Fuel optimal low thrust rendezvous with outer planets via gravity assist

ARTICLE *in* SCIENCE CHINA: PHYSICS, MECHANICS AND ASTRONOMY · APRIL 2011

Impact Factor: 1.14 · DOI: 10.1007/s11433-011-4291-3

4 AUTHORS:

Tieding Guo
Hunan University

Fanghua Jiang
Tsinghua University

Hexi Baoyin
Tsinghua University

Li Junfeng
Tsinghua University

SCIENCE CHINA
Physics, Mechanics & Astronomy

• Research Paper •
Special Issue: Forward for the Department of Engineering Mechanics, Tsinghua University

April 2011 Vol.54 No.4: 756–769
doi: 10.1007/s11433-011-4291-3

Fuel optimal low thrust rendezvous with outer planets via gravity assist

GUO TieDing, JIANG FangHua, BAOYIN HeXi* & LI JunFeng

School of Aerospace, Tsinghua University, Beijing 100084, China

Received December 12, 2010; accepted February 15, 2011; published online March 7, 2011

Low thrust propulsion and gravity assist (GA) are among the most promising techniques for deep space explorations. In this paper the two techniques are combined and treated comprehensively, both on modeling and numerical techniques. Fuel optimal orbit rendezvous via multiple GA is first formulated as optimal guidance with multiple interior constraints and then the optimal necessary conditions, various transversality conditions and stationary conditions are derived by Pontryagin's Maximum Principle (PMP). Finally the initial orbit rendezvous problem is transformed into a multiple point boundary value problem (MPBVP). Homotopic technique combined with random searching globally and Particle Swarm Optimization (PSO), is adopted to handle the numerical difficulty in solving the above MPBVP by single shooting method. Two scenarios in the end show the merits of the present approach.

low thrust, fuel optimal trajectory, maximum principle, homotopic technique, gravity assist

PACS: 95.55.Pe, 95.40.+s

1 Introduction

The past decade has witnessed the low thrust technique, like electrical and ion propulsion, gaining more and more support in aerospace, especially deep space explorations, for example mission Deep Space 1 [1] and HAYABUSA [2]. It is definitely the high specific impulse of low thrust that makes it so attractive, for this will render itself more efficient than traditional propulsion, i.e., chemical propulsion with high thrust magnitude. Most of low thrust trajectory optimization problems are resolved by the direct method [3–5], which formulates the optimal guidance problem as a parameter optimization problem by discretizing the whole continuous trajectory into many phases and finally resorts to nonlinear programming, like GALLOP [3], one widely used program based on parameter optimization for finding fuel-optimal low-thrust trajectory. The essence of the direct method in optimization is actually searching globally from a very broad admissible trajectory category and rendering the trajectories satisfying corresponding various constraints. This is really demanding in the respect of computation costs, considering the number of optimization parameters in most mission design is very large. And what's worse, without an account of the physics underlying the optimal guidance problem, it's not reasonable to expect too much from the direct method with regard to the global optimality of its results, as GALLOP claims, i.e., gravity-assisted low-thrust local optimization program. Indirect method [6–10] in trajectory optimization, on the other hand, taking into account more intrinsic optimal structures, unveiled by the optimal control theory [11–13], especially PMP, is believed likely to provide one with more reliable global optimal results at much fewer computation costs. However, difficulties from the indirect method seem formidable when the fuel optimal low thrust trajectory is treated. These difficulties lie in two respects, i.e., severe sensibility to initial interaction value, maybe leading to local optimal results, and small, even

*Corresponding author (email: baoyin@tsinghua.edu.cn)

irregular convergence domain, maybe leading to divergence of the single shooting method.

Recently, homotopic technique, one numerical continuation method, was introduced to handle the difficulties in fuel-optimal low thrust trajectory optimization [14–17]. The basic idea of the homotopic technique is to obtain the desired results by staring from the solution of a related, while easy-handling problem [14–18]. In the present context, this idea will become clear if one notes that the fuel optimal problem is less smooth, hence more difficult than the energy optimal problem, which can be solved first to provide an initial solution 'continued' then by the homotopic technique to the desired one.

The fight time is sometimes too long to make a mission feasible, however, if only the low thrust technique is employed [3]. Planetary gravity assist (GA), is believed to be one of the most promising techniques for outer planet explorations, which can be utilized to save fuel and shorten the flight time, especially for deep space missions to Jupiter, Uranus or further [3]. Actually some missions like Mariner 10 and Voyager 1 have utilized intermediate GA and various problems associated with GA are treated in refs. [19–22]. However, most of the researchers formulate the problem as a parameter optimization problem or consider only the impulsive case. Based upon the work in the aforementioned references, PMP and homotopic technique are employed in the present paper to treat this problem, i.e., fuel-optimal low-thrust rendezvous with an outer planet via GA, comprehensively in two respects, modeling and numerical technique.

The present paper extends the work in ref. [14] to multiple GA case and thus finishes the comprehensive treatment of the problem, i.e., fuel optimal low thrust rendezvous via GA. Based on Pontryagin's Maximum Principle (PMP), fuel-optimal low-thrust rendezvous via multiple GA is formulated as an optimal guidance problem with various interior point constraints. Then the optimal thrust is derived and various constraints, i.e., stationary conditions and transversality conditions, especially the ones introduced by each GA, are presented in detail. It is then straightforward to obtain a MPBVP, which is solved by the single shooting method, though daunted by severe sensibility to the initial interaction guess. This intrinsic difficulty due to local characteristics of gradient algorithms, say Newton's method or Powell's method [14–17], was even aggravated by the reduction of smoothness of shooting function caused by the less smooth performance index such as fuel optimal performance index. Thus the homotopic technique is employed to cope with this tough situation, by starting from one smoother optimization problem, say, energy optimal problem and then varying the so-called homotopic parameter to finally obtain the desired results of the fuel optimal problem. Actually the meaning of 'varying the homotopic parameter' here to solve a series of optimal guidance problems defined by the homotopic mapping, constructed here to relate the initial smooth optimization problem with the discontinuous one. With initialization difficulty in the energy optimal problem, i.e., sensibility to initial guess, being handled by the global searching algorithm PSO [14,23,24] or random searching globally [14,15], the fuel optimal rendezvous problem is finally solved. At the end of the paper, two scenarios, i.e., Earth- Earth-Jupiter and Earth-Earth-Jupiter-Uranus, are presented to illustrate the ideas and techniques in the paper.

2 Background: low thrust trajectory optimization by the indirect method

2.1 System dynamical model

The dynamics of spacecraft subject to Sun's gravity and electrical propulsion only, can be written as follows:

$$\dot{\boldsymbol{r}} = \boldsymbol{v}, \quad \dot{\boldsymbol{v}} = -\frac{\mu}{r^3}\boldsymbol{r} + \frac{T_{\max}u}{mm_0}\boldsymbol{\alpha}, \tag{1}$$

$$\dot{m} = -\frac{T_{\max}u}{I_{sp}g_0m_0}, \tag{2}$$

where $\boldsymbol{r},\boldsymbol{v}$ are respectively the spacecraft's position and velocity vectors in the heliocentric ecliptic reference frame (HERF). The dimensionless mass of fuel m can be obtained by making $m(t_0)=1$, and here m_0 is the initial total mass. Denoting the maximal thrust magnitude, dimensionless thrust magnitude and direction of thrust by $T_{\max}, u, \boldsymbol{\alpha}$ respectively, the thrust vector can now be denoted by $\boldsymbol{T} = (T_{\max}u)\boldsymbol{\alpha}$, here $0 \leqslant u \leqslant 1$, $\|\boldsymbol{\alpha}\| = 1$. Other constants like μ, I_{sp}, g_0 are Sun's gravitational constants, engine's specific impulse and acceleration of gravity at see level respectively. With the fuel mass m combined together with spacecraft's mechanical state variables $\boldsymbol{r},\boldsymbol{v}$, the state variables of the mass-varying spacecraft system can be denoted by $\boldsymbol{x} = [\boldsymbol{r}, \boldsymbol{v}, m]$. And the system dynamical equation can now be written in a compact form

$$\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u, \boldsymbol{\alpha}). \tag{3}$$

2.2 Fuel optimal low thrust rendezvous

Fuel optimal low thrust rendezvous can be formulated as an optimal control problem and the associated optimal thrust program can be derived by PMP [14,15]. This will be presented briefly in the following.

The performance index of fuel optimal rendezvous is

$$\tilde{J} = -m(t_f). \tag{4}$$

The optimal thrust program should make the above index as small as possible. In other words, the final fuel $m(t_f)$ should be as much as possible.

Note that multiplying the performance index $\tilde{J}$ by any plus constant has no influence on the optimal control and the optimal trajectory. The following equivalent and easy-handling performance index J will be adopted here

$$J = \lambda_0 \frac{T_{\max}}{I_{sp} g_0 m_0} \int_{t_0}^{t_f} u \mathrm{d}t, \tag{5}$$

where $\lambda_0 > 0$ is a factor. And the equivalence will be clear after the mass differential equation (2) is integrated.

For optimal control, Hamiltonian H is denoted by

$$\begin{aligned} H = {} & \boldsymbol{\lambda}_r \cdot \boldsymbol{v} + \boldsymbol{\lambda}_v \cdot \left(-\frac{\mu}{r^3}\boldsymbol{r} + \frac{T_{\max} u}{m m_0}\boldsymbol{\alpha} \right) \\ & + \lambda_m \left(-\frac{T_{\max} u}{I_{sp} g_0 m_0} \right) + \lambda_0 \frac{T_{\max}}{I_{sp} g_0 m_0} u, \end{aligned} \tag{6}$$

where $\boldsymbol{\lambda}(t)$ is the costate, associated with the state equation constraint $\boldsymbol{f}(\boldsymbol{x},\boldsymbol{\alpha},u) - \dot{\boldsymbol{x}} = 0$.

According to PMP, the optimal control can be derived

$$(u, \boldsymbol{\alpha}) = \arg \min_{\|\boldsymbol{\alpha}\|=1, u\in[0 \quad 1]} H. \tag{7}$$

For $u \geqslant 0$, the optimal $\boldsymbol{\alpha}$ can be obtained directly

$$\boldsymbol{\alpha} = -\frac{\boldsymbol{\lambda}_v}{\|\boldsymbol{\lambda}_v\|}. \tag{8}$$

The above optimal thrust direction is the well-known prime vector, introduced by Lawden [25], which indicates that the optimal thrust should always be in the opposite direction of costate associated with velocity, say $\boldsymbol{\lambda}_v$, and this conclusion is independent of the thrust magnitude u.

It is more involved to get the optimal u, however. The Hamiltonian H should be first transformed into a cleaner form utilizing optimal thrust direction (8)

$$\begin{aligned} H = {} & \frac{T_{\max}}{I_{sp} g_0 m_0} \left(\lambda_0 - \lambda_m - \frac{\|\boldsymbol{\lambda}_v\| I_{sp} g_0}{m} \right) u \\ & + \boldsymbol{\lambda}_r \cdot \boldsymbol{v} - \boldsymbol{\lambda}_v \cdot \frac{\mu}{r^3}\boldsymbol{r}. \end{aligned} \tag{9}$$

Note that state differential equations (1) (2) and Hamiltonian (9) are all linear with respect to the magnitude of thrust u, whose admissible domain is closed, i.e., $u \in [0,1]$. Thus the optimal thrust magnitude u is derived as bang-bang control in the following. With the auxiliary so-called switching function S denoted by

$$S = \lambda_0 - \lambda_m - \frac{\|\boldsymbol{\lambda}_v\| I_{sp} g_0}{m}, \tag{10}$$

the optimal control can now be derived as

$$u = \begin{cases} 1 & S < 0, \\ 0 & S > 0, \\ \text{indefinite} & S = 0. \end{cases} \tag{11}$$

If $S \equiv 0$ in a certain continuous domain, the situation becomes singular and subtle. However, one will be assured by numerical results in the end, displaying that the switching function S reaches zero only at some finite points. Note that the above bang-bang control (11) is discontinuous and some troublesome problems will be expected in numerical computation. Actually this is one of the elements that renders the fuel-optimal problem less smooth, hence more difficult than the energy optimal problem.

The costate equation can then be derived by $\dot{\boldsymbol{\lambda}} = -\partial H / \partial \boldsymbol{x}$

$$\begin{aligned} & \dot{\boldsymbol{\lambda}}_r = \boldsymbol{\lambda}_v \frac{\mu}{r^3} - \frac{3\mu(\boldsymbol{\lambda}_v \cdot \boldsymbol{r})}{r^5}\boldsymbol{r}, \quad \dot{\boldsymbol{\lambda}}_v = -\boldsymbol{\lambda}_r, \\ & \dot{\lambda}_m = -\frac{T_{\max}\|\boldsymbol{\lambda}_v\| u}{m_0 m^2}, \end{aligned} \tag{12}$$

where $\boldsymbol{\lambda} = [\boldsymbol{\lambda}_r \quad \boldsymbol{\lambda}_v \quad \lambda_m]$ are the costate variables associated with the state variables of the spacecraft system.

It becomes clear that state differential equation (3) and costate differential equations (12) are really coupled together via the optimal control u and $\boldsymbol{\alpha}$ through some inspection of eqs. (8) (10) (11). Thus state equations (3) and costate equations (12), together with various boundary conditions (initial or terminal conditions) completely define the fuel optimal rendezvous problem.

Explicitly, in guidance problem (3) (12) initial conditions are only partially imposed

$$\boldsymbol{r}(t_0) = \boldsymbol{r}_0, \boldsymbol{v}(t_0) = \boldsymbol{v}_0, m(t_0) = 1. \tag{13}$$

And part of the terminal conditions are

$$\boldsymbol{r}(t_f) = \boldsymbol{r}_f, \boldsymbol{v}(t_f) = \boldsymbol{v}_f. \tag{14}$$

The remaining boundary conditions are supplied by the so-called transversallity conditions

$$\lambda_m(t_f) = 0. \tag{15}$$

Thus the optimal rendezvous problem leads to a typical two point boundary value problem (TPBVP) (3) (12)–(15), totally different from common ordinary differential equations. Solving the above TPBVP efficiently is definitely nontrivial, and at most time one has to rely on the single shooting method, based on some local algorithms like Newton's or Powell's method.

3 Fuel optimal low thrust rendezvous via gravity assist

3.1 Impulsive planetary gravity assist model

Some digestion of planetary gravity assist (GA) is needed before treating optimal rendezvous via GA. In the present paper, impulsive planetary gravity assist model is adopted, that is the time spent in the sphere of influence of a planet is neglected. With the superscripts − and + denoting the time just before and just after GA respectively, we can summarize the fundamental properties of planetary gravity assist model [26] as follows.

Noting the time spent in the process of gravity assist is neglected and assuming the position of spacecraft just before and after gravity assist is equal to the position of planet, one can obtain

$$t_a^- = t_a^+ = t_a, \tag{16}$$

$$\boldsymbol{r}\left(t_a^-\right) = \boldsymbol{r}\left(t_a^+\right) = \boldsymbol{r}_p\left(t_a\right), \tag{17}$$

where t_a is the time for GA and the velocities of spacecraft just before and after gravity assist in HERF are denoted by $\boldsymbol{v}\left(t_a^-\right)$, $\boldsymbol{v}\left(t_a^+\right)$ respectively. Thus the so-called incoming and outgoing hyperbolic excess velocity with respect to the planet can be introduced as $\boldsymbol{v}_\infty^- = \boldsymbol{v}\left(t_a^-\right) - \boldsymbol{v}_p\left(t_a\right), \boldsymbol{v}_\infty^+ = \boldsymbol{v}\left(t_a^+\right) - \boldsymbol{v}_p\left(t_a\right)$, where $\boldsymbol{v}_p\left(t_a\right)$ is the velocity of the planet at GA time t_a. Then, the most important formulas with regard to GA can be written as

$$\left\|\boldsymbol{v}_\infty^-\right\| = \left\|\boldsymbol{v}_\infty^+\right\| = v_\infty, \tag{18}$$

$$\boldsymbol{v}_\infty^- \cdot \boldsymbol{v}_\infty^+ = v_\infty^2 \cos\delta, \tag{19}$$

$$\sin\left(\frac{\delta}{2}\right) = \frac{1}{1+\left(v_\infty^2 r_{\text{periapsis}} / \mu_p\right)}, \tag{20}$$

where δ denotes the so-called turn angle between the hyperbolic excess velocity before and after gravity assist, and $r_{\text{periapsis}}$, μ_p stand for the periapsis radius of the GA hyperbolic orbit and gravitational constants of planet respectively. The trade-off between the turn angle δ and the magnitude of hyperbolic excess velocity v_∞, and the periapsis radius $r_{\text{periapsis}}$, is clear in eq. (20), that is the decrease of v_∞, $r_{\text{periapsis}}$ is always accompanied by the increase of turn angle δ. However, with the hyperbolic excess velocity v_∞ specified, the maximum turn angle is always limited by the minimum periapsis radius $r_{\min}$ of the GA hyperbolic orbit, if the surface or the extent of the atmosphere of the planet are considered. So one more inequality constraint, in design rather than physics, should be introduced here

$$r_{\text{periapsis}} \geqslant r_{\min}. \tag{21}$$

3.2 Optimal rendezvous via multiple gravity assist

All that stated in sect. 2.2 are standard optimal control and the associated TPBVP encountered in low thrust trajectory optimization by the indirect method. Now, various interior point constraints, i.e., multiple GA, will be introduced and the situation becomes a little tough. Keep in mind that various constraints at some specific time or state, like the terminal point or time, or some intermediate time etc, affect only the associated stationary or transversality conditions at that time or state, with no effect on costate equations or first order optimal necessary conditions. This means all that above about the TPBVP is still right in the present case and the new elements here are the associated stationary and transversality conditions introduced by various interior point constraints at that time.

Let various equality and inequality constraints introduced by GA at successive time points, i.e., $t_i, i = 1 \cdots N$, be denoted by

$$\begin{aligned} &\boldsymbol{\varphi}_i\left(\boldsymbol{x}_d\left(t_i^-\right), \boldsymbol{x}_d\left(t_i^+\right), \boldsymbol{x}_c\left(t_i\right), t_i\right) = 0, \\ &\psi_i\left(\boldsymbol{x}_d\left(t_i^-\right), \boldsymbol{x}_d\left(t_i^+\right), \boldsymbol{x}_c\left(t_i\right), t_i\right) \leqslant 0, \end{aligned} \tag{22}$$

where $\boldsymbol{\varphi}_i$ and ψ_i in the present case can be expressed explicitly , i.e., for each GA

$$\begin{aligned} \boldsymbol{\varphi}_i &= \left[\boldsymbol{r}\left(t_i\right) - \boldsymbol{r}_p\left(t_i\right), \left\|\boldsymbol{v}_\infty^-\right\| - \left\|\boldsymbol{v}_\infty^+\right\|\right] = \boldsymbol{0}, \\ \psi_i &= r_{\min} - r_{\text{periapsis}} \leqslant 0. \end{aligned} \tag{23}$$

And here $\boldsymbol{x}_d\left(t_i^-\right), \boldsymbol{x}_d\left(t_i^+\right)$ denote the discontinuous part of state variables just before and after the interior constraint respectively, and $\boldsymbol{x}_c\left(t_i\right)$ the continuous ones. In other words, here we have adopted such a notation convention $\boldsymbol{x} = \left[\boldsymbol{x}_d, \boldsymbol{x}_c\right]$. One should be warned that this partition of state variables is indispensable, though absent in ref. [21], and actually this is one of the characteristics of interior constraints introduced by GA, i.e., part of the state variables are discontinuous while the remaining are not. The above notation conventions will be used in the following. The superscripts − and + denote the quantities just before and after the interior constraints respectively and state variables are partitioned into two distinct parts in the sense of continuity.

With terminal constraints denoted by $\boldsymbol{\phi}\left(\boldsymbol{x}\left(t_f\right)\right) = 0$ and constant Lagrange multipliers associated with each equality constraint $\boldsymbol{\varphi}_i$ and inequality constraint ψ_i denoted by $\boldsymbol{\kappa}_i$ and ν_i respectively, one can obtain the augmented performance index as follows:

$$\begin{aligned}\bar{J} = &\sum_{i=1}^{N}\Big[\boldsymbol{\kappa}_i \cdot \boldsymbol{\varphi}_i\left(\boldsymbol{x}_d\left(t_i^-\right), \boldsymbol{x}_d\left(t_i^+\right), \boldsymbol{x}_c\left(t_i\right), t_i\right) \\ &+ \boldsymbol{v}_i \cdot \psi_i\left(\boldsymbol{x}_d\left(t_i^-\right), \boldsymbol{x}_d\left(t_i^+\right), \boldsymbol{x}_c\left(t_i\right), t_i\right)\Big] \\ &+ \boldsymbol{\mu} \cdot \boldsymbol{\phi}\left(\boldsymbol{x}\left(t_f\right)\right) + \int_{t_0}^{t_1^-}\left(H - \boldsymbol{\lambda} \cdot \dot{\boldsymbol{x}}\right)\mathrm{d}t \\ &+ \sum_{i=1}^{N-1}\int_{t_i^+}^{t_{i+1}^-}\left(H - \boldsymbol{\lambda} \cdot \dot{\boldsymbol{x}}\right)\mathrm{d}t + \int_{t_N^+}^{t_f}\left(H - \boldsymbol{\lambda} \cdot \dot{\boldsymbol{x}}\right)\mathrm{d}t. \end{aligned} \tag{24}$$

Actually here Kuhn-Tucker condition is used. In other words multiplier v_i should satisfy

$$v_i \geqslant 0,\ v_i \cdot \psi_i\left(\boldsymbol{x}_d\left(t_i^-\right), \boldsymbol{x}_d\left(t_i^+\right), \boldsymbol{x}_c\left(t_i\right), t_i\right) = 0. \tag{25}$$

Let $\mathrm{d}t_m$, $\delta\boldsymbol{x}_d(t_i^-)$, $\delta\boldsymbol{x}_d(t_i^+)$, $\delta\boldsymbol{x}_c(t_i)$, δu, $\delta\boldsymbol{x}(t)$ denote the first variation of t_i, $\boldsymbol{x}_d(t_i^-)$, $\boldsymbol{x}_d(t_i^+)$, $\boldsymbol{x}_c(t_i)$, u, $\boldsymbol{x}(t)$ respectively. The first variation of performance index $\bar{J}$ can now be derived as follows:

$$\begin{aligned}\delta\bar{J} = &\sum_{i=1}^{N}\left\{\frac{\partial\Psi_i}{\partial t_i}\mathrm{d}t_i + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^-)}\left[\delta\boldsymbol{x}_d(t_i^-) + \dot{\boldsymbol{x}}_d(t_i^-)\mathrm{d}t_i\right] + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^+)}\right. \\ &\left.\times\left[\delta\boldsymbol{x}_d(t_i^+) + \dot{\boldsymbol{x}}_d(t_i^+)\mathrm{d}t_i\right] + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_c(t_i)}\left[\delta\boldsymbol{x}_c(t_i) + \dot{\boldsymbol{x}}_c(t_i)\mathrm{d}t_i\right]\right\} \\ &+ \sum_{i=1}^{N}\left\{-\left(H - \boldsymbol{\lambda}\cdot\dot{\boldsymbol{x}}\right)\Big|_{t_i^+}\mathrm{d}t_i + \left(H - \boldsymbol{\lambda}\cdot\dot{\boldsymbol{x}}\right)\Big|_{t_i^-}\mathrm{d}t_i\right\} - \boldsymbol{\lambda}\cdot\delta\boldsymbol{x}\Big|_{t_0}^{t_1^-} \\ &- \sum_{i=1}^{N-1}\left[\boldsymbol{\lambda}\cdot\delta\boldsymbol{x}\Big|_{t_i^+}^{t_{i+1}^-}\right] - \boldsymbol{\lambda}\cdot\delta\boldsymbol{x}\Big|_{t_N^+}^{t_f} + \frac{\partial\left(\boldsymbol{\mu}\cdot\boldsymbol{\phi}\right)}{\partial\boldsymbol{x}(t_f)}\delta\boldsymbol{x}(t_f) \\ &+ \int_{t_0}^{t_f}\left[\frac{\partial H}{\partial u}\delta u + \left(\dot{\boldsymbol{\lambda}} + \frac{\partial H}{\partial\boldsymbol{x}}\right)\delta\boldsymbol{x}\right]\mathrm{d}t, \end{aligned} \tag{26}$$

where $\Psi_i = \boldsymbol{\kappa}_i \cdot \boldsymbol{\varphi}_i + v_i \cdot \psi_i$.

Note here for continuous state variables $\boldsymbol{x}_c$ the following rules have been utilized [11,13]

$$\mathrm{d}\boldsymbol{x}_c(t_i) = \begin{cases}\delta\boldsymbol{x}_c(t_i^-) + \dot{\boldsymbol{x}}_c(t_i^-)\mathrm{d}t_i, \\ \delta\boldsymbol{x}_c(t_i^+) + \dot{\boldsymbol{x}}_c(t_i^+)\mathrm{d}t_i.\end{cases} \tag{27}$$

And for discontinuous parts, as pointed out in ref. [11], $\mathrm{d}\boldsymbol{x}_c(t_i^-) \neq \mathrm{d}\boldsymbol{x}_c(t_i^+)$, which should be both treated as independent variations.

With the costates $\boldsymbol{\lambda}$ classified into two distinct categories: one associated with continuous state variables, say $\boldsymbol{\lambda}_c$, and the other with discontinuous ones, say $\boldsymbol{\lambda}_d$, the above seemingly formidable expression can be written in a more compact form after some simplifications

$$\begin{aligned}\delta\bar{J} = &\sum_{i=1}^{N}\left\{\left[-\boldsymbol{\lambda}_d\left(t_i^-\right) + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^-)}\right]\left[\delta\boldsymbol{x}_d(t_i^-) + \dot{\boldsymbol{x}}_d(t_i^-)\mathrm{d}t_i\right]\right. \\ &+ \left[\boldsymbol{\lambda}_d\left(t_i^+\right) + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^+)}\right]\left[\delta\boldsymbol{x}_d(t_i^+) + \dot{\boldsymbol{x}}_d(t_i^+)Et_i\right] \\ &\left.+ \left[\frac{\partial\Psi_i}{\partial\boldsymbol{x}_c(t_i)} - \boldsymbol{\lambda}_c\left(t_i^-\right) + \boldsymbol{\lambda}_c\left(t_i^+\right)\right]\left[\delta\boldsymbol{x}_c(t_i) + \dot{\boldsymbol{x}}_c(t_i)\mathrm{d}t_i\right]\right\} \\ &+ \sum_{i=1}^{N}\left\{\left[H\left(t_i^-\right) - H\left(t_i^+\right) + \partial\Psi_i/\partial t_i\right]\mathrm{d}t_i\right\} + \left[\frac{\partial\left(\boldsymbol{\mu}\cdot\boldsymbol{\phi}\right)}{\partial\boldsymbol{x}(t_f)} - \boldsymbol{\lambda}\right] \\ &\times\delta\boldsymbol{x}(t_f) + \int_{t_0}^{t_f}\left[\frac{\partial H}{\partial u}\delta u + \left(\dot{\boldsymbol{\lambda}} + \frac{\partial H}{\partial\boldsymbol{x}}\right)\delta\boldsymbol{x}\right]\mathrm{d}t. \end{aligned} \tag{28}$$

For variations δu and $\delta\boldsymbol{x}$ in the integrand to be free, costates $\boldsymbol{\lambda}$ in the above should be chosen to make coefficients of the other variations vanish at each GA t_i, $i = 1\cdots N$. And then the so-called transversality and stationary conditions introduced by each GA, can be obtained like

$$-\boldsymbol{\lambda}_d\left(t_i^-\right) + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^-)} = 0, \tag{29}$$

$$\boldsymbol{\lambda}_d\left(t_i^+\right) + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^+)} = 0, \tag{30}$$

$$-\boldsymbol{\lambda}_c\left(t_i^-\right) + \boldsymbol{\lambda}_c\left(t_i^+\right) + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_c(t_i)} = 0\,, \tag{31}$$

$$H\left(t_i^-\right) - H\left(t_i^+\right) + \frac{\partial\Psi_i}{\partial t_i} = 0\,, \tag{32}$$

where the first three are transversal conditions at interior points $\boldsymbol{r}_p\left(t_i\right)$ and the last one is stationary conditions associated with each interior constraint time t_i. Setting $\delta\bar{J} = 0$, then the terminal transversality conditions $\boldsymbol{\lambda}\left(t_f\right) = \partial\left(\boldsymbol{\mu}\cdot\boldsymbol{\phi}\right)/\partial\boldsymbol{x}(t_f)$, costate equations $\dot{\boldsymbol{\lambda}} = -\partial H/\partial\boldsymbol{x}$, and the first order necessary condition for optimality, i.e., $\partial H/\partial u = 0$, can all be obtained.

Although the above derivations apply to the classical case only, i.e., without constraints on control variables or mathematically, the admissible control domain U is an open set, the results or conclusions presented here are right in both cases, i.e., with or without constraints on control variables. And even more, in the latter case, the present local conclusions, say the first order necessary condition of extremum $\partial H/\partial u = 0$, can be strengthened to a global one, say strong minimum in the whole admissible domain, that is $u = \arg\min_{u\in U} H$, which has been utilized before in the paper. The strengthened version of the optimality condition shows the power of PMP and anyone interested in the whole story behind this amazing result is referred to the classic [12].

3.3 Constraints introduced by multiple GA

Various constraints introduced by multiple GA, both on state variables $\boldsymbol{x}(t)$ and costate variables $\boldsymbol{\lambda}(t)$, will be treated completely in this section. Following Jiang's work in the one GA case [14], one can derive constraints in the present case. The interesting point here is that some symmetric properties are utilized and the derivations are simplified.

The first part, i.e., constraints imposed on the state variables are evident, that is $\boldsymbol{r}(t_i)=\boldsymbol{r}_p(t_i)$ and $\|\boldsymbol{v}_\infty^-\|=\|\boldsymbol{v}_\infty^+\|=v_\infty$. Note that here the constraints are all on mechanical quantities, for one can assume that the mass variable m is continuous, that is $m(t_i^-)=m(t_i^+)$, considering the time during GA the planet is so short in comparison with the total flight time $TOF=t_f-t_0$. The last one is associated with the design instead of physics, i.e., $r_{\text{periapsis}}\geqslant r_{\min}$.

Anyway, the aforementioned constraints are all relatively intuitive and easy to grasp or write down. The other part, i.e., constraints on the costates, however, are not so evident, the derivation of which is actually much more involved. This will be finished in the following based on the results in the last section.

In accordance with the terminology used in the last section, here position vector $\boldsymbol{r}$ and mass m are continuous state variables, i.e., $\boldsymbol{x}_c=[\boldsymbol{r},m]$, while the remaining velocity vector $\boldsymbol{v}$ the discontinuous ones, i.e., $\boldsymbol{x}_d=\boldsymbol{v}$. Accordingly, one has $\boldsymbol{\lambda}_c=[\boldsymbol{\lambda}_r,\lambda_m]$ and $\boldsymbol{\lambda}_d=\boldsymbol{\lambda}_v$. The constraint terms for each GA, i.e., $\Psi_i=\boldsymbol{\kappa}_i\cdot\boldsymbol{\varphi}_i+\nu_i\cdot\psi_i$, can be presented explicitly like

$$\begin{aligned}\Psi_i=&\boldsymbol{\kappa}_{i1}\cdot\left(\boldsymbol{r}(t_i)-\boldsymbol{r}_p(t_i)\right)\\&+\kappa_{i2}\cdot\left(\|\boldsymbol{v}_\infty^-\|-\|\boldsymbol{v}_\infty^+\|\right)+\nu_i\cdot\left(r_{\min}-r_{\text{periapsis}}\right),\end{aligned}\tag{33}$$

where

$$\boldsymbol{\varphi}_i=\left[\boldsymbol{r}(t_i)-\boldsymbol{r}_p(t_i),\|\boldsymbol{v}_\infty^-\|-\|\boldsymbol{v}_\infty^+\|\right]=\boldsymbol{0}$$

and $\psi_i=r_{\min}-r_{\text{periapsis}}\leqslant 0$ are the equality and inequality constraints associated with each GA respectively. Note here $\boldsymbol{\kappa}_i=[\boldsymbol{\kappa}_{i1},\kappa_{i2}]$ and $\nu_i\geqslant 0$.

Now with some care one can deduce the corresponding transversality and stationary conditions introduced by each GA. From inspection of eqs. (29)–(32), one can know all that needed are the follows:

$$\frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^-)},\ \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^+)},\ \frac{\partial\Psi_i}{\partial\boldsymbol{x}_c(t_i)},\ \frac{\partial\Psi_i}{\partial t_i}.$$

And actually all the following derivations are straightforward

$$\begin{aligned}&\frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^-)}=\kappa_{i2}\cdot\frac{\boldsymbol{v}_\infty^-}{\|\boldsymbol{v}_\infty^-\|}-\nu_i\cdot\frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^-)},\\&\frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^+)}=-\kappa_{i2}\cdot\frac{\boldsymbol{v}_\infty^+}{\|\boldsymbol{v}_\infty^+\|}-\nu_i\cdot\frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^+)},\\&\frac{\partial\Psi_i}{\partial\boldsymbol{x}_c(t_i)}=[\boldsymbol{\kappa}_{i1},0],\\&\frac{\partial\Psi_i}{\partial t_i}=\kappa_{i2}\cdot\left(-\frac{\boldsymbol{v}_\infty^-}{\|\boldsymbol{v}_\infty^-\|}+\frac{\boldsymbol{v}_\infty^+}{\|\boldsymbol{v}_\infty^+\|}\right)\cdot\frac{\partial\boldsymbol{v}_p(t_i)}{\partial t_i}\\&\qquad-\boldsymbol{\kappa}_{i1}\cdot\frac{\partial\boldsymbol{r}_p(t_i)}{\partial t_i}-\nu_i\cdot\frac{\partial r_{\text{periapsis}}}{\partial t_i},\end{aligned}\tag{34}$$

Note that differentiation with respect to GA time t_i in the above should only apply to the target's states , or specified reference quantities or signals, if borrowing terminology from control theory, like the position and velocity of planet in the present case, i.e., $\boldsymbol{r}_p(t_i)$ and $\boldsymbol{v}_p(t_i)$, while not to the state variables of spacecraft. This remark becomes clear if close inspection is given to the derivation of first variation of augmented performance index $\delta\overline{J}$ in eqs. (24) (26).

The key point now is the differentiation of periapsis radius of the GA hyperbolic orbit $r_{\text{periapsis}}$ with respect to GA time t_i, incoming and outgoing velocities $\boldsymbol{v}(t_i^-)$, $\boldsymbol{v}(t_i^+)$. Rearrange expression (20)

$$r_{\text{periapsis}}=\frac{\mu_p}{v_\infty^2}\left(\frac{1}{\sin(\delta/2)}-1\right),\tag{35}$$

where turn angle δ satisfies $\cos\delta=\boldsymbol{v}_\infty^-\cdot\boldsymbol{v}_\infty^+/v_\infty^2$ and $v_\infty=\|\boldsymbol{v}_\infty^-\|=\|\boldsymbol{v}_\infty^+\|$. Warnings should be given that the auxiliary term v_∞^2 in the above expression is actually misleading, as one can be really tempted to replace v_∞^2 with $\|\boldsymbol{v}_\infty^-\|^2$, or $\|\boldsymbol{v}_\infty^+\|^2$, or $\|\boldsymbol{v}_\infty^+\|\cdot\|\boldsymbol{v}_\infty^-\|$ while differentiating partially with respect to $\boldsymbol{v}(t_i^-)$ and $\boldsymbol{v}(t_i^+)$. Simple computation trials show that the above three are far from equivalent in the sense of partial differentiation, however. This paradox, originating in deviations of impulsive GA model from actual physics, can be removed by replacing the misleading terms v_∞^2 in (35) with $\|\boldsymbol{v}_\infty^+\|\cdot\|\boldsymbol{v}_\infty^-\|$, consisting of fundamental variables $\boldsymbol{v}(t_i^-),\boldsymbol{v}(t_i^+)$ in a symmetric form, which is actually a suggestion based on the symmetry of GA hyperbolic orbit in physics.

Now the more physics-involving periapsis radius $r_{\text{periapsis}}$ can be written as

$$r_{\text{periapsis}} = \frac{\mu_p}{\left\|\boldsymbol{v}_\infty^+\right\| \cdot \left\|\boldsymbol{v}_\infty^-\right\|}\left(\frac{1}{\sin(\delta/2)} - 1\right), \tag{36}$$

where $\cos\delta = \boldsymbol{v}_\infty^- \cdot \boldsymbol{v}_\infty^+ \Big/ \left(\left\|\boldsymbol{v}_\infty^+\right\| \cdot \left\|\boldsymbol{v}_\infty^-\right\|\right)$. The symmetric relation between $\boldsymbol{v}_\infty^-$ and $\boldsymbol{v}_\infty^+$ in the above expression is really clear, hence $\boldsymbol{v}(t_i^-) = \boldsymbol{v}_\infty^- + \boldsymbol{v}_p(t_i)$ and $\boldsymbol{v}(t_i^+) = \boldsymbol{v}_\infty^+ + \boldsymbol{v}_p(t_i)$. This means $\partial/\partial\boldsymbol{v}(t_i^-)$ and $\partial/\partial\boldsymbol{v}(t_i^+)$ will have totally the same form, if we exchange $\boldsymbol{v}(t_i^-)$ with $\boldsymbol{v}(t_i^+)$.

With regard to $\boldsymbol{v}_p(t_i)$, one notes that

$$\frac{\partial}{\partial\boldsymbol{v}_p(t_i)} = -\frac{\partial}{\partial\boldsymbol{v}(t_i^-)} - \frac{\partial}{\partial\boldsymbol{v}(t_i^+)}. \tag{37}$$

Furthermore, keeping in mind the aforementioned note or warning about partial differentiation with respect to constraint time t_m, and recalling

$$\frac{\partial r_{\text{periapsis}}}{\partial t_i} = \frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}_p(t_i)} \cdot \frac{\partial\boldsymbol{v}_p(t_i)}{\partial t_i}, \tag{38}$$

one will arrive at the conclusion that all that is needed now is either one of the two: $\dfrac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^-)}$, $\dfrac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^+)}$. With some care, one can derive $\dfrac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^-)}$ explicitly

$$\frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^-)} = \frac{1}{4}\frac{\mu_p}{v_\infty^4 \sin^3(\delta/2)} \times\left[\left(1 + 2\sin^2(\delta/2) - 4\sin^3(\delta/2)\right)\boldsymbol{v}_\infty^- + \boldsymbol{v}_\infty^+\right]. \tag{39}$$

Now the above just tiny observation will save one from tedious computations, say, one can obtain $\dfrac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^+)}$ directly by just exchanging $\boldsymbol{v}(t_i^-)$ with $\boldsymbol{v}(t_i^+)$, and $\boldsymbol{v}(t_i^+)$ with $\boldsymbol{v}(t_i^-)$ in eq. (38),

$$\frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^+)} = \frac{1}{4}\frac{\mu_p}{v_\infty^4 \sin^3(\delta/2)} \times\left[\left(1 + 2\sin^2(\delta/2) - 4\sin^3(\delta/2)\right)\boldsymbol{v}_\infty^+ + \boldsymbol{v}_\infty^-\right]. \tag{40}$$

And $\dfrac{\partial r_{\text{periapsis}}}{\partial t_i}$ is obtained using eqs. (37) and (38),

$$\begin{aligned}\frac{\partial r_{\text{periapsis}}}{\partial t_i} &= \frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}_p(t_i)} \cdot \frac{\partial\boldsymbol{v}_p(t_i)}{\partial t_i} \\ &= -\left[\frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^-)} + \frac{\partial r_{\text{periapsis}}}{\partial\boldsymbol{v}(t_i^+)}\right] \cdot \frac{\partial\boldsymbol{v}_p(t_i)}{\partial t_i}.\end{aligned} \tag{41}$$

Note here

$$\frac{\partial\boldsymbol{v}_p(t_i)}{\partial t_i} = -\frac{\mu_p}{\left\|\boldsymbol{r}_p(t_i)\right\|^3}\boldsymbol{r}_p(t_i). \tag{42}$$

Substituting the above (39)–(42) into (34), one finally obtains the transversality and stationary conditions introduced by each GA, i.e., (29) –(32).

3.4 Formulation of the MPBVP

To see the whole picture, one should summarize all the points up to now and then formulate the optimal rendezvous problem with multiple GA as MPBVP.

The initial conditions at t_0 are clear

$$\boldsymbol{r}(t_0) = \boldsymbol{r}_0,\ \boldsymbol{v}(t_0) = \boldsymbol{v}_0,\ m(t_0) = 1. \tag{43}$$

The constraints on state variables introduced by each GA are

$$\begin{gathered}\boldsymbol{r}\left(t_i^-\right) - \boldsymbol{r}_p(t_i) = 0;\ \left\|\boldsymbol{v}_\infty^-\right\| - \left\|\boldsymbol{v}_\infty^+\right\| = 0, \\ \nu_i \cdot \left(r_{\min} - r_{\text{periapsis}}\right) = 0,\ \nu_i \geqslant 0.\end{gathered} \tag{44}$$

The constraints on costate variables introduced by each GA at t_i^-, i.e., transversality conditions are

$$-\boldsymbol{\lambda}_d\left(t_i^-\right) + \frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^-)} = 0. \tag{45}$$

The constraint on optimal GA time t_i, in other words, stationary condition is

$$H\left(t_i^-\right) - H\left(t_i^+\right) + \frac{\partial\Psi_i}{\partial t_i} = 0. \tag{46}$$

The constraints at terminal time t_f are

$$\boldsymbol{r}(t_f) - \boldsymbol{r}_f = 0,\ \boldsymbol{v}(t_f) - \boldsymbol{v}_f = 0,\ \lambda_m\left(t_f\right) = 0. \tag{47}$$

State equations (3) together with costate equations (12), and all the conditions or constraints at initial, terminal and each GA time, i.e., t_0, t_f, t_i, $i = 1\cdots N$, finally constitute the MPBVP. And the basic ideas of solving the MPBVP by the single shooting method is to transforming the former into an initial value problem of ODE (ordinary differential equation) by assigning or guessing enough so-called shooting variables, like initial conditions, corner conditions, if interior points exist, and maybe some other parameters such as GA time t_i, to make sure that the transformed initial value problem, maybe multiple stages or legs involved, is well-defined.

In the present case, to start integrating forward at initial time t_0, one should first assign or guess the initial values of costates: $\boldsymbol{\lambda}(t_0)=\left[\boldsymbol{\lambda}_r(t_0),\boldsymbol{\lambda}_v(t_0),\lambda_m(t_0)\right]$. At t_i^+, i.e., just after GA, to carry up integration, velocity $\boldsymbol{v}(t_i^+)$, or equivalently, actually more conveniently, velocity difference $\Delta\boldsymbol{v}_i=\boldsymbol{v}(t_i^+)-\boldsymbol{v}(t_i^-)$ should be assigned, considering that $\boldsymbol{\lambda}_d(t_i^+)=\boldsymbol{\lambda}_v(t_i^+)$, $\boldsymbol{\lambda}_c(t_i^+)=\left[\boldsymbol{\lambda}_r(t_i^+),\lambda_m(t_i^+)\right]$, $\boldsymbol{r}(t_i^+)$ and $m(t_i^+)$ can be supplied by the following relations

$$\boldsymbol{r}\left(t_i^+\right)=\boldsymbol{r}\left(t_i^-\right),\ m\left(t_i^+\right)=m\left(t_i^-\right), \tag{48}$$

$$\boldsymbol{\lambda}_d\left(t_i^+\right)+\frac{\partial\Psi_i}{\partial\boldsymbol{x}_d(t_i^+)}=0, \tag{49}$$

$$-\boldsymbol{\lambda}_c\left(t_i^-\right)+\boldsymbol{\lambda}_c\left(t_i^+\right)+\frac{\partial\Psi_i}{\partial\boldsymbol{x}_c(t_i)}=0 \tag{50}$$

Here on assumes that the Lagrange multipliers $\boldsymbol{\kappa}_i$, ν_i have been assigned already. Thus one has the following shooting variables now, i.e., $\boldsymbol{\lambda}(t_0)$, and $\Delta\boldsymbol{v}_i$, $\boldsymbol{\kappa}_i$, ν_i for each GA, which are not complete, however. Each GA time t_i must also be included here to make the problem well-defined. Finally one obtains $7+9N$ shooting variables if N GA are employed in the rendezvous process. And the corresponding shooting equations $\boldsymbol{\Phi}(\boldsymbol{q})=\boldsymbol{0}$ are the above listed constraints at t_0, t_f, t_i, $i=1\cdots N$, i.e., (43)–(47). Finding roots of this high dimensional shooting equations $\boldsymbol{\Phi}(\boldsymbol{q})=\boldsymbol{0}$, which is definitely nontrivial, will be treated in detail in the next section.

Note in the above, partial differentiations of Ψ_i with respect to $\boldsymbol{x}_d(t_i^-)$, $\boldsymbol{x}_d(t_i^+)$, $\boldsymbol{x}_c(t_i)$ and t_i have already been derived in the last part, i.e., (34) (39)–(42).

4 Numerical techniques

4.1 Normalization, GA time & random searching

The sensibility to initial guess in solving the above equations $\boldsymbol{\Phi}(\boldsymbol{q})=\boldsymbol{0}$ by single shooting method, due to the local characteristics of algorithm, urges one to try to understand the underlying physics of the optimal rendezvous problem as much as possible, in light of which one can set a reasonable admissible domain of shooting variables, rather than in darkness. For example, to solve shooting equation $\boldsymbol{\Phi}(\boldsymbol{q})=\boldsymbol{0}$ initial costates are usually allowed to vary in a very broad domain. This admissible domain is far from reasonable in our sense. Actually, the admissible domain should shrink to a bounded unit sphere, a much smaller, hence more reasonable searching domain. Note that a reasonable admissible domain is of vital importance for obtaining reasonable initial interaction value of shooting variables by searching algorithms, even global one like PSO or random searching globally. In other words, to reduce the sensibility to interaction initial guess, one should first set a reasonable admissible domain for shooting variables and then employ the global searching algorithm.

Jiang [14] noted the homogeneousness of state equations (3), costate equations (12) and various constraints or boundary conditions with respect to the hybrid multipliers, which include positive factor λ_0, initial costate $\boldsymbol{\lambda}(t_0)$ and constant Lagrange multipliers $\boldsymbol{\kappa}$, ν associated with equality and inequality constraints introduced by GA. And then he proposed a normalization technique to confine the hybrid multipliers in the bounded unit sphere. Here, what 'normalization technique' means is dividing a vector by its magnitude. The basic idea underlying normalization technique is that it is the normalized costates that really matter. If the performance index is multiplied by a positive factor λ_0, the ratio (term by term) of the new costates to the old ones will be λ_0 automatically. However, optimal control and optimal trajectory will definitely remain the same. This insight is very important in the sense that it will simplify the initialization of the shooting variables $\boldsymbol{\lambda}(t_0)$, $\boldsymbol{\kappa}$, $\boldsymbol{\nu}$ through parameterization of high dimensional abstract unit sphere. Here this important technique is adopted and extended to multiple GA case. That is, after normalization, multipliers should satisfy

$$\sqrt{\lambda_0^2+\boldsymbol{\lambda}\cdot\boldsymbol{\lambda}+\sum_{i=1}^{N}\left(\boldsymbol{\kappa}_i\cdot\boldsymbol{\kappa}_i+\nu_i^2\right)}=1\,. \tag{51}$$

where $\boldsymbol{\kappa}_i$, ν_i are constant Lagrange multipliers associated with equality and inequality constraints introduced by each GA. These parameters defining the above unit sphere (51) will be the actual shooting variables, which fall all in the bounded domain. The parameterization of high dimensional abstract sphere itself is a very interesting mathematical trick and the details can be found in refs. [14,27].

The shooting variables $\Delta\boldsymbol{v}_i$, velocity increment owing to each GA, will be assigned within unit ball

$$\left\|\frac{\Delta\boldsymbol{v}_i}{\sqrt{\mu_p/r_{\min}}}\right\|\leqslant 1\,, \tag{52}$$

where $\sqrt{\mu_p/r_{\min}}$ is the possible maximum magnitude of velocity increment $\Delta\boldsymbol{v}_i$ for all the admissible GA altitudes satisfying $r_{\text{periapsis}}\geqslant r_{\min}$ [26]. It is the convenience of parameterizing the above unit ball, that $\Delta\boldsymbol{v}_i$ rather than $\boldsymbol{v}(t_i^+)$, is finally chosen as shooting variables in the present paper. One has to admit, however, that the above admissible domain is larger than the actual one by noting that various main is larger than the actual one by noting that various constraints on velocity before and after GA, i.e., eqs. (18)–(20) are relaxed at this moment. It is one of the advantages of indirect method in trajectory optimization, i.e., the formulation of shooting equations, i.e., $\boldsymbol{\Phi}(\boldsymbol{q})=\mathbf{0}$, guarantees the optimality of the original problem and the satisfaction of various constraints at the same time, that one can assume here that the relaxation of some constraints when setting admissible domain will not affect the optimality, but bring much convenience. The situation will be definitely not the same when it comes to parameter optimization problems, like direct methods in trajectory optimizations, where relaxation of constraints totally changes the original optimization problem, i.e., in respect of both optimality and constraints.

To get a preliminary, while reasonable, estimate of the admissible domain of GA time t_i, however, is more involved. At most time, one has to rely on experience and corresponding results in an impulsive case. The basic idea here is to formulate the problem as a parameter optimization problem based on the results of classical Lambert problem.

For outer planet GA, like Mars and Jupiter, it is straightforward to formulate the optimization problem. Let launch time, rendezvous time and successive GA time be t_0, t_f, t_i, $i=1\cdots N$, respectively and then the positions of spacecraft at the above time are also specified. Solving a series of Lambert problems can yield proper velocities at corresponding positions, i.e., $\tilde{\boldsymbol{v}}(t_0)$, $\tilde{\boldsymbol{v}}(t_f)$, $\tilde{\boldsymbol{v}}(t_i^-)$, $\tilde{\boldsymbol{v}}(t_i^+)$, $i=1\cdots N$. Now taking the total velocity increment as performance index and each GA time t_i, $i=1\cdots N$ as optimization parameters, one finally formulates the optimal impulsive rendezvous problem as

$$\min_{\{t_i\}}\left(\left\|\tilde{\boldsymbol{v}}(t_0)-\boldsymbol{v}_0\right\|+\left\|\tilde{\boldsymbol{v}}(t_f)-\boldsymbol{v}_f\right\|+\sum_{i=1}^{N}\left\|\tilde{\boldsymbol{v}}(t_i^-)+\Delta\boldsymbol{v}_i-\tilde{\boldsymbol{v}}(t_i^+)\right\|\right). \tag{53}$$

Note here in the above three terms, $\sum_{i=1}^{N}\left\|\tilde{\boldsymbol{v}}(t_i^-)+\Delta\boldsymbol{v}_i-\tilde{\boldsymbol{v}}(t_i^+)\right\|$ is the velocity increment or maneuver impulse at patching positions apart from optimal velocity impulse supplied by GA, i.e., $\Delta\boldsymbol{v}_i$, which can be determined uniquely with income velocity $\tilde{\boldsymbol{v}}(t_i^-)$ and outgoing velocity $\tilde{\boldsymbol{v}}(t_i^+)$ specified [14].

For Earth GA, i.e., spacecraft launched from Earth coming back to encounter Earth again, the situation here is a little different, where the problem can not be formulated as a Lambert problem directly. The usual approach to Earth GA is first to launch the spacecraft into a heliocentric orbit, whose period is slightly larger than an integer number of years and perihelion is the radius of Earth's orbit, i.e., 1 AU, and then to apply a retrograde velocity maneuver at aphelion to lower the perihelion to intercept Earth [26]. It's clear that launch velocity $\boldsymbol{v}_\infty(t_0)=\boldsymbol{v}(t_0)-\boldsymbol{v}_E(t_0)$ in this case should never be zero, which should also be treated as an optimization parameter. With launch velocity specified, where and when to apply retrograde velocity maneuver can then be determined, say $\boldsymbol{r}(t_r)$, $\boldsymbol{v}(t_r)$, $t_r=t_0+T_{\text{out}}/2$, here T_{out} denoting the outgoing orbit's period. The present problem can now be transformed into the same formulation as that of outer planet GA, if $\boldsymbol{r}(t_r)$ is treated as initial position. All the followings will be parallel to those for outer planet GA. Warnings should be given that all the above about GA time are just a preliminary estimate or guess. Both experience and preliminary guess should be taken into account on setting an admissible domain.

With the reasonable admissible domain for shooting variables, PSO is then adopted in the paper to search globally the preliminary initial interaction value for solving energy optimal problem, whose object function is performance index plus residual of shooting equations $\boldsymbol{\Phi}(\boldsymbol{q})=\mathbf{0}$. However, PSO is found less and less effective as dimension of shooting variables increases, when the multiple GA case is treated for example. Instead of PSO, multiple random searching globally is employed to treat the high dimension case. It is found more efficient than PSO and this will become clear in the followed numerical examples.

4.2 Homotopic technique

The aforementioned sensibility to initial guess in solving the above equations $\boldsymbol{\Phi}(\boldsymbol{q})=\mathbf{0}$ by the single shooting method is just one of the various difficulties encountered in fuel-optimal low-thrust trajectory optimization, which is actually the intrinsic difficulty in solving high dimensional nonlinear equations, a typical topic in applied mathematics.

Another difficulty, however, is more daunting, which is totally rooted in the different physics underlying various performance indexes. Recalling aforementioned fuel optimal performance index

$$J=\lambda_0\frac{T_{\max}}{I_{sp}g_0m_0}\int_{t_0}^{t_f}u\,\mathrm{d}t, \tag{54}$$

one should note here the integrand $u=c\|\boldsymbol{T}\|$ is proportional to thrust vector's norm, the smoothness of which is not satisfying. And even worse, the bang-bang control (11), due to the linearity of state equations (3), hence Hamilton function H (9), with respect to the just mentioned normalized magnitude of thrust vector, i.e., u, is discontinuous. These will be bound to reduce the smoothness of shooting functions $\boldsymbol{\Phi}(\boldsymbol{q})=\mathbf{0}$ and thus the convergence radius of associated gradient algorithm badly.

This difficulty, i.e., small and irregular convergence domain mainly due to the loss of smoothness of shooting function $\Phi(\boldsymbol{q})$, can be removed to some degree, if the so-called homotopic technique, one kind of numerical continuation method, is explored properly [14–18]. Actually it is the linearity of the integrand in performance index and of state equations with respect to normalized thrust magnitude u that lead to discontinuous bang-bang control, which finally reduces the smoothness of shooting functions $\Phi(\boldsymbol{q})$. This observation motivates one to find smoother shooting function $\tilde{\Phi}(\boldsymbol{q})$, from some new performance index $\tilde{J}$, in which the integrand is not linear with respect to u, say quadratic form u^2, here

$$\tilde{J} = \lambda_0 \frac{T_{\max}}{I_{sp} g_0 m_0} \int_{t_0}^{t_f} u^2 \mathrm{d}t. \tag{55}$$

It's easy to derive the corresponding optimal control associated with the above performance index by PMP. Quadratic form u^2 in $\tilde{J}$ leads to quadratic Hamilton function H with respect to u and one can finally get a continuous optimal control, making shooting function $\tilde{\Phi}(\boldsymbol{q})$ smoother than $\Phi(\boldsymbol{q})$. It seems at first that the new continuous optimal control, leading to smoother shooting function, has nothing to do with the former discontinuous one. It is homotopic technique that saves one from this embarrassing situation, suggesting that one should find some continuous mapping $F(\varepsilon,\boldsymbol{q}) = \Phi_\varepsilon(\boldsymbol{q})$, called homotopic mapping, which gradually transitions from the smoother shooting function to the less smooth one by varying the so-called homotopic parameter ε continuously from 0 to 1, specifically in our case here, from energy optimal problem $\tilde{\Phi}(\boldsymbol{q}) = 0$ to fuel optimal one $\Phi(\boldsymbol{q}) = 0$. In other words, the homotopic mapping $F(\varepsilon,\boldsymbol{q})$ satisfies $F(0,\boldsymbol{q}) = \tilde{\Phi}(\boldsymbol{q})$, $F(1,\boldsymbol{q}) = \Phi(\boldsymbol{q})$. The underlying key idea of the above remarks is that the solution sequence $\boldsymbol{q}(\varepsilon_n)$ of $\Phi_{\varepsilon_n}(\boldsymbol{q}) = 0$ will converge finally to solution $\boldsymbol{q}(1)$ of $\Phi(\boldsymbol{q}) = 0$ starting by the solution $\boldsymbol{q}(0)$ of energy optimal problem $\tilde{\Phi}(\boldsymbol{q}) = 0$, if sequence ε_n converges to 1. The details will be neglected here and any mathematically oriented readers can consult ref. [15] and related references there in.

Noting that shooting function $\Phi(\boldsymbol{q})$ means actually all the listed various constraints (43)–(47), which is definitely not explicitly presented, nor could be in fact, one realizes the above mentioned homotopic mapping $F(\varepsilon,\boldsymbol{q}) = \Phi_\varepsilon(\boldsymbol{q})$ is difficult to be constructed. Actually the one explicitly constructed here is homotopic mapping $f(\varepsilon)$ between the two integrands u^2, u of the corresponding performance indexes $\tilde{J}$, J, and the desired one $F(\varepsilon,\boldsymbol{q}) = \Phi_\varepsilon(\boldsymbol{q})$ between shooting functions will be induced from $f(\varepsilon)$. More explicitly, choosing homotopic mapping $f(\varepsilon) = \varepsilon u + (1-\varepsilon)u^2$, one then obtains a series of optimal guidance problems

$$J_\varepsilon = \lambda_0 \frac{T_{\max}}{I_{sp} g_0 m_0} \int_{t_0}^{t_f} \left(\varepsilon u + (1-\varepsilon) u^2\right) \mathrm{d}t. \tag{56}$$

Now applying PMP to every optimal control problem associated with J_ε, one can obtain the corresponding shooting functions $\Phi_\varepsilon(\boldsymbol{q}) = 0$, which actually define the desired homotopic mapping $F(\varepsilon,\boldsymbol{q}) = \Phi_\varepsilon(\boldsymbol{q})$.

With homotopic technique in hand, one can expect and finally obtain the fuel optimal solution starting with the energy optimal problem. The difficulty in obtaining the energy optimal solution is handled with techniques introduced in sect. 4.1.

5 Numerical results

For comparison, two rendezvous missions, which were designed by program GALLOP in ref. [3], are treated here with the present approach.

5.1 Earth to Jupiter rendezvous mission via Earth GA

To illustrate rendezvous via Earth GA, which is different from other outer planet GA as claimed in sect. 4.1, we present a rendezvous mission from Earth to Jupiter via Earth gravity assist. The parameters are listed in Table 1.

The corresponding shooting variables here are $\boldsymbol{q} = \left[\boldsymbol{\lambda}(t_0), \boldsymbol{\kappa}_1, \nu_1, \Delta\boldsymbol{v}_1, t_1\right]$ and shooting equations include (43)–(46). The admissible domain of GA time t_1 is set to be $t_1 \in [0.72, 1.72]$ and this domain is actually very broad, also very intuitive, if close inspection is paid to sect. 4.1 about Earth GA. With PSO searching globally, the preliminary initial interaction value is obtained. And then the

Table 1 Parameters in the E-E-J rendezvous mission

Parameters	Value	Units
Launch date on Earth	Sep. 29, 2015	–
Launch v_∞ on Earth	0.71	km/s
Arrival date on Jupiter	Apr. 5, 2020	–
Total TOF	4.52	year
Initial mass	19820	kg
Earth GA $r_{\min}$	6378.145+500	km

smooth energy-optimal problem is solved by the single shooting method. It takes about 0.75 hours for this process in the present mission example.

The solutions to the shooting equation associated with the fuel-optimal problem are finally obtained through homotopic mapping starting with the energy-optimal results. This homotopic process is amazingly quick, say less than 10 seconds in the present example. In Table 2 results here are compared with those of GALLOP [3]. It seems at first that the approach here is almost the same as GALLOP in the sense of optimality. However, one should note that the present approach is much more efficient than GALLOP, which usually relies on parallel computations due to a huge number of optimization parameters. And the difference between the results here and GALLOP's is so slight that it is probably due to numerical error, only about 1.3‰ of fuel consumptions.

Recall that bang-bang control (11) is derived assuming that the problem will not be a singular case. It's clear now from Figure 1, that switching function S reaches zeros only at some discrete points, not on the continuous domain and this excludes the singularity case from the present mission example, thus finally confirming the earlier assumptions.

Figure 2 illustrates the optimal thrust, in the form of bang-bang control, with thrust in two phases, i.e., Earth to Earth and Earth to Jupiter depicted by solid lines and dashed ones respectively. And Figure 3 shows the optimal low thrust trajectory in the E-E-J rendezvous mission, where solid lines (red) denote burn arcs, dotted ones (black) denote coast arcs, and the remaining dashdotted lines are Earth's and Jupiter's orbits respectively.

Table 2 Comparison with GALLOP for the E-E-J rendezvous mission

	Our results	GALLOP
Earth GA date	Dec. 14, 2016	Dec. 13, 2016
Earth GA altitude (km)	500	500
Earth GA v_∞ (km/s)	7.55	7.49
Final Mass (kg)	16176	16181

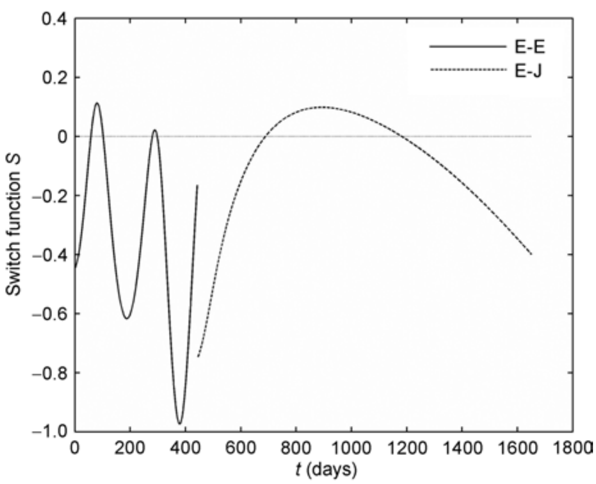

Figure 1 Switching function S in the E-E-J rendezvous mission.

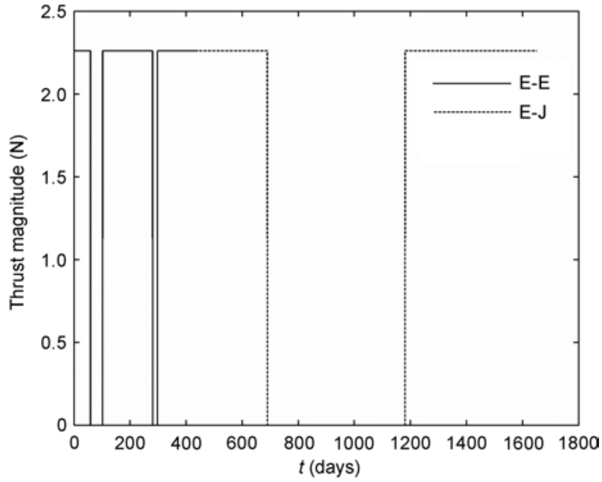

Figure 2 Optimal thrust profile in the E-E-J rendezvous mission.

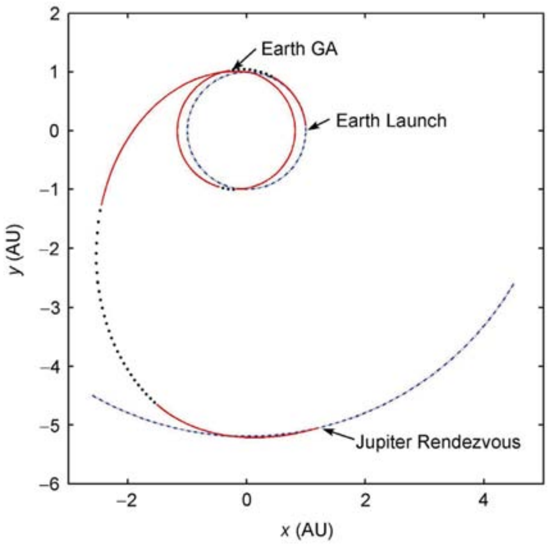

Figure 3 Trajectory plot of the E-E-J rendezvous mission.

5.2 Earth to Uranus rendezvous mission via Earth GA & Jupiter GA

To illustrate the present approach applied to multiple GA case, we present a rendezvous mission from Earth to Uranus via Earth GA and Jupiter GA successively. The parameters are listed in Table 3.

Like the former mission example, energy optimal results should be obtained first before starting homotopic programs. Note that here the shooting variables have increased by 9 due to one more GA, that is to say $\boldsymbol{q}=\left[\boldsymbol{\lambda}\left(t_0\right),\boldsymbol{\kappa}_1,\nu_1,\Delta\boldsymbol{v}_1,t_1,\boldsymbol{\kappa}_2,\nu_2,\Delta\boldsymbol{v}_2,t_2\right]$, where the latter 9×2 shooting variables are

Table 3 Parameters in the E-E-J-U rendezvous mission

Parameters	Value	Units
Launch date on Earth	Feb. 11, 2020	–
Launch v_∞ on Earth	0.94	km/s
Arrival date on Uranus	June 3, 2029	–
Total TOF	9.31	year
Initial mass	19682	kg
Earth GA $r_{\min}$	6378.145+500	km
Jupiter GA $r_{\min}$	71492	km

associated with the two GA, i.e., Earth GA and Jupiter GA. The admissible domain of Earth GA time t_1 and Jupiter GA time t_2 is set to be $t_1 \in [0.5, 1.5]$ and $t_2 \in [2.5, 3.0]$, based on both the aforementioned preliminary estimate in the impulsive case and experience. Numerical computation confirms the earlier worry that PSO becomes less effective as the number of optimization parameters increases. Thus here the random searching globally is adopted to find the solution to the energy optimal problem. The situation becomes tougher and it takes longer time to obtain the results, nearly 1.0 hour in the present mission example. Once energy optimal results obtained, like the former mission example, the following homotopic process is still very fast, say less than 5.0 minutes. And fuel optimal results are finally obtained as listed in the Table 4, where results provided by GALLOP [3] are also listed for comparison. It's clear from Table 4 that the results obtained by the present approach is better than GALLOP's, with 158 kg more final fuel. This also confirms the usual assertion that the indirect method is better than the direct method in the sense of optimality, though many numerical difficulties in the former one are formidable. The optimal rendezvous process is illustrated in the following.

In Figure 4, switching function S is depicted to confirm the earlier assumption that it will only reach zeros at certain discrete points, not on the continuous domain.

In Figure 5, optimal thrust in three phases, i.e., Earth to Earth, Earth to Jupiter and Jupiter to Uranus, are depicted by solid lines, dashed ones and dashdotted ones respectively. The optimal thrust is in bang-bang form in each phase. And the whole optimal rendezvous process is plotted in Figure 6,

Table 4 Comparison with GALLOP for the E-E-J-U rendezvous mission

	Our results	GALLOP
Earth GA date	May 4, 2021	May 1, 2021
Earth GA attitude (km)	500	500
Earth GA v_∞ (km/s)	8.65	8.25
Jupiter GA date	Nov. 22, 2022	Dec. 3, 2022
Jupiter GA attitude $(\times R_J)$	17.86	16.7
Jupiter GA v_∞ (km/s)	11.72	11.8
Final mass (kg)	13187	13029

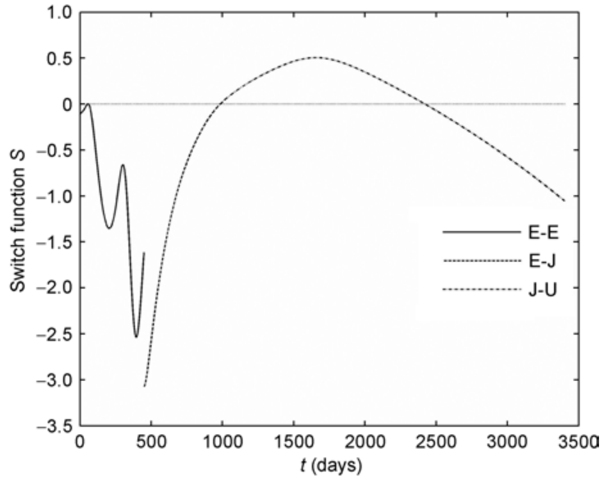

Figure 4 Switching function S in the E-E-J-U rendezvous mission.

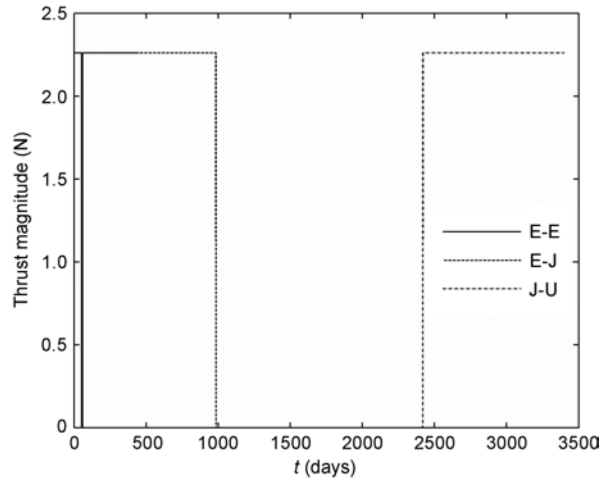

Figure 5 Optimal thrust profile in the E-E-J-U rendezvous mission.

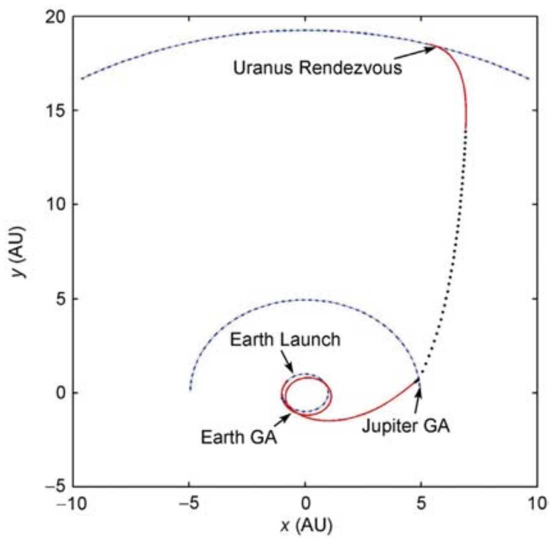

Figure 6 Trajectory plot of the E-E-J-U rendezvous mission.

where solid lines (red) denote burn arcs, dotted ones (black) denote coast arcs, and the orbits of Earth, Jupiter, and Uranus are all denoted by dashdotteded lines.

The warning here is that the aforementioned time taken by the preliminary searching and solving energy optimal problem, i.e., nearly 0.75 hours in mission1 and 1.0 hours in mission 2, is presented here mainly for showing the feasibility of the present approach. Actually a more complete, hence more reasonable, measure of efficiency should also take into account the probability distribution of obtaining desired results [14,15]. The statistical characteristics of the present approach stem from random algorithms like PSO or random searching directly employed in the preliminary searching phase.

All the above results in both mission 1 and mission 2 are obtained from numerical computations based on programs coded by C++ and executed in Microsoft Visual C++ 6.0, which operates on personal computer with the hardware of CPU 2.0 GHz and memory 2.0 GB.

6 Conclusions

The present paper extends previous work [14] to multiple GA case and thus finishes the complete treatment of the problem, i.e., fuel optimal low thrust rendezvous via GA, both on modeling and numerical techniques. The optimal rendezvous via multiple GA is formulated as an optimal control problem with multiple interior constraints, and then leads to a MPBVP. To cope with various numerical difficulties in solving the above MPBVP, we use the homotopic technique combined with random searching globally and PSO. Finally, the problem is resolved successfully and the numerical results in the end show good performance of the present approach, both on global optimality and computation costs in comparison with those provided by GALLOP. Two difficulties, however, are also found on the way. As one more GA adds another 9 shooting variables to the previous ones, the dimension of shooting equations increases fast if multiple GA is employed. Searching preliminary initial shooting variables of the high-dimensional nonlinear equations becomes more and more troublesome. The other difficulty is about estimating GA time in the multiple GA case. A preliminary estimate of GA time here is obtained from both experience and corresponding impulsive results while the underlying reasoning is actually not so clear. More links between low thrust and impulsive case are demanded. Further investigations into these issues are needed.

This work was supported by the National Natural Science Foundation of China (Grant Nos. 10832004 and 11072122). The authors sincerely thank Dr. GONG ShengPing for his helpful discussions and comments.

1 Rayman M D, Varghese P, Lehman D H, et al. Results from the Deep Space 1 technology validation mission. Acta Astron, 2000, 9: 475–487
2 Kuninaka H, Nishivama K, Funakai I, et al. Asteroid rendezvous of HAYABUSA explorer using microwave discharge ion engines. In: the 29th International Electric Propulsion Conference. Princeton: Princeton University, 2005
3 Yan C H, McConaghy T T, Chen K J, et al. Preliminary design of nuclear electric propulsion missions to the outer planets. In: AIAA/AAS Astrodynamics Specialist Conference and Exhibit, 16-19, August 2004. Providence, Rhode Island, 2004
4 Hargraves C, Paris S. Direct trajectory optimization using nonlinear programming and collocation. J Guid Control Dyn, 1987, 4: 338–342
5 Enright P J, Conway B A. Discrete approximations to optimal trajectories using direct transcription and nonlinear programming. J Guid Control Dyn, 1992, 4: 994–1002
6 Kechichian J A. Optimal low-earth-orbit-geostationary-earth-orbit intermediate acceleration orbit transfer. J Guid Control Dyn, 1997, 4: 803–811
7 Sang-Young P, Kyu-Hong C. Optimal low-thrust intercept/rendezvous trajectories to earth-crossing objects. J Guid Control Dyn, 2005, 5: 1049–1055
8 Ranieri C L, Ocampo C A. Indirect optimization of spiral trajectories. J Guid Control Dyn, 2006, 6: 1360–1366
9 Ranieri C L, Ocampo C A. Indirect optimization of three-dimensional finite-burning interplanetary transfers including spiral dynamics. J Guid Control Dyn, 2009, 2: 444–454
10 Mantia M L, Casalino L. Indirect optimization of low-thrust capture trajectories. J Guid Control Dyn, 2006, 4: 1011–1014
11 Jr Arthur E B, Yu-Chi H. Applied Optimal Control Theory. Revised edition. New York: Taylor & Francis, 1975
12 Pontryagin L S. Mathematical Theory in Optimal Control Process. Boca Raton: CRC Press, 1987
13 Mcintryre J E. The Pontryagin Maximum Principle, Guidance, Flight Mechanics and Trajectory Optimization, v.2. NASA CR-1006, 1968
14 Jiang F H, Li J F, BaoYin. Practical techniques for low-thrust trajectory optimization with homotopic approach. J Guid Control Dyn, 2010, submitted
15 Bertrand R, Epenoy R. New smoothing techniques for solving bang-bang optimal control problems-numerical results and statistical interpretation. Opt Control Appl Methods, 2002, 4: 171–197
16 Haberkorn T, Martinon P, Gergaud J. Low-thrust minimum-fuel orbital transfer: A homotopic approach. J Guid Control Dyn, 2004, 6: 1046–1060
17 Mischa K. Continuous low-thrust trajectory optimization: Techniques and applications. Doctoral Degree Dissertation. Virginia: Virginia Polytechnic Institute and State University, 2005
18 Allgower E L, Georg L. Introduction to Numerical Continuation Methods. Philadelphia: Society for Industrial Mathematics, 1987
19 McConaghy T T, Debban T J, Petropoulos A E, et al. Design and optimization of low-thrust trajectories with gravity assists. J Spacecr Rockets, 2003, 3: 380–387
20 Grigoriev I S, Zapletin M P. Constructing pontryagin extremals for the optimal control problem of asteroid fly-by. Autom Remote Control, 2009, 9: 1499–1513
21 Casalino L, Colasurdo G, Pastrone D. Optimal low-thrust escape trajectories using gravity assist. J Guid Control Dyn, 1999, 5: 637–642
22 Pascale P D, Vasile M. Preliminary design of low-thrust multiple gravity-assist trajectories. J Spacecr Rockets, 2006, 5: 1065–1076

23 Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of IEEE International Conference on Neural Networks, Perth, WA, Australia. Piscataway: IEEE Service Center, 1995. 4: 1942–1948
24 Das S, Abraham A, Konar A. Particle swarm optimization and differential evolution algorithms: Technical analysis, applications and hybridization perspectives. In: Studies in Computational Intelligence. Liu Y, ed. Berlin: Springer Verlag, 2008. 116: 1–38
25 Lawden D F. Optimal Trajectories for Space Navigation. London: Butterworths, 1963
26 Sims J A. Delta-V gravity-assist trajectory design: Theory and practice. Doctoral Degree Dissertation. West Lafayette: School of Aeronautics and Astronautics, Purdue University, 1997
27 Zorich V A, Cooke R. Mathematical Analysis v.2. Berlin: Springer, 2004