SYMMETRIES IN THE OPTIMAL CONTROL OF SOLAR SAIL SPACECRAFT

M. KIM and C. D. HALL
Department of Aerospace and Ocean Engineering
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061, USA
e–mails: mischa@vt.edu, cdhall@vt.edu

Abstract. The theory of optimal control is applied to obtain minimum–time trajectories for solar sail spacecraft for interplanetary missions. We consider the gravitational and solar radiation forces due to the Sun. The spacecraft is modelled as a flat sail of mass m and surface area A and is treated dynamically as a point mass. Coplanar circular orbits are assumed for the planets. We obtain optimal trajectories for several interrelated problem families and develop symmetry properties that can be used to simplify the solution–finding process. For the minimum–time planet rendezvous problem we identify different solution branches resulting in multiple solutions to the associated boundary value problem. We solve the optimal control problem via an indirect method using an efficient cascaded computational scheme. The global optimizer uses a technique called Adaptive Simulated Annealing. Newton and Quasi–Newton Methods perform the terminal fine tuning of the optimization parameters.

Key words: Solar sail, Optimal control, Interplanetary missions, Symmetry.

1. Introduction

The concept of using solar radiation pressure as a means of propulsion for space vehicles was first introduced in the 1920s by Tsiolkovsky (1921) and Tsander (1967 (quoting a 1924 report by the author)). About fifty years later, the effect of solar radiation on spacecraft attitude dynamics was first experienced with the Mariner 10 mission to Mercury and Venus. Mariner 10 was also the first spacecraft to use a gravity assist trajectory, accelerating as it entered the gravitational influence of Venus, then using the planet's gravity field to move onto a slightly different course to reach Mercury. Since then, there have been several attempts to realize a solar sail mission. In his book Wright (1992) presents a detailed analysis on some possible solar sail applications. During his time at the Jet Propulsion Laboratory (JPL) Wright was actively involved in the planning of a rendezvous mission to comet Halley using solar sail technology. In 1977 a solar electric propulsion concept was selected instead, primarily because of technology maturity. Not long thereafter, the Halley rendezvous mission was dropped by NASA.

During the development of the solar sail spaceflight concept, numerous studies of the

associated dynamics problem were presented. The first solar sail trajectories were calculated by Tsu (1959) and London (1960). Tsu investigated various means of propulsion and showed that in many cases solar sails show superior performance when compared to chemical and ion propulsion systems. The author used approximated heliocentric motion equations to obtain spiraling trajectories for a "*... fixed sail setting*". London presented similar spiral solutions for Earth–Mars transfers with constant sail orientation using the exact equations of motion. Optimal solar sail trajectories were first computed by Zhukov and Lebedev (1964) for interplanetary missions between coplanar circular orbits. In 1980 Jayaraman (1980) published similar minimum–time trajectories for transfers between the Earth and Mars. Two years later, Wood et al. (1982) presented an analytical proof to show that the orbital transfer times obtained by Jayaraman (1980) were incorrect due to the incorrect application of a transversality condition of variational calculus and an erroneous control law. Powers et al. (1999) and Powers and Coverstone (2001) obtained results similar to those reported in Wood's paper, and obtained solutions for transfers to synchronous orbits. The more general time–optimal control problem of three–dimensional, inclined and elliptic departure and rendezvous planet orbits was discussed by Sauer (1976). Hughes and McInnes (2001) used Genetic Algorithms and Sequential Quadratic Programming to obtain interplanetary trajectories via a direct method. In a recent paper Dachwald (2004) presented a novel approach based on Evolutionary Neurocontrollers (ENC) to calculate optimal solar sail trajectories for interplanetary missions.

The paper is organized as follows: In the following sections we introduce the system model and the corresponding motion equations. Subsequently, we define the minimum–time optimal control problem and present various optimal control problem families. We discuss system symmetries and show how these symmetries can be used to great advantage to find optimal control solutions. We outline the numerical algorithm with focus on the global optimization algorithm, namely, Simulated Annealing (SA). Finally we present simulation results for each of the optimal control problem families and discuss the relationship between the various families.

2. System Model And Motion Equations

The system model is illustrated in Figure 1. The spacecraft is modelled as a perfectly flat solar sail of mass m and surface area A and is treated dynamically as a point mass. We define the body–fixed reference frame $\{\boldsymbol{b}_x, \boldsymbol{b}_y\}$ with the $\boldsymbol{b}_x$ axis identifying the solar sail surface normal vector. The sail orientation angle α is defined as the angle between the sail normal ($= \boldsymbol{b}_x$) and the solar flux direction $\boldsymbol{S}$. A positive angle rotates $\boldsymbol{b}_x$ anti–clockwise into $\boldsymbol{S}$. For convenience, we define $\boldsymbol{S}_\perp$ as the unit vector orthogonal to $\boldsymbol{S}$ such that $\boldsymbol{S}_\perp$ and $\boldsymbol{b}_y$ are aligned for $\alpha = 0$. The environmental forces acting on the spacecraft system are due to the gravitational field and the radiation pressure of the Sun. The initial and target spacecraft trajectories are modelled as heliocentric, circular, and coplanar orbits.

We define the generalized coordinate vector as $\boldsymbol{r} \triangleq (r, \theta)^T$ and the corresponding velocity vector as $\boldsymbol{v} \triangleq (v_r, v_\theta)^T$, where $v_r = \dot{r}$ and $v_\theta = r\dot{\theta}$. In the inertial reference frame the motion

equations for a perfectly reflective solar sail spacecraft are then (McInnes, 1999)

$$\begin{aligned} \dot{r} &= v_r && (1)\\ \dot{\theta} &= v_\theta / r && (2)\\ \dot{v}_r &= \beta \cos^3\alpha / r^2 + v_\theta^2 / r - 1/r^2 && (3)\\ \dot{v}_\theta &= \beta \sin\alpha \cos^2\alpha / r^2 - v_r v_\theta / r && (4) \end{aligned}$$

where β is the nondimensional characteristic acceleration. The motion equations can be rewritten in condensed form as $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, u)$, where the state vector and control input are defined as $\boldsymbol{x} = (r, \theta, v_r, v_\theta)^T$ and $u = \alpha$. Typical normalized boundary conditions for the system (1–4) are

$$r(t_0) = r_0 \qquad \theta(t_0) = \text{free} \qquad v_r(t_0) = 0 \qquad v_\theta(t_0) = 1/\sqrt{r_0} \qquad (5)$$

$$r(t_f) = r_f \qquad \theta(t_f) = \theta_f \text{ or free} \qquad v_r(t_f) = 0 \qquad v_\theta(t_f) = 1/\sqrt{r_f} \qquad (6)$$

Note that depending on the particular optimal control problem family the angular boundary condition has to be chosen appropriately. For the minimum–time orbit transfer problem, $\theta(t_f) = \text{free}$, whereas for the minimum–time planet rendezvous problem $\theta(t_f) = \theta_f$. Therefore, we choose as the end–point constraint for the general minimum–time optimal control problem

$$\boldsymbol{\psi}\left(\boldsymbol{x}(t_f), t_f\right) = \left(r(t_f) - r_f,\ \theta(t_f) - \theta_f,\ v_r(t_f),\ v_\theta(t_f) - 1/\sqrt{r_f}\right)^T = \boldsymbol{0} \qquad (7)$$

In the next section we formulate the optimal control problem.

3. Optimal Control Problem Formulation

The optimal control problem is to find an optimal control input $\boldsymbol{u}^\star$ for a generally nonlinear system $\dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)$ such that the associated performance index

$$\mathcal{J} = \phi\left(\boldsymbol{x}(t_f), t_f\right) + \int_{t_0}^{t_f} \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, t)\, dt \qquad (8)$$

is minimized, and such that the constraint at final time t_f

$$\boldsymbol{\psi}\left(\boldsymbol{x}(t_f), t_f\right) = \boldsymbol{0} \qquad (9)$$

is satisfied. In equations (8,9) $\boldsymbol{x}$ is the n–dimensional state vector, $\boldsymbol{u}$ is the m–dimensional control input, and ϕ and $\mathcal{L}$ are the terminal and accumulated costs, respectively. Instead of solving a constrained optimization problem, it is usually advantageous to consider the corresponding unconstrained optimization problem using the augmented performance index

$$\mathcal{J}^+ = \phi\left(\boldsymbol{x}(t_f), t_f\right) + \boldsymbol{\nu}^T \boldsymbol{\psi}\left(\boldsymbol{x}(t_f), t_f\right) + \int_{t_0}^{t_f} \left\{\mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, t) + \boldsymbol{\lambda}^T \left(\boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t) - \dot{\boldsymbol{x}}\right)\right\} dt \qquad (10)$$

Defining the *Hamiltonian function* $\mathcal{H}$ as

$$\mathcal{H} = \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, t) + \boldsymbol{\lambda}^T \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t) \qquad (11)$$

the state and costate equations are obtained as

$$\dot{\boldsymbol{x}} = \frac{\partial \mathcal{H}}{\partial \boldsymbol{\lambda}} = \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t) \quad \text{and} \quad \dot{\boldsymbol{\lambda}} = -\frac{\partial \mathcal{H}}{\partial \boldsymbol{x}} = -\frac{\partial \mathcal{L}(\boldsymbol{x}, \boldsymbol{u}, t)}{\partial \boldsymbol{x}} - \frac{\partial \boldsymbol{f}(\boldsymbol{x}, \boldsymbol{u}, t)}{\partial \boldsymbol{x}} \boldsymbol{\lambda} \tag{12}$$

Equations (12), combined with a set of appropriate boundary conditions for the states and costates, yield a well–defined Boundary Value Problem (BVP). By specifying sets of end– and (where necessary) interior–point constraints, various control problem families can be identified:

$\mathcal{P}_0$. Minimum–time transfer between an initial and a target orbit. There is no angular end–point constraint; that is $\theta(t_f) = \text{free}$.

$\mathcal{P}_1$. $\mathcal{P}_0$ transfer with subsequent $\mathcal{P}_0$ return transfer to initial orbit. There are no angular interior– and end–point constraints; that is, $\theta(t_i) = \text{free}$ and $\theta(t_f) = \text{free}$.

$\mathcal{P}_2$. Minimum–time rendezvous with a target planet. The angular end–point constraint depends on the initial angular separation between initial and target planet, and the synodic period of the target planet; that is $\theta(t_f) = \theta_f$.

$\mathcal{P}_3$. Minimum–time transfer including a target planet rendezvous and subsequent return to initial planet. The angular constraints are $\theta(t_i) = \theta_i$ and $\theta(t_f) = \theta_f$.

In the following we present the optimality conditions and discuss specifics of each of the control problem families introduced above.

3.1. Optimality conditions

For the general minimum–time transfer problem using the Lagrange formulation, the Hamiltonian is given by

$$\mathcal{H} = \lambda_1 v_r + \lambda_2 v_\theta / r + \lambda_3 \left(\beta \cos^3 \alpha / r^2 + v_\theta^2 / r - 1/r^2 \right) + \lambda_4 \left(\beta \sin \alpha \cos^2 \alpha / r^2 - v_r v_\theta / r \right) \tag{13}$$

The corresponding costate equations are defined by $\dot{\boldsymbol{\lambda}} = -\partial \mathcal{H} / \partial \boldsymbol{x}$ and are given as

$$\begin{aligned}
\dot{\lambda}_1 &= \lambda_2 v_\theta / r^2 + \\
&\quad \lambda_3 \left(2\beta \cos^3 \alpha / r^3 + v_\theta^2 / r^2 - 2/r^3 \right) + \lambda_4 \left(2\beta \sin \alpha \cos^2 \alpha / r^3 - v_r v_\theta / r^2 \right) && (14) \\
\lambda_2 &= \text{const.} \begin{cases} = 0 & \textit{iff} \quad \theta(t_f) = \text{free} \\ \neq 0 & \textit{iff} \quad \theta(t_f) = \theta_f \end{cases} && (15) \\
\dot{\lambda}_3 &= -\lambda_1 + \lambda_4 v_\theta / r && (16) \\
\dot{\lambda}_4 &= -\lambda_2 / r - 2\lambda_3 v_\theta / r + \lambda_4 v_r / r && (17)
\end{aligned}$$

Applying Pontryagin's Minimum Principle (Bryson and Ho, 1975), the optimal control $u^\star \equiv \alpha^\star$ is chosen such that the Hamiltonian $\mathcal{H}$ is minimized; that is,

$$u^\star = \arg \min_{u \in \mathcal{U}} \mathcal{H} \left(\boldsymbol{x}^\star, \boldsymbol{\lambda}^\star, u \right) \ , \qquad \forall t \geqslant 0 \tag{18}$$

where $\boldsymbol{x}^\star$ and $\boldsymbol{\lambda}^\star$ denote the optimal state and costate vector. Therefore the stationary condition yields

$$\partial\mathcal{H}/\partial\alpha = 0 = -3\lambda_3\beta\,\sin\alpha\,\cos^2\alpha/r^2 + \lambda_4\beta\left(\cos^3\alpha - 2\sin^2\alpha\,\cos\alpha\right)/r^2 \tag{19}$$

which is satisfied if

$$\begin{cases} \cos\alpha^\star = 0 \\ \cos\alpha^\star \neq 0 \quad \text{and} \quad \tan^2\alpha^\star + \dfrac{3\lambda_3}{2\lambda_4}\tan\alpha^\star - \dfrac{1}{2} = 0 \end{cases} \tag{20}$$

The optimal control angle $\alpha^\star$ which minimizes the Hamiltonian in equation (13) is given by (Wood et al., 1982)

$$\alpha^\star = \begin{cases} \tan^{-1}\left\{\left(-3\lambda_3 - \sqrt{9\lambda_3^2 + 8\lambda_4^2}\right)/(4\lambda_4)\right\} & \textit{if} \quad \lambda_4 \neq 0 \\ 0 & \textit{if} \quad \lambda_4 = 0\,, \lambda_3 < 0 \\ \pm\pi/2 & \textit{if} \quad \lambda_4 = 0\,, \lambda_3 > 0 \end{cases} \tag{21}$$

Anticipating one of the system symmetry properties discussed in a later section, we point out that the transformation $\lambda_4 \to -\lambda_4 \;\Rightarrow\; \alpha^\star \to -\alpha^\star$ in control law (21).

3.2. Control problem family specifics

The costate equations (15–17) and the control law (21) are valid irrespective of the specific problem family. However, transversality conditions for the Hamiltonian as well as interior– and end–point constraints require further analysis.

3.2.1. The minimum–time orbit transfer problem – the $\mathcal{P}_0$ family

The minimum–time orbit transfer problem is the most well–documented problem family in the literature (Tsu, 1959; London, 1960; Zhukov and Lebedev, 1964; Sauer, 1976; Jayaraman, 1980; Wood et al., 1982; Powers et al., 1999; Powers and Coverstone, 2001; Dachwald, 2004). Note that for the orbit transfer problem the angle θ is not only an ignorable coordinate (no differential constraint) but also does not appear in any algebraic constraint. As an immediate consequence, solutions to the angularly unconstrained $\mathcal{P}_0$ family are true global minima of the more general and angularly constrained problem family $\mathcal{P}_2$. Without loss of generality the Lagrange multiplier associated with the ignorable coordinate can be set $\lambda_2 = 0$. The remaining reduced–order system leads to a well–defined Two–Point Boundary Value Problem (TPBVP) with three (constraint) equations (7) in three unknowns. Note that the final time t_f is unspecified; therefore for problem families $\mathcal{P}_0$ and $\mathcal{P}_1$ the transversality condition $\mathcal{H}(t_f) = -a,\ a > 0$ must be satisfied. Moreover, since the motion equations are autonomous

$$d\mathcal{H}/dt = \partial\mathcal{H}/\partial t = 0 \quad \Rightarrow \quad \mathcal{H} = \text{const.} = -a \tag{22}$$

Additionally, since $\boldsymbol{\lambda}$ appears linearly in the Hamiltonian, scaling freedom allows for fixing one of the Lagrange multipliers at $t = t_0$ such that $\lambda_i(t_0) = b,\ b \neq 0$, and $\text{sign}\,(b)$ chosen such that condition (22) is satisfied.

3.2.2. The minimum–time two–way orbit transfer problem – the $\mathcal{P}_1$ family

Minimum–time two–way orbit transfer problems present a challenging extension to the $\mathcal{P}_0$ family in that additional interior–point constraints have to be satisfied. One approach to solve the resulting Three–Point Boundary Value Problems is to convert the problem to an equivalent TPBVP. By doing so, the interior point and boundary points of the Three–Point Boundary Value Problem are transformed to initial and final points of the TPBVP. The resulting TPBVP is of higher dimension but less challenging to solve numerically. We solve the Three–Point Boundary Value Problem following the procedure outlined in Bryson and Ho (1975). At the interior point $t_0 < t_i < t_f$ boundary conditions of the form

$$\boldsymbol{\upsilon}(\boldsymbol{x}_i, t_i) = (r(t_i) - r_i,\ v_r(t_i),\ v_\theta(t_i) - 1/\sqrt{r_i})^T = \mathbf{0} \tag{23}$$

have to be satisfied. The corresponding jump conditions for the Lagrange multipliers then yield

$$\boldsymbol{\lambda}_r(t_i^-) = \boldsymbol{\lambda}_r(t_i^+) + \boldsymbol{\pi}^T \frac{\partial \boldsymbol{\upsilon}}{\partial \boldsymbol{x}(t_i)} = \boldsymbol{\lambda}_r(t_i^+) + \boldsymbol{\pi} \tag{24}$$

$$\mathcal{H}(t_i^-) = \mathcal{H}(t_i^+) - \boldsymbol{\pi}^T \frac{\partial \boldsymbol{\upsilon}}{\partial t_i} = \mathcal{H}(t_i^+) \tag{25}$$

where $\boldsymbol{\lambda}_r = (\lambda_1, \lambda_3, \lambda_4)^T$ are the Lagrange multipliers for the reduced system ($\lambda_2 = 0$), $\boldsymbol{\pi}$ are the (constant) adjoined multipliers, and t_i^- and t_i^+ signify the immediate points in time before and after $t = t_i$, respectively. The increased complexity in solving for solutions of the $\mathcal{P}_1$ problem family stems from the additional unknowns: the adjoined multipliers $\boldsymbol{\pi}$ and the optimal return transfer time. Note that a particular $\mathcal{P}_1$ solution can always be separated into two independent $\mathcal{P}_0$–type solutions due to missing angular constraints. Additionally, system symmetry can be used to great advantage to obtain one of the $\mathcal{P}_0$–type solutions given the other. Symmetry properties are discussed in the following section.

3.2.3. The minimum–time planet rendezvous problem – the $\mathcal{P}_2$ family

For the rendezvous problem, the motion of the corresponding planets has to be taken into account. The angular end–point constraint in equation (7) yields

$$\theta(t_f) - \theta_f = 0\ , \quad \text{where} \quad \theta_f = T\dot{\Theta}_t - \Delta\Theta(t_0) = Tv_t^3 - \Delta\Theta(t_0) \tag{26}$$

In equation (26), $T = t_f - t_0$ is the transfer time, $\dot{\Theta}_t$ and v_t are angular rate and velocity of the target planet, and $\Delta\Theta(t_0)$ is the initial angular separation between initial and target planet. Note that due to system symmetry

$$\Delta\Theta(t_0) = -\text{sign}\left(\dot{\Theta}_i - \dot{\Theta}_t\right)(\Theta_i(t_0) - \Theta_t(t_0))\ ; \tag{27}$$

that is, $\Delta\Theta(t_0)$ in equation (26) only depends on the relative orientation of initial and target planet. Also, with constraint (26) the transversality condition for the Hamiltonian for problem families $\mathcal{P}_2$ and $\mathcal{P}_3$ becomes $\mathcal{H}(t_f) = \mathcal{H} = \breve{a},\ \breve{a}$ indefinite (as opposed to $\mathcal{H}(t_f) = -a,\ a > 0$ for the angularly unconstrained $\mathcal{P}_0$ and $\mathcal{P}_1$ families). As before, with scaling freedom, $\lambda_i(t_0) = \breve{b},\ \breve{b} \neq 0$ and sign $\left(\breve{b}\right)$ chosen to satisfy $\mathcal{H} = \breve{a}$.

3.2.4. The minimum–time two–way planet rendezvous problem – the $\mathcal{P}_3$ family

Unlike for the $\mathcal{P}_1$ problem family we conjecture that $\mathcal{P}_3$ transfers can not be separated into two independent $\mathcal{P}_2$–type solutions. As mentioned before, the standard approach to obtain two–way trajectories is to solve the associated Three–Point Boundary Value Problems with appropriately chosen interior– and end–point constraints.

In the following section we discuss system symmetries as a means to assist in computing optimal trajectories.

4. Symmetries

In the following we present and prove two system symmetries summarized in Theorem 1 and Theorem 2. Theorem 1 provides an effective tool to determine optimal return trajectories, which can be used to compute solutions of the $\mathcal{P}_1$ family efficiently. As pointed out in Section 3.2.2, solving the associated Three–Point Boundary Value Problem is not only numerically challenging but also a time–consuming process. Another interesting symmetry property is formulated in Theorem 2. Using nondimensional analysis one can show similarity of minimum–time solutions for the case when $r(t_f)/r(t_0) = \text{const}$. Moreover, there exists a simple relationship between transfer times and initial (or final) orbit radii for trajectories with $r(t_f)/r(t_0) = \text{const}$.

Definition A function $\boldsymbol{\Omega} = (\boldsymbol{X}, \boldsymbol{\Lambda}, U, T)^T$ is called a solution trajectory of the optimal control problem if $\boldsymbol{X}$ and $\boldsymbol{\Lambda}$ are compatible solutions to the state equations (1–4) and costate equations (15–17), respectively, with control history U and transfer time T, for a given set of boundary values.

Theorem 1. *Let $\boldsymbol{\Omega}$ be a $\mathcal{P}_0$ solution trajectory satisfying the boundary conditions $\boldsymbol{X}(t_0) = \boldsymbol{X}_0$, $\boldsymbol{X}(t_f) = \boldsymbol{X}_f$, $\boldsymbol{\Lambda}(t_0) = \boldsymbol{\Lambda}_0$, and $\boldsymbol{\Lambda}(t_f) = \boldsymbol{\Lambda}_f$, then the costate solution for the corresponding return trajectory $\boldsymbol{\Omega}^* = (\boldsymbol{X}^*, \boldsymbol{\Lambda}^*, U^*, T^*)^T$ satisfies*

$$\Lambda_1^*(t^*) = -\Lambda_1(t)\,, \quad \Lambda_2^*(t^*) = \Lambda_2(t) = 0\,, \quad \Lambda_3^*(t^*) = \Lambda_3(t)\,, \quad \Lambda_4^*(t^*) = -\Lambda_4(t)\,,$$

with

$$t^* = T - t$$

and where $T = t_f - t_0 = t_f^* - t_0^* = T^*$.

Proof. The boundary conditions for the states are trivially compatible under the symmetry transformation for the independent variable t. With $(d/dt^*) = -(d/dt)$ it follows from equation (1) that $v_r^*(t^*) = -v_r(t)$ and from equation (3) that $r^*(t^*) = r(t)$. For system invariance the control angle satisfies $\alpha^*(t^*) = -\alpha(t)$ [equation (4)], which is compatible with the symmetry transformations for the Lagrange multipliers and the control law (21). Similarly, with the proposed symmetry transformations the costate equations are rendered invariant. □

Remark. Obviously, $(-\Lambda_1(t_f), \Lambda_3(t_f), -\Lambda_4(t_f)) \mapsto (\Lambda_1^*(t_0^*), \Lambda_3^*(t_0^*), \Lambda_4^*(t_0^*))$ and the return trajectory is readily propagated forward in time. For $\mathcal{P}_0$ solution trajectories θ is an ignorable coordinate and does not appear in any algebraic constraints; therefore, $\Lambda_2 = \Lambda_2^* = 0$ without loss of generality.

Theorem 2. *Let $\mathbf{\Omega}$ be a $\mathcal{P}_0$ solution trajectory with $\Gamma \triangleq r(t_f)/r(t_0)$ then the equivalent $\mathcal{P}_0$ solution trajectory $\mathbf{\Omega}^*$ with $\Gamma^* = \Gamma$ and $r^*(t_0^*) \neq r(t_0)$ satisfies*

$$X_1^*(t^*) = \xi X_1(t)\,, \qquad X_2^*(t^*) = X_2(t)\,, \qquad X_3^*(t^*) = \xi^{-1/2} X_3(t)\,, \quad X_4^*(t^*) = \xi^{-1/2} X_4(t)\,,$$
$$\Lambda_1^*(t^*) = \xi^{-3/2}\Lambda_1(t)\,, \quad \Lambda_2^*(t^*) = \Lambda_2(t) = 0\,, \quad \Lambda_3^*(t^*) = \Lambda_3(t)\,, \qquad \Lambda_4^*(t^*) = \Lambda_4(t)\,,$$

with

$$t^* = \sigma t \qquad \textit{and therefore} \qquad T^*/\,T = \sigma \equiv \xi^{3/2}$$

where $\xi = r^*(t_0^*)/r(t_0) = r^*(t_f^*)/r(t_f)\,,\ T = t_f - t_0\,,$ and $T^* = t_f^* - t_0^*\,.$

Proof. Using as reference distance units 1 DU $= r(t_0)$ and 1 DU* $= r^*(t_0^*)$ to nondimensionalize motion equations and to obtain the corresponding solution trajectories $\mathbf{\Omega}$ and $\mathbf{\Omega}^*$ the state and costate transformations render the systems of differential equations equivalent provided that

$$\tilde{\mu}\,\mathrm{TU}^{\,2}/\,\mathrm{DU}^{\,3} = \tilde{\mu}\,\mathrm{TU}^{*\,2}/\,\mathrm{DU}^{*\,3}$$

□

Remark. For solution trajectories that are not readily obtained with Theorem 1 and Theorem 2, for example, $\Gamma^* \neq \Gamma$ and $r_0^* \neq r_0$, Theorem 2 can be used to reduce the two–parameter continuation problem to a one–parameter continuation problem as follows:

1. Compute $\hat{\mathbf{\Omega}}$ with $\hat{\Gamma} = \Gamma$, $\hat{r}_0 = r_0^*$ (or $\hat{r}_f = r_f^*$), using Theorem 2
2. Use homotopy to calculate $\mathbf{\Omega}^*$ with $r_0^* = \hat{r}_0$ (or $r_f^* = \hat{r}_f$) fixed

Also note the similarity between Kepler's Third Law and the equation in Theorem 2 describing the relationship between minimum transfer times and corresponding initial (or equivalently final) radial distances.

5. Numerical Approach

The inherent difficulty of global optimization problems lies in finding the best optimum from a possible multitude of local optima. In using indirect methods to solve optimum control problems, the difficulty stems from the fact that the initial conditions of the Lagrange multipliers of the associated TPBVP cannot be estimated – not even approximately – without extensive analysis. von Stryk and Bulirsch (1992) and later Seywald and Kumar (1996) introduced the idea of combining direct and indirect methods to obtain approximate solutions with the direct method and to generate accurate solutions with the indirect method. An obvious drawback of this approach is that by using two fundamentally different solution methodologies, the control problem must be formulated twice, as well. Also, an interface is necessary to communicate between the two algorithms (Lagrange multipliers). In this paper the optimization problem is solved with a cascaded computational scheme using an indirect method.

5.1. The computational scheme

Figure 2 illustrates the computational scheme. The qualitative performance of each of the three methods used is indicated in the left part of the figure, with one star indicating relatively poor performance, and three stars indicating relatively good performance. Simulated Annealing (SA) is a global, statistical optimization algorithm, which was first introduced by Kirkpatrick et al. (1983) to solve discrete optimization problems such as computer chip packing and wiring, and to analyze classical problems such as the travelling salesman problem (Cerny, 1985). We use a variant of the SA algorithm, namely Adaptive Simulated Annealing (Ingber and Rosen, 1992), as the initial optimization tool to obtain approximate estimates for the costates and the optimal transfer time $T = t_f - t_0$. Since statistical algorithms are in general neither efficient nor accurate, the global algorithm is used to identify the region in the parameter space that contains the true global minimum.

Once the algorithm has located a set of parameters in close vicinity of the optimal set, a Quasi–Newton method (Gill et al., 1999) is used to further refine the parameter set. A crucial aspect of Quasi–Newton methods is the computation of second–order derivative information. Rather than calculating the Hessian of the objective function accurately at every iteration step or even just every so often we found that approximate Hessian information obtained using update formulas can significantly increase the algorithm effectiveness. In particular the Inverse Rank–One update (Gill et al., 1999) and the Inverse–Broyden–Fletcher–Goldfarb–Shanno (IBFGS) update (Gill et al., 1999) provide satisfactory optimization performance.

Newton's method is known to be the most efficient zero–finding algorithm provided the starting guess of the unknowns lies within the region of attraction of the algorithm. Used in combination with a Quasi–Newton method Newton's method presents an efficient approach to obtain accurate solutions for the present problem.

5.2. Simulated annealing – A global, statistical optimization algorithm

The term simulated annealing (SA) derives from the analogous physical process of thermal annealing (metallurgy) to obtain a defect–free (and so in some sense optimized) crystalline structure. In an annealing process a melt, initially at high temperature and disordered, is cooled in a controlled, slow manner to keep the system in an approximate state of thermodynamic equilibrium (adiabatic cooling). As cooling proceeds, the system becomes ordered and approaches a ground state. In a SA optimization algorithm, the annealed substance corresponds to the system being optimized. Similarly, the current "energy" state of the substance corresponds to the current value of the system cost function, with the goal of identifying the ground state of the system, the global minimum. The internal microscopic interactions that keep the substance in a state of thermodynamical equilibrium are simulated in SA by a sequence of parameter perturbations described by Markov chains. One major difficulty of implementing a SA algorithm is that there is no obvious analog to the

temperature in the physical process. The corresponding SA control parameter serves as a reference energy defining the boundary between the local and global vicinity of the current optimal parameter set in parameter space.

In general statistical optimization methods such as Simulated Annealing differ from deterministic techniques in that the iteration procedure need not converge towards a local optimum since transitions thereout are always possible. Another feature is that an adaptive divide–and–conquer occurs: coarse features of the optimal parameter set appear at higher temperatures, fine details develop at lower temperatures. For a detailed analysis on SA we refer to Kirkpatrick et al. (1983), Cerny (1985), and Ingber and Rosen (1992).

6. Simulation Results

Results obtained for Earth–Mars $\mathcal{P}_0$ transfers are in excellent agreement with data published by Wood et al. (1982). Table I shows transfer times and initial and final costates for two different characteristic accelerations. We obtained high–accuracy results with $|\boldsymbol{\psi}\left(\boldsymbol{x}(t_f), t_f\right)| < 10^{-14}$, which result in slightly improved transfer times in the order of 10 to 15 hours compared to reported results in Wood et al. (1982). Note that the nondimensional characteristic accelerations of $\beta = 0.16892$ and $\beta = 0.33784$ correspond to nominal values of $\tilde{\beta} = 1$ mm/s^2 and $\tilde{\beta} = 2$ mm/s^2. The minimum transfer times correspond to 323.87 and 407.62 days, respectively (1 TU $= 365.25/(2\pi)$ days $= 58.1313$ days).

Figure 3 shows minimum transfer time versus target orbit radius for $\mathcal{P}_0$ orbit transfers starting at 1 AU and for three different characteristic accelerations. According to McInnes (1999), state–of–the–art solar sail spacecraft achieve reasonable characteristic accelerations on the order of $\tilde{\beta} \lessapprox 2$ mm/s^2; thus $\beta = 3.3784$ is only of pedagogical value. Note the increased "sail effectiveness" due to the $1/r^2$ potential field for inbound trajectories when compared to outbound trajectories; that is, inbound transfers take comparatively less time than outbound transfers.

Symmetry properties discussed in Section 4 are illustrated in Figures 4 and 5 and Table II. Figure 4 shows two outbound trajectories, $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$, with $\Gamma_1 = \Gamma_2 = 2$, and $r_1(t_1 = 0) = 1$ AU and $r_2(t_2 = 0) = 1.5$ AU, respectively, and one inbound trajectory, $\boldsymbol{\Omega}_3$, with $\Gamma_3 = 1/2$ and $r_3(t_3 = 0) = 1.5$ AU. As shown in Theorem 2, $\alpha_1(t_1) = \alpha_2(t_2)$ or equivalently, $\bar{\alpha}_1(\tau) = \bar{\alpha}_2(\tau)$ on the unit time interval with $\tau \in [0, 1]$. Also, for the outbound trajectories $\sigma = 1.8371 = 1.5^{3/2}$ with $T_2 = 16.6176$ and $T_1 = 9.0455$. Using both Theorems the inbound trajectory satisfies $\alpha_2(t_2) = -\alpha_3(t_3)$. The inbound transfer time results are $T_3 = 5.8752$ and $\sigma = T_2/\,T_3 = 2^{3/2}$, as expected. As pointed out in Section 3.2.2, $\mathcal{P}_1$ solution trajectories can be obtained by either solving the associated Three–Point Boundary Value Problem or by solving for the corresponding $\mathcal{P}_0$ solution and using Theorem 1. Figure 5 shows a $\mathcal{P}_1$ Earth–Mars–Earth transfer for $\beta = 0.33784$ obtained by solving the Three–Point Boundary Value Problem. Similarly to Figure 4 the sail orientation angle clearly shows system symmetry. Accompanying simulation data are listed in Table II. Note that the initial, and therefore also the final, Lagrange multipliers are scaled such that $\lambda_1(t_0) = -1$ and

$\lambda_1(t_f) = +1$. The adjoined multipliers $\boldsymbol{\pi}$, are essential to solve the boundary value problem (BVP), but dispensable when using symmetry (dashed–lined arrows).

Multiple solutions to nonlinear BVP are possible. Figure 6 shows two globally minimal Earth–Mars $\mathcal{P}_2$ solutions for an initial angular separation between Mars and Earth of $\Delta\Theta(t_0) = 1.117$ rad ($\beta = 0.135136$). For the two significantly different trajectories, we refer to the spacecraft as being in either "sleep" or "catch–up" mode. Sleep–mode solutions are characterized by a dramatic initial increase of the radial distance of the spacecraft reducing its angular rate. Midway through the transfer the angular rate is further decreased by reducing the sail orientation angle, gravity becomes dominant, and the spacecraft naturally falls back into the target orbit. Comparing sleep and catch–up type trajectories, the latter are more dynamically "active". As illustrated in Figure 6 the control angle history for the sleep mode trajectory shows a distinctively more moderate functional behavior. Additionally, catch–up mode trajectories typically include one or more solar gravity/radiation assists. The tangential velocity v_θ is generally higher for catch–up mode than for corresponding sleep mode trajectories. Not surprisingly, once close to the object planet (Earth) the spacecraft in catch–up mode approaches its target from "behind". In sleep mode the spacecraft slows down appropriately to be approached by its target.

Obviously, the two trajectories in Figure 6 are members of two fundamentally different solution branches. For fixed initial angular separation $\Delta\Theta(t_0)$ of the planets, there exist in general an infinite number of locally optimal $\mathcal{P}_2$ solutions but only one or two globally optimal solution trajectories as illustrated in Figure 7. Varying $\Delta\Theta(t_0)$, the minimum–time $\mathcal{P}_2$ trajectory to transfer between arbitrary points on the initial and target orbits coincides with the corresponding $\mathcal{P}_0$ solution; that is, $T^{\min}_{\mathcal{P}_2} = T_{\mathcal{P}_0}$. In Figure 7 the corresponding $\mathcal{P}_0$–type solution is located at $T^{\min}_{\mathcal{P}_2} = 7.01204$ and $\Delta\Theta(t_0) = 2.67402$ rad with $\beta = 0.16892$ according to Table I. Starting at the $\mathcal{P}_0$–type solution, the $\mathcal{P}_2$ solution family consists of two solution branches: for decreasing $\Delta\Theta(t_0)$ we find catch–up type solutions, whereas for increasing $\Delta\Theta(t_0)$ sleep mode turns out to be more time–efficient. The two branches represent globally optimal solutions for $T_{\mathcal{P}_2} \leqslant T^{\max}_{\mathcal{P}_2} = 12.59186$, where the branches intersect (first crossover = Darboux point) and the global optimum switches from sleep– to catch–up–type solution trajectories, and vice versa. Note that the Darboux point marks solution trajectories for pessimal $\Delta\Theta(t_0)$. A further increase or decrease in $\Delta\Theta(t_0)$ yields locally optimal solutions, with faster, globally optimal solutions readily apparent.

Figures 8 and 9 illustrate minimum transfer time as a function of initial angular separation $\Delta\Theta(t_0)$ for different characteristic accelerations for Mars–Earth and Earth–Mars $\mathcal{P}_2$ transfers. Unlike for $\mathcal{P}_0$ solutions, symmetries of the $\mathcal{P}_2$ family are less distinct. Nevertheless, minimum $T^{\min}_{\mathcal{P}_2}$ and maximum $T^{\max}_{\mathcal{P}_2}$ transfer times agree for both Mars–Earth and Earth–Mars $\mathcal{P}_2$ solution families. Lower and upper bounds on transfer time as a function of characteristic acceleration are illustrated in Figure 10 for Mars–Earth and Earth–Mars transfers. We point out again that reasonable characteristic accelerations for solar sail spacecraft are in the order of $\tilde{\beta} \lessapprox 2$ mm/s^2. Figure 11 shows optimal $\Delta\Theta^{\text{opt}}(t_0)$ and pessimal initial angular separation $\Delta\Theta^{\text{pes}}(t_0)$ versus characteristic acceleration. Note as $\beta \to \infty$ the difference $(\Delta\Theta^{\text{opt}}(t_0) - \Delta\Theta^{\text{pes}}(t_0)) \to \pi$.

7. Summary and Conclusions

We study the optimal control problem of solar sail spacecraft for planar interplanetary missions in detail. The optimization problem is solved using an indirect method. The cascaded computational scheme is divided into two optimization levels. On the first level a global statistical algorithm based on Adaptive Simulated Annealing is used to find an approximate guess for the Lagrange multipliers and the transfer time. The optimization parameters are then refined using a Quasi–Newton method. The final optimization stage is realized using a Newton's method. The composite algorithm proves extremely efficient finding highly accurate solutions to the minimum–time control problem.

We obtain optimal trajectories for several interrelated problem families that are described as Multi–Point Boundary Value Problems. We present and prove two theorems describing system symmetries. We demonstrate how these symmetry properties can be used to significantly simplify the solution–finding process. For the minimum–time transfer between two planetary orbits with subsequent return transfer, only a Two–Point Boundary Value Problem has to be solved when using symmetry as opposed to the associated Three–Point Boundary Value Problem. Another system symmetry allows for efficient computation of solution trajectories by replacing a two–parameter continuation problem by a corresponding one–parameter continuation problem.

8. Future Work

In the future we intend to analyze more complex optimal control problems such as non–planar transfers between planets in elliptical orbits. From a dynamics point of view it is unlikely that symmetries exist for minimum–time transfer problems with arbitrary planetary orbits. Nevertheless, the approach presented in this paper can be used to generate initial guesses for the costate vector for the more general class of optimization problems.

Choosing the solar sail orientation angle as the control variable yields analytically simple control laws for spacecraft modelled as point masses. In practice, however, a rigid–body model might be more appropriate. Consequently, a combination of control forces and torques replaces the solar sail orientation angle as the control variable. As an interesting consequence, the particular choice of control variables might yield bang–type control laws with the controls (forces and torques) appearing linearly in the Hamiltonian, which in turn could entail controllability issues.

Acknowledgements

The authors would like to thank Dr. Eugene Cliff for several fruitful discussions during the preparation of this work.

References

Bryson, Jr. A. E. and Ho, Yu-Chi. 1975. *Applied Optimal Control*. Taylor & Francis. Revised Printing.

Cerny, V., 1985. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm. *Journal of Optimization Theory and Applications*, **45**(1), 41–51.

Dachwald, B., 2004. Optimization of interplanetary solar sailcraft trajectories using evolutionary neurocontrol. *Journal of Guidance, Control, and Dynamics*, **27**(1), 66–72.

Gill, P. E., Murray, W., and Wright, M. H. 1999. *Practical Optimization*. Academic Press Limited, 24/28 Oval Road, London NW1 7DX.

Hughes, G. W. and McInnes, C. R. 2001. Solar sail hybrid trajectory optimization. *Proceedings of the 2001 AAS/AIAA Astrodynamics Specialist Conference, Quebec City, Quebec, Canada*, volume 109 of *Advances in the Astronautical Sciences*, pages 2369–2380.

Ingber, L. and Rosen, B., 1992. Genetic algorithms and very fast simulated reannealing: A comparison. *Mathematical and Computer Modelling*, **16**(11), 87–100.

Jayaraman, T. S., 1980. Time–optimal orbit transfer trajectory for solar sail spacecraft. *Journal of Guidance and Control*, **3**(6), 536–542.

Kirkpatrick, S., Gelatt, Jr., C. D., and Vecchi, M. P., 1983. Optimization by simulated annealing. *Science*, **220**(4598), 671–680.

London, H. S., 1960. Some exact solutions of the equations of motion of a solar sail with constant sail setting. *American Rocket Society Journal*, **30**, 198–200.

McInnes, C. R. 1999. *Solar Sailing: Technology, Dynamics, and Mission Applications*. Space Science and Technology. Springer. Published in association with Praxis Publishing.

Powers, R. B. and Coverstone, V. L., 2001. Optimal solar sail orbit transfers to synchronous orbits. *The Journal of the Astronautical Sciences*, **49**(2), 269–281.

Powers, R. B., Coverstone-Carroll, V. L., and Prussing, J. E. 1999. Solar sail optimal orbit transfers to synchronous orbits. *Proceedings of the 1999 AAS/AIAA Astrodynamics Specialist Conference, Girdwood, Alaska*, volume 103 of *Advances in the Astronautical Sciences*, pages 523–538.

Sauer, C. G. 1976. Optimum solar–sail interplanetary trajectories. Paper presented at AIAA/AAS Astrodynamics Conference, San Diego, CA, AIAA Paper 76–792.

Seywald, H. and Kumar, R. R., 1996. Method for automatic costate calculation. *Journal of Guidance, Control, and Dynamics*, **19**(6), 1252–1261.

Tsander, K., 1967 (quoting a 1924 report by the author). From a scientific heritage. *NASA Technical Translation TTF–541*.

Tsiolkovsky, K. E., 1921. Extension of man into outer space. Cf. Also K. E. Tsiolkovskiy, Symposium on Jet Propulsion, No. 2, United Scientific and Technical Presses (NIT), 1936 (in Russian).

Tsu, T. C., 1959. Interplanetary travel by solar sail. *American Rocket Society Journal*, **29**, 422–427.

von Stryk, O. and Bulirsch, R., 1992. Direct and indirect methods for trajectory optimization. *Annals of Operations Research*, **37**, 357–373.

Wood, L. J., Bauer, T. P., and Zondervan, K. P., 1982. Comment on time–optimal orbit transfer trajectory for solar sail spacecraft. *Journal of Guidance, Control, and Dynamics*, **5**(2), 221–224.

Wright, J. L. 1992. *Space Sailing*. Gordon and Breach Science Publishers. Second printing 1993.

Zhukov, A. N. and Lebedev, V. N., 1964. Variational problem of transfer between heliocentric orbits by means of a solar sail. *Kosmicheskie Issledovaniya (Cosmic Research)*, **2**, 45–50.

Table I
Minimum transfer times and corresponding costates for Earth–Mars transfers.

Analysis Authors	Characteristic acceleration β	Transfer time T	Initial costates $\boldsymbol{\lambda}_r(0)$	Final costates $\boldsymbol{\lambda}_r(T)$
Wood et al. (1982)	0.16892	7.02232	$\begin{pmatrix}-7.40981\\-4.22855\\-7.91115\end{pmatrix}$	$\begin{pmatrix}-21.5481\\+\ 8.5025\\-46.6174\end{pmatrix}$
	0.33784	5.57911	$\begin{pmatrix}-3.68044\\-2.59597\\-2.56421\end{pmatrix}$	$\begin{pmatrix}-14.9271\\+10.1662\\-29.6344\end{pmatrix}$
Kim and Hall	0.16892	7.01204	$\begin{pmatrix}-7.41099\\-4.23234\\-7.90591\end{pmatrix}$	$\begin{pmatrix}-21.5482\\+\ 8.5063\\-46.5737\end{pmatrix}$
	0.33784	5.57134	$\begin{pmatrix}-3.68301\\-2.59741\\-2.56208\end{pmatrix}$	$\begin{pmatrix}-14.9274\\+10.1619\\-29.6045\end{pmatrix}$

Table II
Lagrange and adjoined multipliers for an Earth–Mars–Earth minimum–time transfer for $\beta = 0.33784$.

Solution arc	Lagrange multipliers $\boldsymbol{\lambda}_r(t)$ at Earth orbit	Lagrange multipliers $\boldsymbol{\lambda}_r(t)$ at Mars orbit	Adjoined multipliers $\boldsymbol{\pi}$	Transfer time
Outbound trajectory	$\begin{pmatrix}-1.00000\\-0.70524\\-0.69565\end{pmatrix}$	$\begin{pmatrix}-4.05305\\+2.75912\\-8.03812\end{pmatrix}$		5.57134
Turning point			$\begin{pmatrix}-\ 8.10610\\-\ 0.00000\\-16.07625\end{pmatrix}$	
Inbound trajectory	$\begin{pmatrix}+1.00000\\-0.70524\\+0.69565\end{pmatrix}$	$\begin{pmatrix}+4.05305\\+2.75912\\+8.03812\end{pmatrix}$		5.57134

Figure Captions

Figure 1: System model.

Figure 2: Cascaded numerical algorithm for solving the optimal control problem.

Figure 3: Transfer time as a function of target orbit radius. Initial orbit radius is 1 AU.

Figure 4: Minimum–time orbit transfer system symmetry.

Figure 5: Earth–Mars–Earth minimum–time orbit transfer for $\beta = 0.33784$.

Figure 6: Minimum–time Mars–Earth rendezvous trajectories for $\beta = 0.135136$.

Figure 7: Minimum transfer time for Mars–Earth rendezvous for $\beta = 0.16892$.

Figure 8: Minimum transfer time for Mars–Earth rendezvous.

Figure 9: Minimum transfer time for Earth–Mars rendezvous.

Figure 10: Minimum transfer time for optimal and pessimal initial phase difference between Mars and Earth.

Figure 11: Optimal and pessimal initial phase difference between Mars and Earth versus characteristic acceleration.

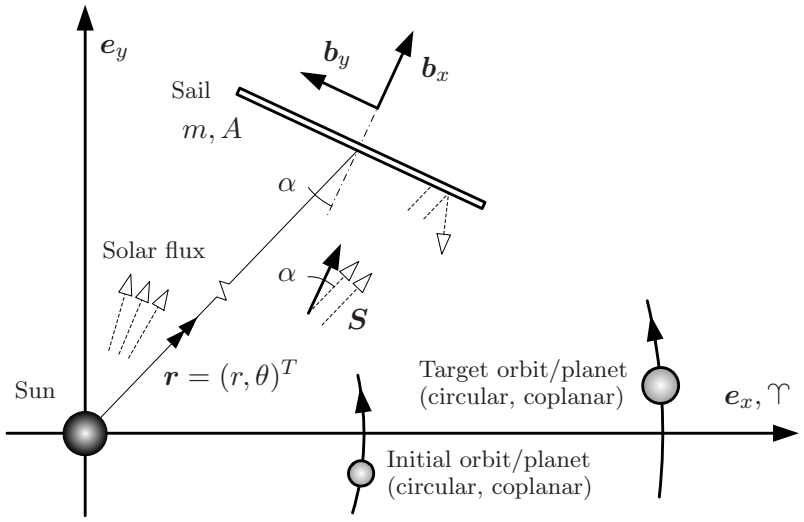

Figure 1. System model.

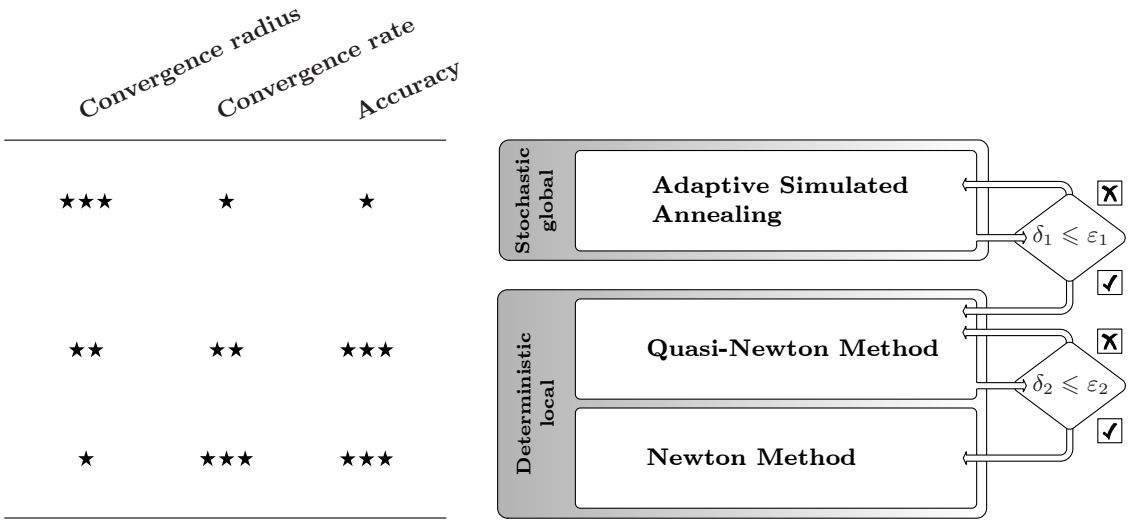

Figure 2. Cascaded numerical algorithm for solving the optimal control problem.

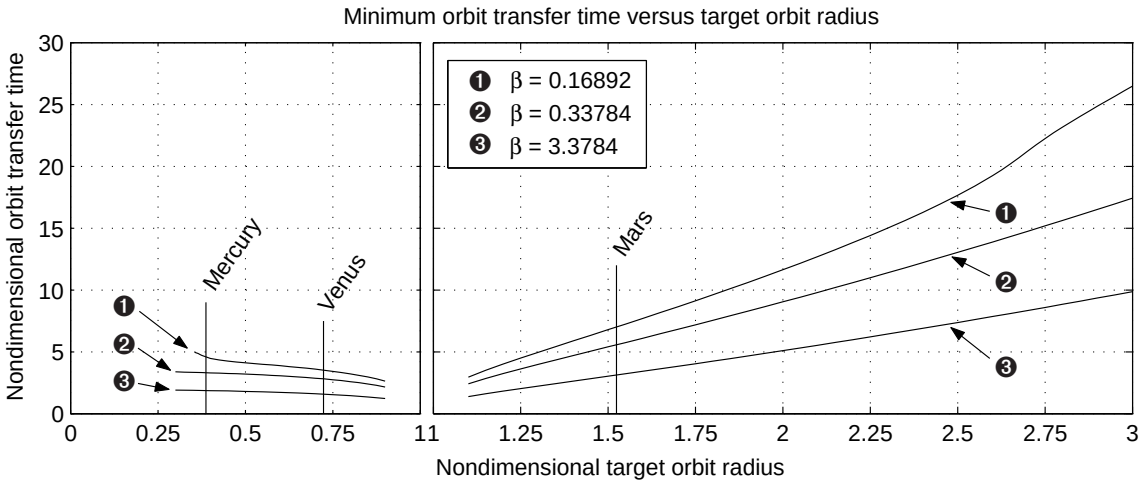

Figure 3. Transfer time as a function of target orbit radius. Initial orbit radius is 1 AU.

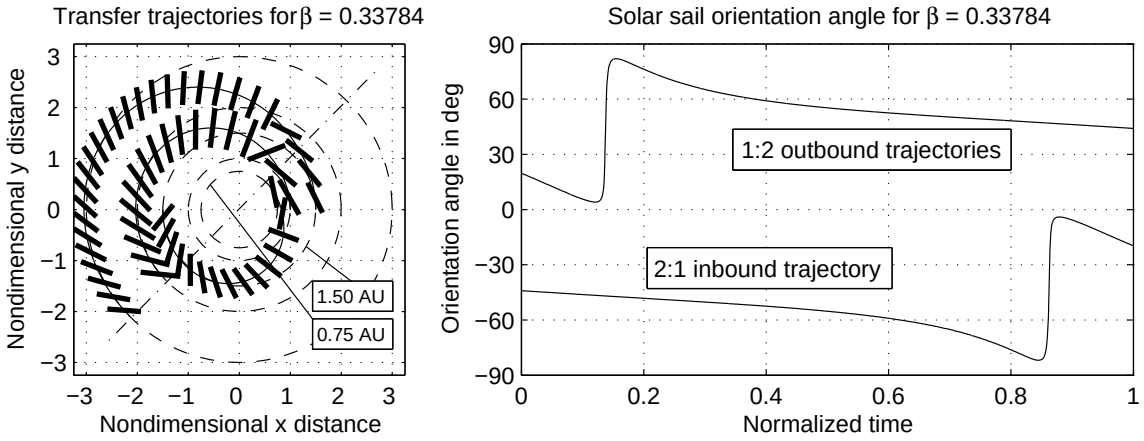

Figure 4. Minimum–time orbit transfer system symmetry.

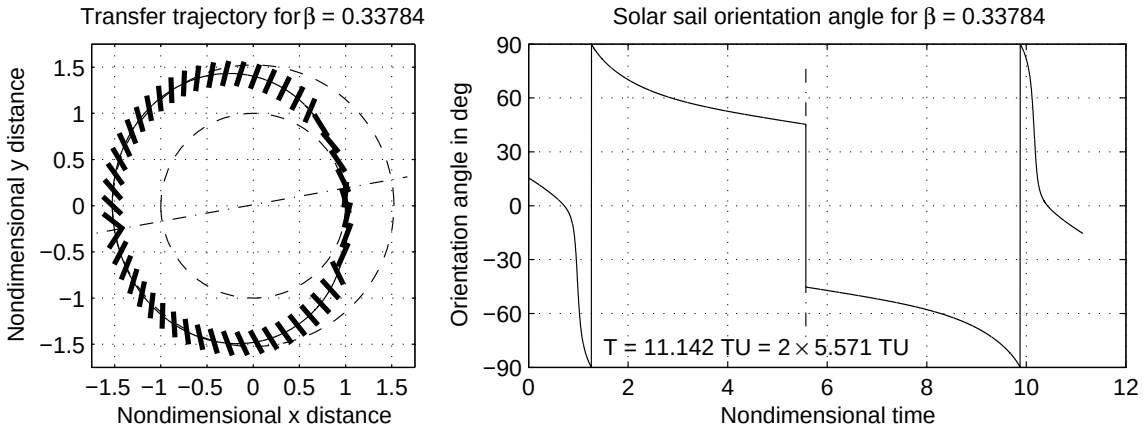

Figure 5. Earth–Mars–Earth minimum–time orbit transfer for $\beta = 0.33784$.

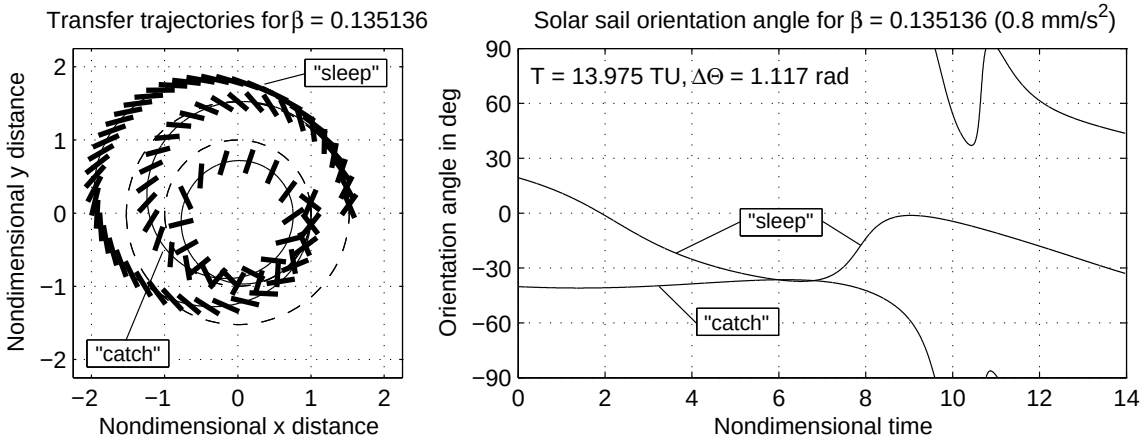

Figure 6. Minimum–time Mars–Earth rendezvous trajectories for $\beta = 0.135136$.

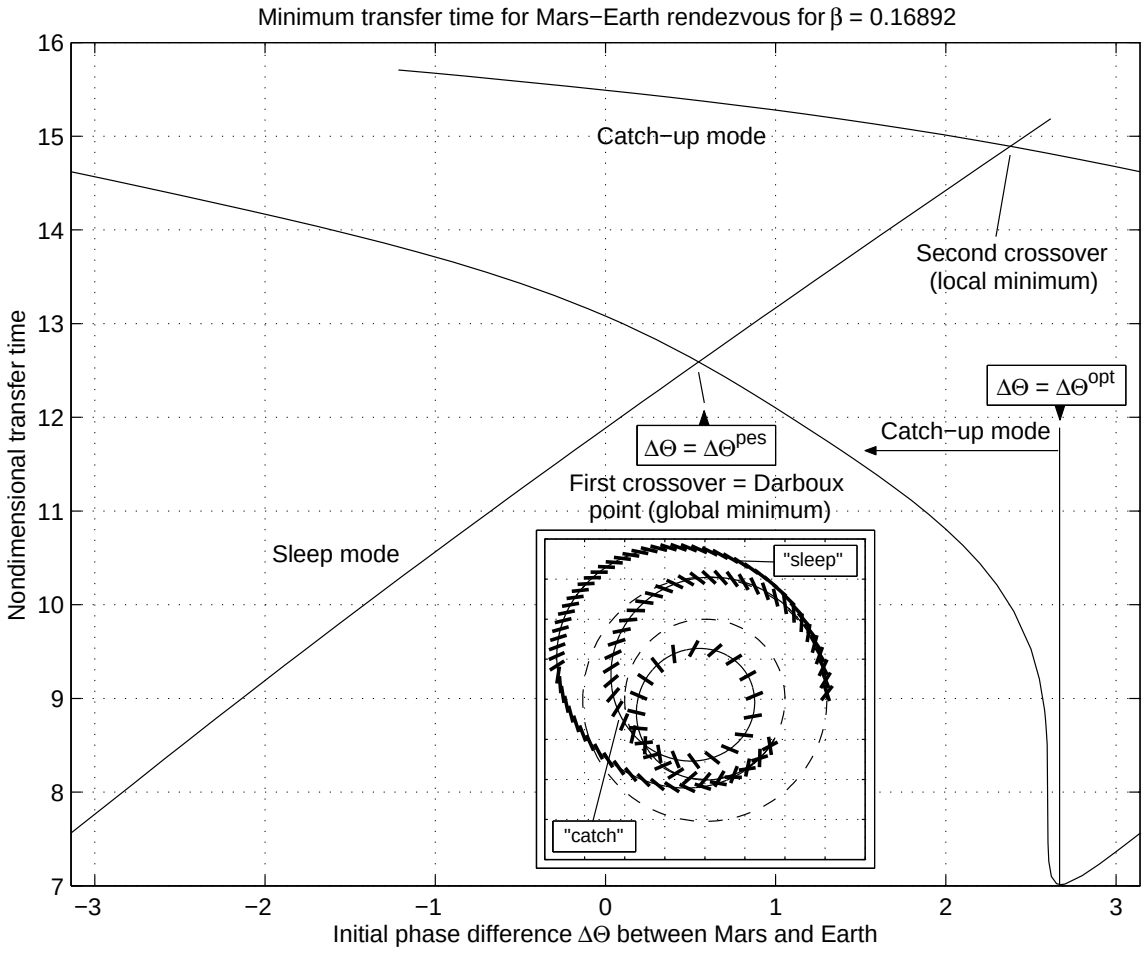

Figure 7. Minimum transfer time for Mars–Earth rendezvous for $\beta = 0.16892$.

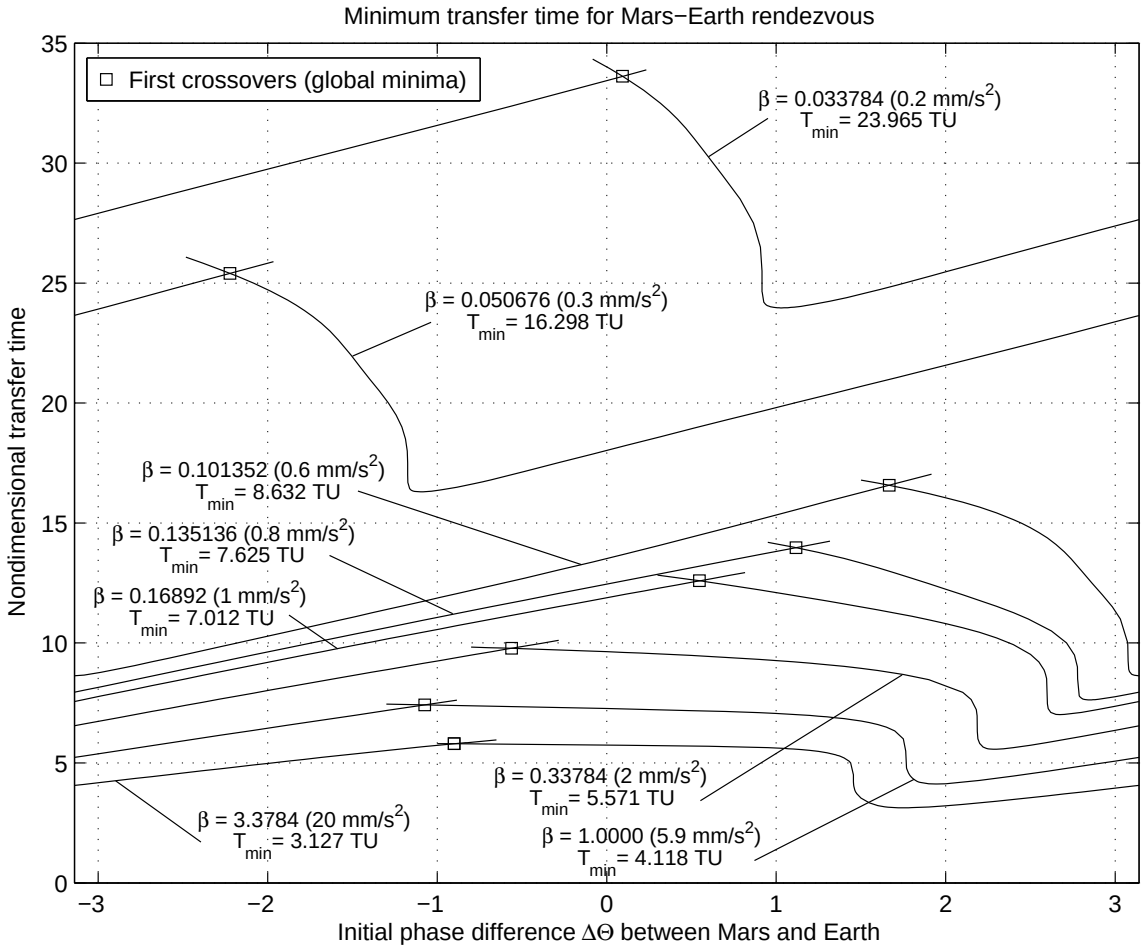

Figure 8. Minimum transfer time for Mars–Earth rendezvous.

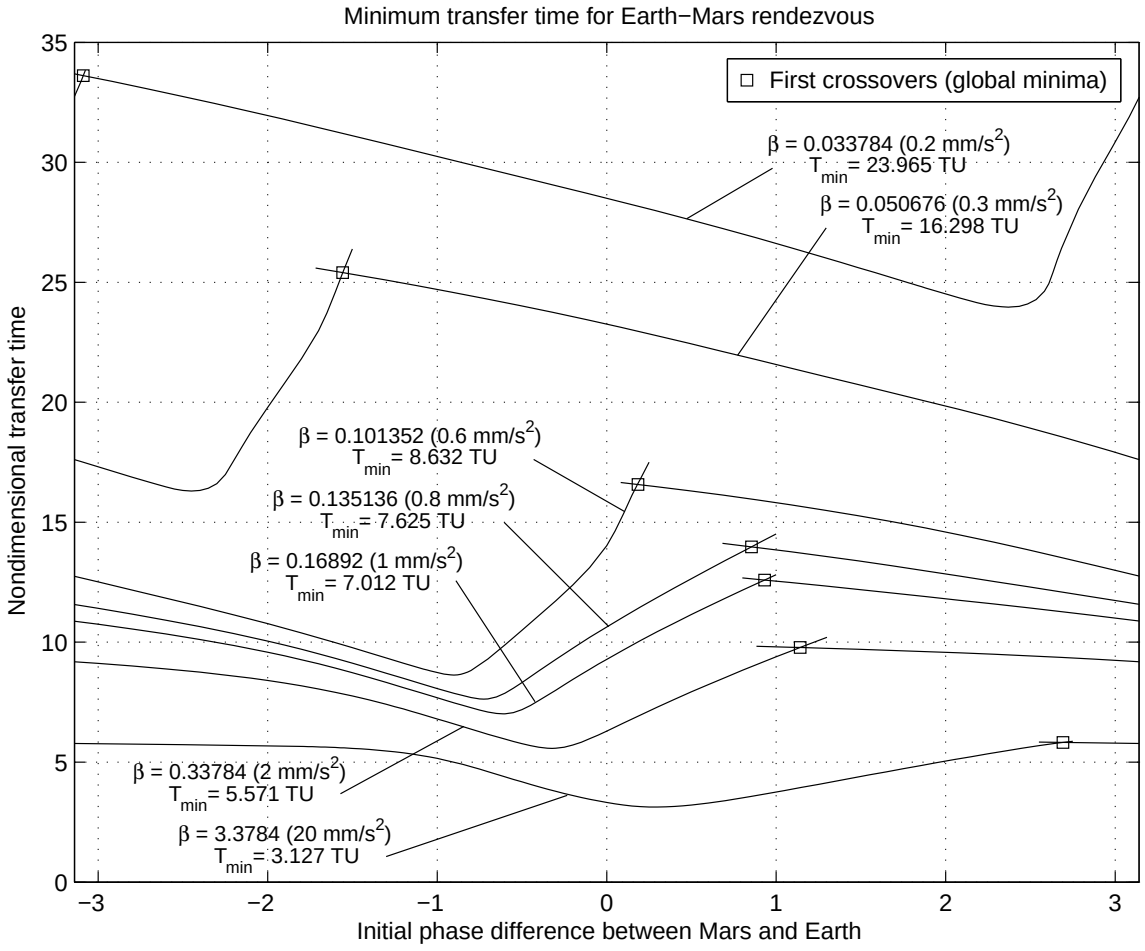

Figure 9. Minimum transfer time for Earth–Mars rendezvous.

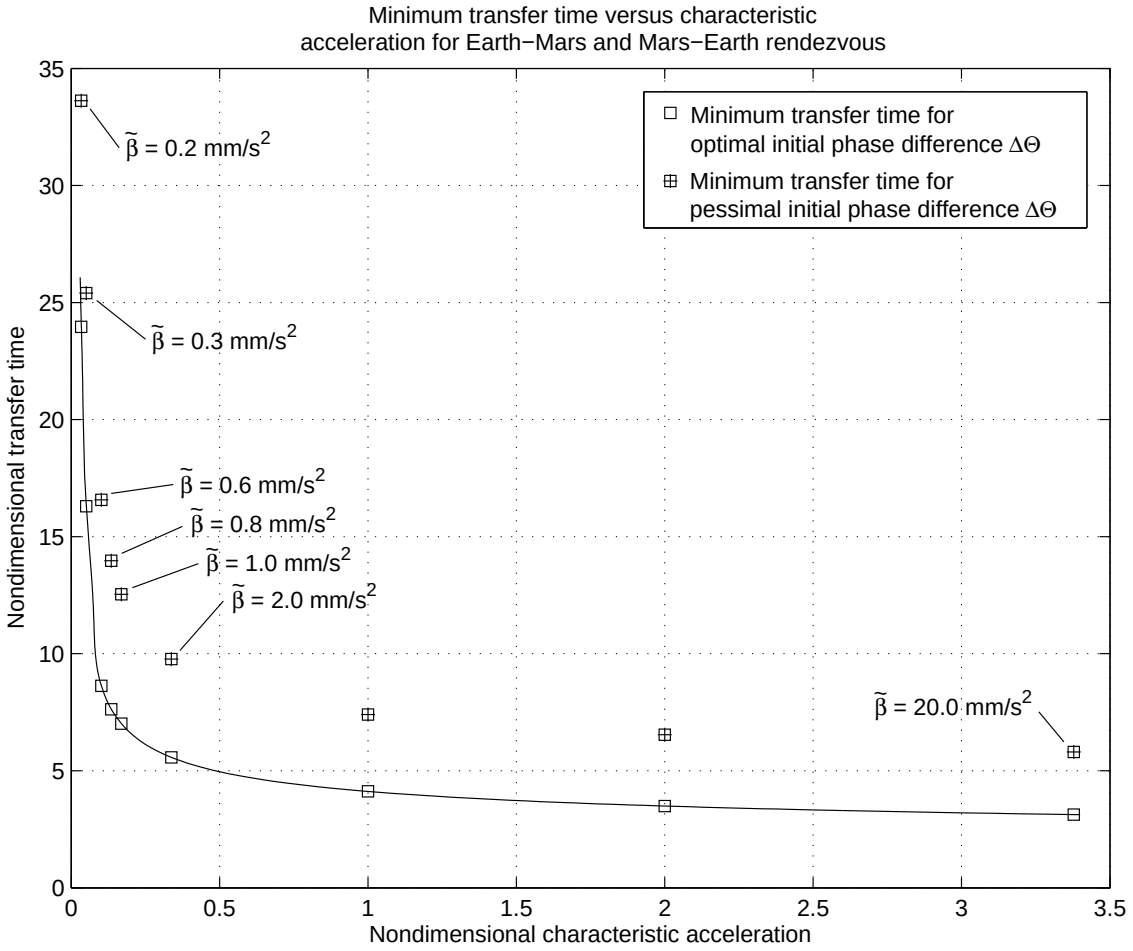

Figure 10. Minimum transfer time for optimal and pessimal initial phase difference between Mars and Earth.

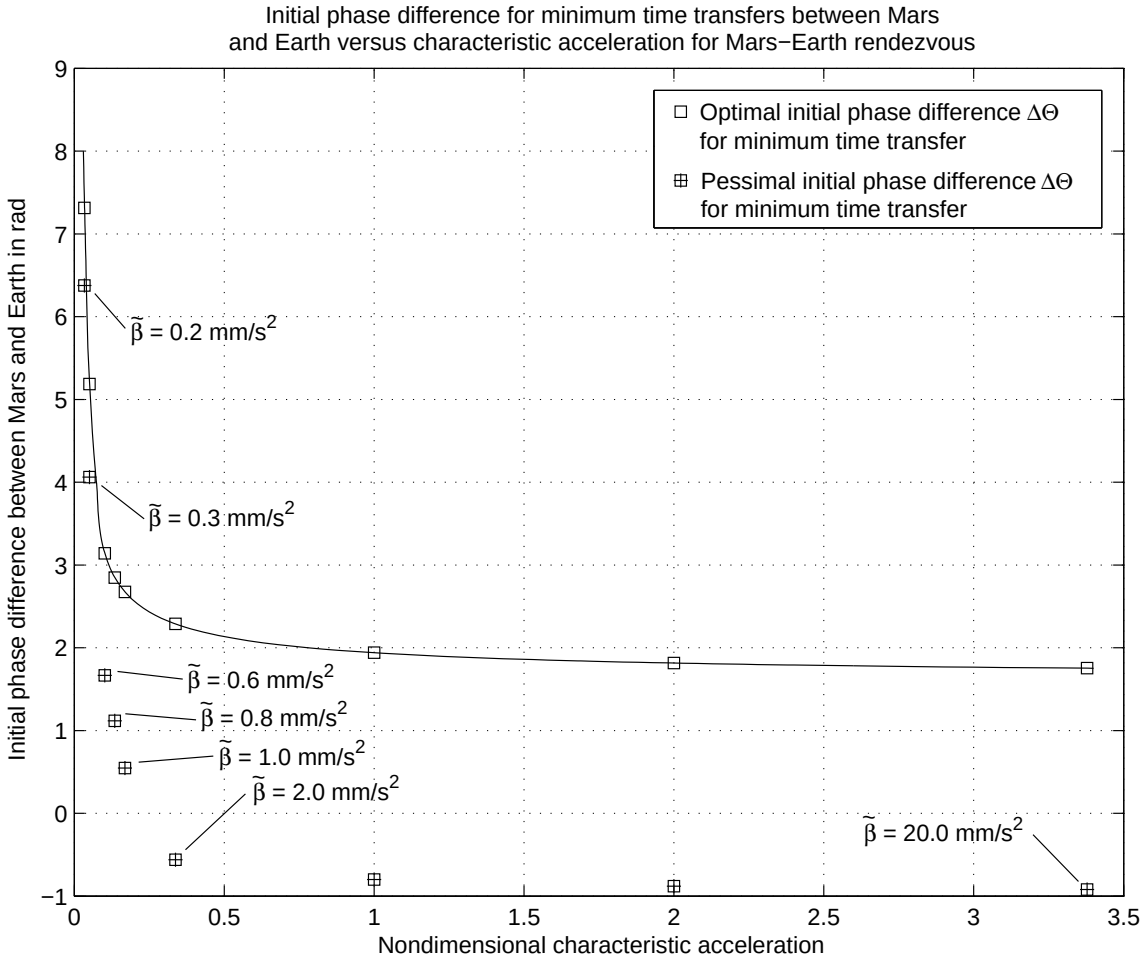

Figure 11. Optimal and pessimal initial phase difference between Mars and Earth versus characteristic acceleration.