and Optimization of Low-Thrust Trajectories with Gravity Assists," *Journal of Spacecraft and Rockets*, Vol. 40, No. 3, 2003, pp. 380–387. doi:10.2514/2.3973

[27] Hull, D. G., *Optimal Control Theory for Applications*, Springer, New York, 2003, p. 167, Chaps. 6–17.

[28] Das, S., Abraham, A., and Konar, A., "Particle Swarm Optimization and Differential Evolution Algorithms: Technical Analysis, Applications and Hybridization Perspectives," *Advances of Computational Intelligence in Industrial Systems*, Studies in Computational Intelligence, Springer–Verlag, Berlin, 2008, pp. 1–34.

[29] Fehlberg, E., "Classical Fifth-, Sixth-, Seventh- and Eighth-Order Runge–Kutta Formulas with Stepsize Control," NASA TR R-287, 1968.

[30] More, J. J., Garbow, B. S., and Hillstrom, K. E., "User Guide for MinPack-1," Argonne National Lab., Rept. ANL-80-74, 1980; also http://www.netlib.org/minpack [retrieved 13 July 2010].

[31] Powell, M. J. D., "A Hybrid Method for Nonlinear Equations,"

Numerical Methods for Nonlinear Algebraic Equations, edited by P. Rabinowitz, Gordon and Breach, London, 1970, pp. 87–114.

[32] Sims, J. A., "Delta-V Gravity-Assist Trajectory Design: Theory and Practice," Ph.D. Dissertation, School of Aeronautics and Astronautics, Purdue Univ., West Lafayette, IN, 1997.

[33] Yan, C. H., McConaghy, T. T., Chen, K. J., and Longuski, J. M., "Preliminary Design of Nuclear Electric Propulsion Missions to the Outer Planets," AIAA/AAS Astrodynamics Specialist Conference and Exhibit, AIAA Paper 2004-5393, Providence, RI, 16–19 Aug. 2004.