The Journal of the Astronautical Sciences, Vol. XXVII, No. 1, pp. 63–84, January–March, 1979

On the Maximization of Orbital Momentum and Energy Using Solar Radiation Pressure[1]

J. C. Van der Ha[2] and V. J. Modi[3]

Abstract

This study is concerned with determination of the time-dependent optimal orientation of a solar sail for maximum increase in the total energy (semi-major axis) and angular momentum (semi-latus rectum) in one revolution of the spacecraft around the sun. The solutions are found in an implicit form involving the state and adjoint variables by means of Pontryagin's 'maximum principle'. Explicit approximate representations of the optimal control strategy are obtained in terms of asymptotic series in the small parameter denoting the ratio of solar radiation to gravity forces. Accuracy of the analytical approximations is assessed through a comparison with results obtained by a numerical iteration technique based on the steepest-ascent method. No restrictions are placed on the eccentricity of the initial and the ensuing osculating orbits, nor on the position of the spacecraft in the starting orbit. The analysis should prove useful in the design of a transfer mission to the distant planets or for reaching an escape trajectory out of the solar system.

Nomenclature

a	semi-major axis
a_e	semi-major axis of earth's orbit, 1 A.U. = 1.496×10^8 km
c	velocity of light, 2.998×10^8 m/s
e	eccentricity

[1] The investigation reported here was supported by the National Research Council of Canada, Grant Number A-2181. Presented at the 1977 AAS/AIAA Astrodynamics Conference, Jackson Lake Lodge, Jackson, Wyoming, September 7-9, 1977.

[2] Graduate Research Fellow, Department of Mechanical Engineering, The University of British Columbia, Vancouver, B.C., Canada, V6T 1W5.

[3] Professor, Department of Mechanical Engineering, The University of British Columbia, Vancouver, B.C., Canada, V6T 1W5.

i	orbital inclination with respect to the ecliptic
ℓ	semi-latus rectum
m	total mass of spacecraft
p	auxiliary orbital element, $e \cos \tilde{\omega}$
q	auxiliary orbital element, $e \sin \tilde{\omega}$
$\mathbf{r}$	radius vector pointing from the sun to spacecraft
t	time
u	inverse radius, $1/r$
$\mathbf{u}^n = (u_1, u_2, u_3)$	unit vector normal to the solar sail with components along the ξ_0, η_0 and ζ_0 axes, Eq. (1)
$\mathbf{u}^s$	unit vector along solar radiation, i.e., radial direction
A	total effective illuminated surface area of spacecraft
C_s	solar constant, 1.35 kW/m^2
E	total energy, $-1/a$
$\mathbf{F} = (F_1, F_2, F_3)$	nondimensional solar radiation force, Eq. (2) with components along the ξ_0, η_0, ζ_0 axes, Eq. (4)
H, H_ℓ	Hamiltonians in maximization of E [Eq. (11)] and L [Eq. (26)], respectively
L	auxiliary variable, $ln(\ell)$
P, Q	auxiliary functions, Eqs. (17)
P_ℓ, Q_ℓ	auxiliary functions, Eqs. (32)
R, S, T	functions of control angles, Eqs. (4)
S'	solar radiation pressure, C_s/c 4.51 X 10^{-6} N/m^2
$W = (W_1, W_2, W_3)$	rotation vector describing motion of the local ξ_0, η_0, ζ_0 frame in the inertial X, Y, Z frame, Eqs. (5), with components along the ξ_0, η_0, ζ_0 axes
X, Y, Z	inertial frame of reference
$\tilde{\alpha} = (\alpha, \beta, \gamma)$	control angles describing arbitrary orientation of the solar sail
ϵ_s	ratio of solar radiation and gravity forces, Eq. (3)
θ	true anomaly, designating the position of the spacecraft in the instantaneous osculating plane, $\nu - \tilde{\omega}$
$\boldsymbol{\lambda} = (\lambda_0, \lambda_1, \lambda_2, \lambda_3)$	vector of adjoint variables, λ_0 belongs to E, λ_1 to L, λ_2 to Φ and λ_3 to Ψ
μ_s	sun's gravitational parameter, 1.326 X 10^{20} m^3/s^2
ν	independent variable, Eq. (7)
ξ, η, ζ	reference frame fixed to the solar sail after rotations α and β, Fig. 2b
ξ_0, η_0, ζ_0	local coordinate axes, along radial, circumferential and orbit-normal directions, respectively, Fig. 2b
ξ_1, η_1, ζ_1	intermediate frame fixed to the solar sail after rotation by α, Fig. 2b

ρ	material parameter characterizing (specular) reflectivity of the solar sail
σ	material parameter $(1\text{-}\tau\text{-}\rho)/2$
τ	material parameter indicating the transmissivity of the solar sail
ϕ	position angle of satellite as measured from the line of nodes, Fig. 2a
ψ	position of the line of nodes measured from the reference line $\nu = 2\pi k$ in the osculating plane, Fig. 2a
$\tilde{\omega}$	position of the perigee measured from the reference line $\nu = 2\pi k$ in the osculating plane, Fig. 2a
Φ, Ψ	auxiliary elements, $e \cos\theta$ $(p \cos\nu + q \sin\nu)$ and $e \sin\theta$ $(p \sin\nu - q \cos\nu)$, respectively
Φ_0, Ψ_0	auxiliary elements, $p_{00} \cos\nu + q_{00} \sin\nu$ and $p_{00} \sin\nu - q_{00} \cos\nu$, respectively
Ω	longitude of the ascending node, measured from the X axis, Fig. 2a

Dots and primes denote differentiation with respect to time and ν, respectively, and initial conditions are identified by the subscripts 00.

Introduction

The steady flux of solar radiation in outer space offers an extremely attractive mode of propulsion (so-called solar sailing) for interplanetary space probes. The abundance of radiation energy ensures a reliable and unremitting source of motive power. The feasibility of the concept is established through various studies, e.g. by Garwin [1], Tsu [2] and Kiefer [3]. In fact, for some deep-space missions, propulsion by solar radiation forces may well constitute the only foreseeable practical means of transfer. In this regard, the technology assessment of a solar sail mission to Halley's comet seriously undertaken by the Jet Propulsion Laboratory may be mentioned: it was planned to launch a scientific package of approximately 850 kg into a trajectory by means of an 850 m × 850 m aluminized 0.1 mil thick mylar sail (Fig. 1) for rendezvous with the comet. Although the project was finally abandoned, the potential of solar sailing had been clearly established.

Since the magnitude of the thrust generated by the solar radiation pressure is proportional to the effective area/mass ratio of the spacecraft, it is imperative to employ the lightest materials available meeting the structural requirements for the sail and supporting booms. The combination of useful payload and solar sail would in most practical cases, lead to an area/mass ratio in the range of 50 to 200 m^2/kg corresponding to characteristic accelerations between about 0.5 and 2 mm/s^2. A propellant force of this order of magnitude may be small in comparison with that from chemical thrusters, but the latter have a very limited lifetime. It is important,

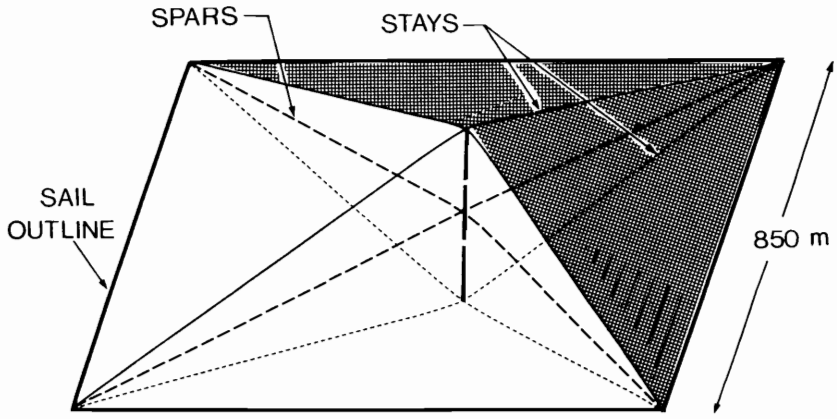

FIG. 1. A Schematic Diagram Showing the Concept of Solar Sail and Its Supports

however, to utilize the available solar radiation power as efficiently as possible in order to limit the duration of the mission to a practical length of time. Since both the magnitude and the direction of the resulting thrust depend upon the orientation of the sail with respect to the radiation, it is interesting to evaluate the time history of the sail setting which would fulfill a certain mission in a prescribed optimal manner.

The literature review suggests that only a few problems of this type have been studied. If the orientation of the sail is kept fixed with respect to the local reference frame, a solution for the resulting trajectory in the shape of a logarithmic spiral is valid provided that the initial velocity vector has a specific magnitude and direction. An expression for the distance as a function of time can be derived for this case and the best (constant) sail setting and corresponding spiral angle maximizing the radial distance can be determined graphically for a given A/m ratio [4]. Recently, an out-of-plane spiral transfer trajectory involving the reversal of the sail orientation with respect to the orbital plane, was presented by the authors [5]: the best sail setting for maximizing the inclination at a prescribed distance (and vice versa) was determined and the response established. Using a numerical Newton-Raphson iteration method, Lebedev et al. [6-8] determined the minimum-time strategy for transfer between two coplanar circular orbits.

In many deep-space missions, e.g. rendezvous with a distant planet or escape from the solar system, it is important to increase the size of the orbit in an effective manner. While the appropriate optimization criterion would depend on the objective of the actual mission under investigation, two specific criteria with a wide range of applicability are selected in the present investigation:

(i) which control strategy leads to the maximum increase in the semi-major axis (and thus total energy) after one revolution?

(ii) which control strategy yields a maximum increase in the semi-latus rectum (and thus angular momentum) after one revolution?

While the control strategy which directs the thrust along the instantaneous velocity vector at all times would likely be very effective in case (i), a formulation in terms of the optimal control theory [9] would evaluate, for instance, the effect of steering the spacecraft relatively closer to the sun initially in order to take advantage of the larger magnitude of the force there. Application of Pontryagin's maximum-principle [10] provides necessary conditions to be satisfied by the optimal control strategy. The solutions are found in an implicit form in terms of the state and adjoint variables and approximate explicit representations can be determined in terms of an asymptotic series in the small parameter denoting the ratio of solar radiation and gravity forces. The approximate analytical results are substantiated by means of a numerical iteration procedure based on the steepest-ascent method [11]. No restrictions are placed on the position of the satellite in the initial orbit nor on the nature of the initial and ensuing osculating ellipses.

Formulation of the Problem

An inertial X, Y, Z reference frame with origin at the center of the sun is introduced in Fig. 2a where the X axis points towards the initial position of the spacecraft and the X, Y plane coincides with the initial osculating plane, usually the ecliptic, while the Z axis is aligned with the initial angular momentum vector. In addition, a local ξ_0, η_0, ζ_0 frame moving with the spacecraft is introduced: the ξ_0, η_0 and ζ_0 axes point along the radial, circumferential and orbit-normal directions,

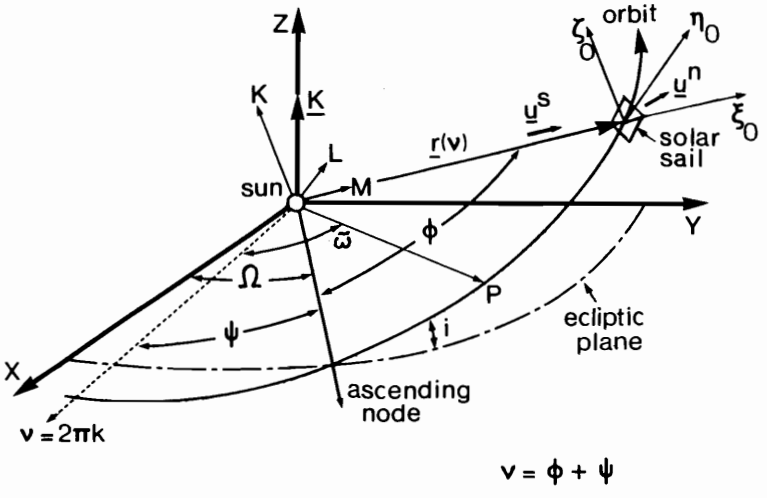

FIG. 2a. Configuration of Sun and Solar Sail in Heliocentric Trajectory

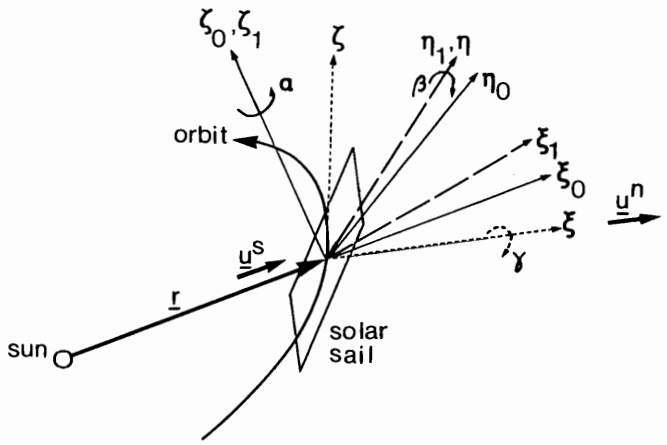

FIG. 2b. Successive Rotations α, β (and γ) for Defining Arbitrary Orientation of Solar Sail

respectively. Any orientation of the solar sail in the ξ, η, ζ frame is described by the following successive Eulerian rotations (Fig. 2b): taking the outward normal to the sail to be directed, initially, along the ξ_0 axis, a first rotation α along the ζ_0 axis produces the ξ_1, η_1, ζ_1 frame and brings the solar sail to the required line of intersection with the orbital plane. A subsequent rotation β along the η_1 axis yields the ξ, η, ζ frame and moves the normal to the sail out of the orbital plane to its prescribed orientation. A final rotation γ along the normal (ξ axis) could be performed for attaining the proper attitude of the sail in its η, ζ plane without affecting the resulting solar radiation force.

The components of the unit sail normal $\mathbf{u}^n$ taken along the local ξ_0, η_0, ζ_0 axes depend on α and β only:

$$\mathbf{u}^n = (u_1, u_2, u_3) = (\cos\alpha \cos\beta, \sin\alpha \cos\beta, -\beta) \tag{1}$$

In nondimensional form (unit of length equals $a_e = 1$ A.U. and unit of time is $1/(2\pi)$ year), the solar radiation force upon a homogeneous solar sail can be written as,

$$\mathbf{F} = \epsilon_s |u_1| \{\sigma \mathbf{u}^s + \rho u_1 \mathbf{u}^n\}/r^2. \tag{2}$$

Here the physical force is nondimensionalized through multiplication by $a_e^2/(m\mu_s)$. The component containing the parameter σ is due to the absorbed radiation and is directed along the vector $\mathbf{u}^s$ while the term with ρ is due to the specularly reflected photons. The absolute sign around $u_1 = (\mathbf{u}^n \cdot \mathbf{u}^s)$ is necessary to ensure that the

force has a non-negative component along the direction of the radiation $\mathbf{u}^s$. The small parameter ϵ_s denotes the ratio of solar radiation over gravity forces,

$$\epsilon_s = 2S' \text{ (A/m)} \, (a_e^2/\mu_s) = 1.52 \times 10^{-3} \text{ (A/m)} \tag{3}$$

It may be noted that $S' = C_s/c$ where the solar constant C_s and thus the force in Eq. (2) varies proportionally to the inverse square of the distance from the sun. This is because the total radiant energy emitted by the sun equals the amount passing through any concentric spherical surface around the sun per unit time (taking the rate of energy output constant). Components of the solar radiation force in the local ξ_0, η_0, ζ_0 frame can be expressed in terms of components of $\mathbf{u}^n$ as:

$$\begin{aligned} F_1 &= \epsilon_s |u_1| \, (\sigma + \rho u_1^2)/r^2 \equiv \epsilon_s \, R/r^2 , \\ F_2 &= \epsilon_s \rho |u_1| \, u_1 \, u_2/r^2 \equiv \epsilon_s \, S/r^2 , \\ F_3 &= \epsilon_s \rho |u_1| \, u_1 \, u_3/r^2 \equiv \epsilon_s \, T/r^2 , \end{aligned} \tag{4}$$

with R, S and T defined in an obvious manner.

In general, when $T \neq 0$, the plane of the orbit will shift. The motion of the ξ_0, η_0, ζ_0 frame relative to the inertial X, Y, Z frame may be described in terms of the rotation vector $\mathbf{W}$ with components W_1, W_2, W_3 along the instantaneous ξ_0, η_0, ζ_0 axes:

$$\mathbf{W} = \begin{bmatrix} W_1 \\ W_2 \\ W_3 \end{bmatrix} = \begin{bmatrix} \dot{\Omega} \sin \phi \sin (i) + (\dot{i}) \cos \phi \\ \dot{\Omega} \cos \phi \sin (i) - (\dot{i}) \sin \phi \\ \dot{\nu} = \dot{\phi} + \dot{\Omega} \cos (i) \end{bmatrix} \tag{5}$$

Since the velocity vector $\dot{\mathbf{r}}$ lies in the instantaneous orbital (ξ_0, η_0) plane (condition of osculation), the vector $\mathbf{W} \times \mathbf{r}$ cannot have a component along the ζ_0 axis, implying that $W_2 \equiv 0$. The motion of the ξ_0, η_0, ζ_0 reference frame can be visualized as the sum of the rotations $W_3 = \dot{\nu}$ along the normal and W_1 along the radial directions, the latter component solely describing the shift in the orientation of the orbital plane. The equations of motion follow from Newton's second law, accounting for rotation, with scalar components along the ξ_0, η_0, ζ_0 axes given as:

$$\begin{aligned} \ddot{r} + 1/r^2 - r\dot{\nu}^2 &= \epsilon_s R/r^2 , \\ r\ddot{\nu} + 2\dot{r}\dot{\nu} &= \epsilon_s S/r^2 , \\ r W_1 \dot{\nu} &= \epsilon_s T/r^2 . \end{aligned} \tag{6}$$

The characteristic feature of this formulation is the fact that the in-plane perturbations appear uncoupled from the out-of-plane force component. It is advantageous to employ the quasi-angle ν, Eq. (5), as the independent variable. Its dependence on time is given by the equation

$$\dot{\nu} = \dot{\phi} + \dot{\Omega} \cos(i) = \ell^{1/2}/r^2, \tag{7}$$

which can be derived from angular momentum considerations ($h = \ell^{1/2}$) in conjunction with the last relation of Eqs. (5). As in the case of the classical Keplerian solution, the in-plane equations can be transformed in terms of $u = 1/r$ and ℓ:

$$\begin{aligned} u''(\nu) + u(\nu) &= (1 - \epsilon_s R)/\ell - \epsilon_s S u'/(u\ell); \\ \ell'(\nu) &= 2\,\epsilon_s\, S/u. \end{aligned} \tag{8}$$

The physical interpretation is familiar: the solar radiation force produces perturbations in the elements of the Keplerian ellipse emerging from Eqs. (8) when $\epsilon_s = 0$.

To have a uniformly valid analysis for all eccentricities including $e = 0$ where the argument of the perigee becomes indeterminate, the inverse radius is written as $u(\nu) = (1 + p \cos \nu + q \sin \nu)/\ell$ with slowly varying elements p, q and ℓ. The auxiliary orbital parameters p and q can be interpreted as the Cartesian components of the eccentricity vector $\mathbf{e}$ (pointing from the origin to the perigee) along and normal to the line $\nu = 2\pi k$ in the instantaneous orbital plane; thus $p = e \cos \tilde{\omega}$, $q = e \sin \tilde{\omega}$. The second-order equation for $u(\nu)$ is transformed into two first-order autonomous equations in terms of $\Phi(\nu) = p \cos \nu + q \sin \nu$ and $\Psi(\nu) = p \sin \nu - q \cos \nu$:

$$\begin{aligned} \Phi'(\nu) &= -\Psi + 2\epsilon_s S; \\ \Psi'(\nu) &= \Phi + \epsilon_s \{R + S\,\Psi/(1 + \Phi)\}. \end{aligned} \tag{9}$$

With these preliminaries, one may proceed to evolve an optimal control strategy with the help of Pontryagin's maximum principle. The equations involved are considerably simplified if two new state variables $E = -1/a$ (i.e., total energy) and $L = ln(\ell)$ are introduced. Defining a set of adjoint variables λ_0, λ_1, λ_2 and λ_3 corresponding to the state variables E, L, Φ and Ψ, respectively, the complete system of in-plane equations can be written as:

$$E'(\nu) = 2\epsilon_s \{R(\tilde{\alpha})\,\Psi + S(\tilde{\alpha})\,(1 + \Phi)\} \exp(-L), \quad E(0) = -1/a_{00};$$

$$L'(\nu) = 2\epsilon_s\, S(\tilde{\alpha})/(1 + \Phi), \qquad L(0) = ln(\ell_{00}); \tag{10}$$

$$\Phi'(\nu) = -\Psi + 2\epsilon_s S(\tilde{\alpha}), \qquad \Phi(0) = \Phi_{00} = e_{00} \cos \tilde{\omega}_{00};$$

$$\Psi'(\nu) = \Phi + \epsilon_s \{R(\tilde{\alpha}) + S(\tilde{\alpha}) \Psi/(1 + \Phi)\}, \qquad \Psi(0) = \Psi_{00} = -e_{00} \sin \tilde{\omega}_{00};$$

$$\lambda_0'(\nu) = 0;$$

$$\lambda_1'(\nu) = 2\epsilon_s \lambda_0 \{R(\tilde{\alpha}) \Psi + S(\tilde{\alpha}) (1 + \Phi)\} \exp(-L); \tag{10}$$

(*Cont.*)

$$\lambda_2'(\nu) = -\lambda_3 - 2\epsilon_s \lambda_0 S(\tilde{\alpha}) \exp(-L) + \epsilon_s S(\tilde{\alpha}) [2\lambda_1 + \lambda_3 \Psi]/(1 + \Phi)^2$$

$$\lambda_3'(\nu) = \lambda_2 - 2\epsilon_s \lambda_0 R(\tilde{\alpha}) \exp(-L) - \epsilon_s \lambda_3 S(\tilde{\alpha})/(1 + \Phi).$$

The behavior of the orbital plane can be described by the usual equations for $(\dot{i})$ and $\dot{\Omega}$ which follow readily from Eqs. (5) and the results for W_1 in Eqs. (6). However, since the out-of-plane motion turns out to be irrelevant for the present investigation they are disregarded here.

Maximization of Total Energy

In this section, an approximate analytical representation for the optimal control strategy $\tilde{\alpha}(\nu)$ maximizing the total energy E (and thus major axis a) at $\nu = 2\pi$ is derived. The Hamiltonian for the present problem, Eqs. (10), becomes:

$$H(\tilde{\alpha}) = \lambda_3 \Phi - \lambda_2 \Psi + \epsilon_s R(\tilde{\alpha}) \{2\lambda_0 \Psi \exp(-L) + \lambda_3\}$$

$$+ \epsilon_s S(\tilde{\alpha}) \{2\lambda_0 (1 + \Phi) \exp(-L) + (2\lambda_1 + \lambda_3 \Psi)/(1 + \Phi) + 2\lambda_2\}. \tag{11}$$

For $\tilde{\alpha}(\nu)$ to be an optimal control vector over the fixed interval $(0, 2\pi)$, the following necessary conditions must be satisified:

(i) $\partial H/\partial\alpha = \partial H/\partial\beta = 0$;

(ii) $H(\tilde{\alpha}) = \text{constant}$;

(iii) $\lambda_j(2\pi) = 0, \quad j = 1, 2, 3$; (transversality)

(iv) $\lambda_0(\nu) = 1$; (12)

according to Pontryagin's Maximum Principle [8, 9]. From the conditions (ii), (iii) and (iv) it follows that

$$H = 2\epsilon_s \{[\Psi R(\tilde{\alpha}) + (1 + \Phi) S(\tilde{\alpha})] \exp(-L)\}_{\nu=2\pi}, \tag{13}$$

which equals $E'(2\pi)$, Eqs. (10). The conditions in (i) lead to the following equation for α,

$$\alpha = [2\,\Psi \exp(-L) + \lambda_3]\,\frac{\partial R}{\partial \alpha} + [2(1+\Phi)\exp(-L) + (2\lambda_1 + \lambda_3\Psi)/(1+\Phi) + 2\lambda_2]\,\frac{\partial S}{\partial \alpha}, \tag{14}$$

and a similar one for β. It follows readily that $\beta(\nu) = 0$ is a solution for the out-of-plane rotation confirming that the optimal trajectory is a planar one since the solar radiation force remains in the plane of the orbit. The equation for the control angle $\alpha(\nu)$ is reduced to the following implicit relation:

$$\frac{\rho \cos\alpha\,(1 - 3\sin^2\alpha)}{\sin\alpha\,(\sigma + 3\,\rho\cos^2\alpha)} = \frac{2\,\Psi + \ell\,\lambda_3}{2(1+\Phi) + (2\lambda_1 + \lambda_3\Phi)\,\ell/(1+\Phi) + 2\ell\,\lambda_2}, \tag{15}$$

with α in the interval $(0, \pi/2)$ on physical grounds. For obtaining approximate solutions for $\alpha(\nu)$ from Eq. (15) it is imperative to assess, carefully, the orders of magnitude of the various terms on the right-hand-side of Eq. (15). Thereto, the orbital elements and adjoint variables are written as a system of coupled integral equations derived from Eqs. (10) by integration while taking the mixed boundary conditions into account:

$$\begin{aligned}
a(\nu) &= a_{00} + \epsilon_s \int_0^\nu \{a^2\,[R\,\Psi + S(1+\Phi)]/\ell\}\,d\tau;\\
\ell(\nu) &= \ell_{00} + 2\epsilon_s \int_0^\nu \{\ell\,S/(1+\Phi)\}\,d\tau;\\
\Phi(\nu) &= \Phi_0(\nu) + \epsilon_s \int_0^\nu \{2\,S\cos(\tau-\nu) + [R + S\,\Psi/(1+\Phi)]\sin(\tau-\nu)\}\,d\tau;\\
\Psi(\nu) &= \Psi_0(\nu) + \epsilon_s \int_0^\nu \{[R + S\,\Psi/(1+\Phi)]\cos(\tau-\nu) - 2S\sin(\tau-\nu)\}\,d\tau;\\
\lambda_1(\nu) &= -2\epsilon_s \int_\nu^{2\pi} \{[R\,\Psi + S(1+\Phi)]/\ell\}\,d\tau;
\end{aligned} \tag{16}$$

$$\lambda_2(\nu) = \epsilon_s \int_{\nu}^{2\pi} \{Q \sin(\tau - \nu) + P \cos(\tau - \nu)\}\, d\tau;$$

$$\lambda_3(\nu) = \epsilon_s \int_{\nu}^{2\pi} \{Q \cos(\tau - \nu) - P \sin(\tau - \nu)\}\, d\tau. \qquad (16)\ (Cont.)$$

Here $\Phi_0(\nu) = p_{00} \cos \nu + q_{00} \sin \nu$ and $\Psi_0(\nu) = p_{00} \sin \nu - q_{00} \cos \nu$ and the auxiliary functions P and Q stand for:

$$P = 2\, S/\ell - S(2\lambda_1 + \lambda_3\, \Psi)/(1 + \Phi)^2;$$

$$Q = 2\, R/\ell + \lambda_3\, S/(1 + \Phi). \qquad (17)$$

An asymptotic series for $\alpha(\nu)$ in terms of the small parameter ϵ_s can now be constructed. By writing $\alpha(\nu) = \alpha_0(\nu) + \epsilon_s\, \alpha_1(\nu) + 0(\epsilon_s^2)$, developing the left-hand-side of Eq. (15) in a Taylor series around α_0 and expanding the right-hand-side using results in Eqs. (16), successive terms in the series for $\alpha(\nu)$ can be established. The leading term satisfies the implicit relation:

$$\frac{\rho \cos \alpha_0\, (1 - 3 \sin^2 \alpha_0)}{\sin \alpha_0\, (\sigma + 3\rho \cos^2 \alpha_0)} = \frac{\Psi_0(\nu)}{1 + \Phi_0(\nu)}. \qquad (18)$$

A good approximation to the solution of Eq. (18) may be obtained by successive substitution with a starting value $\alpha_0^{(1)}(\nu) = 35.26°$, (i.e., the solution of Eq. (18) for an initially circular orbit). The $(n+1)$th approximation is obtained from $\alpha_0^{(n)}(\nu)$ as follows:

$$\alpha_0^{(n+1)}(\nu) = \arcsin\left[\{1/3 - \Psi_0 \tan \alpha_0^{(n)}\, [(\sigma/3\rho) + \cos^2 \alpha_0^{(n)}]/(1 + \Phi_0)\}^{1/2}\right], \qquad (19)$$

$n = 1, 2, 3, \ldots$, which converges rapidly provided that the initial eccentricity is not too large. Geometrically, the steering angle $\alpha_0(\nu)$ in Eq. (19) makes the resulting solar radiation force aligned with the velocity vector of the unperturbed initial osculating ellipse at each instant.

Whereas this may serve as a useful guide for very small values of ϵ_s, it is evident that higher-order terms relating to the slowly varying geometry of the osculating ellipse must be evaluated when practical values of ϵ_s are taken. For the analytical evaluation of the higher-order terms, an explicit relation for $\alpha_0(\nu)$ would be needed. In the special case when the reflection is specular ($\rho = 1$, $\sigma = 0$), a closed-form result for $\alpha_0(\nu)$ can be derived from Eq. (18),

$$\alpha_0(\nu) = \frac{1}{2}\arcsin\left[\frac{(1+\Phi_0)\{[9\,\Psi_0^2 + 8(1+\Phi_0)^2]^{1/2} - \Psi_0\}}{3[\Psi_0^2 + (1+\Phi_0)^2]}\right] \tag{20}$$

On expanding both sides of Eq. (15) as a Taylor series in terms of the small parameter ϵ_s, the first-order term $\alpha_1(\nu)$ becomes now:

$$\alpha_1(\nu) = -\frac{3}{2}\frac{\sin^2(2\alpha_0)}{[3-\cos(2\alpha_0)]}\Big\{\Psi^{(1)} + \ell_{00}\,\lambda_3^{(1)}/2 - \Psi_0\,\ell_{00}[\lambda_1^{(1)} + \Psi_0\,\lambda_3^{(1)}/2]/(1+\Phi_0)^2 - \ell_{00}\,\Psi_0\,\lambda_2^{(1)}/(1+\Phi_0)\Big\}, \tag{21}$$

where the superscript (1) denotes the coefficient of ϵ_s in the expressions in Eqs. (16). The trigonometric terms in Eq. (21) can be eliminated in favor of the orbital variables Φ_0 and Ψ_0 through Eq. (20). Also the integrands in Eqs. (16) can be expanded for small ϵ_s and expressed in terms of ν. Whereas the resulting integrals are unwieldly for arbitrary eccentricity, analytical results can be obtained for near-circular orbits. Thereto, expansions of the trigonometric terms for small e_{00} are needed. These can be derived using the expansion of Eq. (20) for small e_{00}:

$$3\sin[2\alpha_0(\nu)] = 2^{3/2} - \Psi_0(\nu)\{1 - \Phi_0(\nu) + 7(2^{-5/2})\,\Psi_0(\nu) + \Phi_0(\nu)[\Psi_0(\nu) - 7(2^{-3/2})\,\Psi_0(\nu)]\} + 0(e_{00}^4). \tag{22}$$

After developing $R(\alpha)$ and $S(\alpha)$ around $\alpha = \alpha_0$, e.g., $R(\alpha) = R_0(\nu) + 0(\epsilon_s)$, the following near-circular approximations are found:

$$\begin{aligned} R_0(\nu) &= (2/3)^{3/2} + \Psi_0(\nu)\{3^{1/2} - \Psi_0(\nu)[3^{1/2}\Phi_0(\nu) + \Psi_0(\nu)/8]\} + 0(e_{00}^3); \\ S_0(\nu) &= 2(3^{-3/2}) - \Psi_0^2(\nu)/6 + 0(e_{00}^3). \end{aligned} \tag{23}$$

With these results, all integrands in Eqs. (16) can be evaluated and near-circular approximations for $\Psi^{(1)}$, $\lambda_3^{(1)}$, etc., in Eq. (21) are obtained by integration. Finally, the following expression for $\alpha_1(\nu)$ with an error of the order e_{00}^3 is established:

$$\begin{aligned} \alpha_1(\nu) = {} & -3^{-3/2}\{1 - \cos\nu + 3\pi q/2 + (4\pi - 3\nu/2)\Psi_0 - (p + \Phi_0)\cos\nu \\ & + 2p - 3q/2\sin\nu - 9(2^{-3/2})\,\Psi_0\,(1 - \cos\nu)\} \\ & - \nu\{7(3^{1/2})\,(p^2 - q^2)\sin(2\nu) - 2(6^{1/2})\,e^2\} \\ & - \pi\{4(6^{1/2})\,e^2 + 3^{3/2}\,pq\}/18 \\ & - \sin(2\nu)\{(p^2 - q^2)[6^{1/2} - 4(3^{3/2})\pi] - pq/2\}/18 \end{aligned}$$

$$
\begin{aligned}
&- [1 - \cos(2\nu)]\ \{(p^2 - q^2)(3^{-1/2} + 1/4) + 2(6^{1/2})pq\}/18 \\
&- (1 - \cos\nu)\ \{(p^2 - q^2)[2(3^{-1/2} - 1/4 + 4\sin(2\nu)] - 3e^2/4 - (3^{3/2})\pi pq\}/18 \\
&- \sin\nu\ \{pq - 3^{3/2}(2p^2 + q^2)\pi + 3^{1/2}[4pq\cos\nu + (q^2 - 3p^2)\sin\nu] \\
&\quad - 4(3^{1/2})(p^2 - q^2)\sin(2\nu) + 6^{1/2}[e^2 - (p^2 - q^2)\cos(2\nu) \\
&\quad - 2pq\sin(2\nu)]\}/18 \\
&- 6^{1/2}\ \{3\pi q\Psi_0 + (8\pi - 3\nu)\Psi_0^2 - 2(p + \Phi_0)\Psi_0\cos\nu + 4p\Psi_0 \\
&\quad - 3q\Psi_0\sin\nu - \Psi_0\ [2\Phi_0 + 3(2^{-1/2})\Psi_0]\ (1 - \cos\nu)\}/144 + 0(e_{00}^3). \qquad (24)
\end{aligned}
$$

Here the subscripts 00 are omitted for brevity. It follows that the first-order correction $\epsilon_s\alpha_1(\nu)$ for an initially circular orbit is at most $22\epsilon_s$ degrees (at $\nu = \pi$) below the constant $\alpha_0 = 35.26°$ control program. It is interesting to evaluate the response of the major axis under the optimal control strategy. For a near-circular orbit, $a(\nu)$ can be written as

$$
\begin{aligned}
a(\nu) = a_{00}\ \exp\ &\{2\epsilon_s(1 + e_{00}^2)\ [2(3^{-3/2})\ (\nu + p_{00}\sin\nu + q_{00} - q_{00}\cos\nu) \\
&+ 2(6^{1/2})\ (p_{00} - p_{00}\cos\nu - q_{00}\sin\nu)/9 + e_{00}^2\nu/2 \\
&+ (q_{00}^2 - p_{00}^2)\sin(2\nu)/4 + p_{00}q_{00}\cos(2\nu) - p_{00}q_{00} + 0(e_{00}^3)] \\
&+ 0(\epsilon_s^2)\}. \qquad (25)
\end{aligned}
$$

If $e_{00} = 0$ this result can be reduced considerably yielding $a(2\pi) = a_{00}\ \exp\ [4.8368\ \epsilon_s + 0(\epsilon_s^2)]$ after one revolution.

Maximization of Angular Momentum

Here, the optimal control strategy for maximum increase in angular momentum (and thus semi-latus rectum) per revolution is determined. This corresponds with maximization of $L(2\pi)$. The system of Eqs. (10) remains valid provided that the equations for E and λ_0 are ignored and the equation for λ_1 is replaced by $\lambda_1'(\nu) = 0$. Now the Hamiltonian becomes

$$
H_{\mathrm{Q}}(\tilde{\alpha}) = \lambda_3\Phi - \lambda_2\Psi + \epsilon_s\,\lambda_3\,R(\tilde{\alpha}) + \epsilon_s\,S(\tilde{\alpha})\ \{(2\lambda_1 + \lambda_3\Psi)/(1 + \Phi) + 2\lambda_2\}. \qquad (26)
$$

Application of Pontryagin's maximum principle leads to results as in Eqs. (12) with $\lambda_1 = 1$ now. It follows that $H_{\mathrm{Q}} = L'(2\pi)$ and the out-of-plane rotation $\beta(\nu) = 0$ while the optimal control angle $\alpha(\nu)$ is given by the implicit relation,

$$\frac{\rho \cos \alpha (1 - 3 \sin^2 \alpha)}{\sin \alpha (\sigma + 3\rho \cos^2 \alpha)} = \frac{\lambda_3 (1 + \Phi)}{2 + \lambda_3 \Psi + 2 \lambda_2 (1 + \Phi)} \tag{27}$$

The orbital elements ℓ, Φ and Ψ can be written in the form of Eqs. (16) while the adjoint variables $\lambda_2(\nu)$ and $\lambda_3(\nu)$ become

$$\begin{aligned} \lambda_2(\nu) &= \epsilon_s \int_{\nu}^{2\pi} \{Q_\ell \sin(\tau - \nu) + P_\ell \cos(\tau - \nu)\} \, d\tau, \\ \lambda_3(\nu) &= \epsilon_s \int_{\nu}^{2\pi} \{Q_\ell \cos(\tau - \nu) - P_\ell \sin(\tau - \nu)\} \, d\tau, \end{aligned} \tag{28}$$

with P_ℓ and Q_ℓ defined by

$$\begin{aligned} P_\ell &= -S(2 + \lambda_3 \Psi)/(1 + \Phi)^2, \\ Q_\ell &= \lambda_3 \, S/(1 + \Phi). \end{aligned} \tag{29}$$

The right-hand-side of Eqs. (27) is of the order ϵ_s so that $\alpha(\nu)$ is written as $\alpha(\nu) = \alpha_0 + \epsilon_s \, \alpha_1 + 0(\epsilon_s^2)$ with $\alpha_0 = \arcsin(3^{-1/2}) = 35.26°$. The first-order term $\alpha_1(\nu)$ is determined by expanding both sides of Eqs. (27) in Taylor series for small ϵ_s yielding the following explicit result,

$$\begin{aligned} \alpha_1(\nu) = &-(\sigma + 2\rho) \, 3^{-3/2}/2 \, [1 + \Phi_0(\nu)] \\ &\cdot \left[2 \, \Psi_0(\nu) \left\{ \pi - \arctan\left[\frac{(1 - e_{00}^2)^{1/2} \tan(\nu/2)}{1 + p_{00} + q_{00} \tan(\nu/2)} \right] \right\} /(1 - e_{00}^2)^{1/2} \right. \\ &\left. + 1 - [\Phi_0(\nu) + \cos \nu]/(1 + p_{00}) \right] /(1 - e_{00}^2). \end{aligned} \tag{30}$$

The resulting response $\ell(\nu)$ under the optimal sail setting can be approximated by integrating $L'(\nu)$ in Eqs. (10) (up to order ϵ_s),

$$\ell(\nu) = \ell_{00} \exp \left[8\epsilon_s \rho \, 3^{-3/2} \arctan \frac{(1 - e_{00})^{1/2} \tan(\nu/2)}{1 + p_{00} + q_{00} \tan(\nu/2)} \, /(1 - e_{00}^2)^{1/2} \right] \tag{31}$$

Considering an initially circular orbit, it follows that $\ell(2\pi) = \ell_{00} \exp\{4.8368 \, \rho\epsilon_s + 0(\epsilon_s^2)\}$. This result is identical to the one found in the previous section while maximizing the semi-major axis for a near-circular starting orbit. Obviously, also the

control programs in Eqs. (24) and (30) are identical for $e_{00} = 0$ in the present approximation.

Discussion of Results

The accuracy of the analytical solution obtained in the section entitled "Maximization of Total Energy" is now assessed by comparison with results from a numerical iteration procedure based upon the steepest-ascent method [11]. An arbitrary nominal control strategy is selected and the influence of a small variation in that control program upon the response is investigated. The variation leading to the maximum increase in major axis under a prescribed step-length (i.e., the integral from 0 to 2π of the square of the variation in the control function) can be determined in terms of the derivatives of the system of Eqs. (10) with respect to the control angle. Thus, a generally more effective new control strategy is obtained and the procedure is repeated. While the algorithm converges rapidly to a near-optimal control strategy, care must be taken in the neighborhood of the optimum due to the weakness of the gradient field. By making both the steplength and the error parameter in the Runge-Kutta integration routine proportional to the length of the gradient, satisfactory results are obtained. In the present case, the initial control program is taken as $\alpha(\nu) = (2\pi - \nu)/6$ and the optimal strategy is established to within, approximately, 0.1 degree in less than 30 iterations (Fig. 3). A relatively small value of the solar parameter (based on $A/m = 10\ m^2/kg$) is taken in this example. The first-order analytical result of Eq. (24) for a near-circular initial orbit in conjunction with the exact zeroth-order term in Eq. (20) yields an extremely close approximation when $e_{00} = 0.2$ (Fig. 3a): the maximum discrepancy is less than 0.1 degree. On the other hand, if $e_{00} = 0.4$ (Fig. 3b), the near-circular analytical solution is in error by almost three degrees around $\nu = 270°$, while still providing a valid representation for the optimal strategy in the remaining portion of the orbit. The breakdown in accuracy must be attributed to two reasons: first, it should be recognized that the first-order analytical result developed here does not contain terms of order e_{00}^3 and higher which are likely to be influential when the eccentricity is as high as 0.4. Secondly, the state and adjoint variables are represented as perturbation series in terms of ϵ_s and only the first-order solutions are taken into account leading to a rapidly growing error when away from the initial and final points.

Figure 4 shows the results for a higher value of ϵ_s, namely $\epsilon_s = 0.09$, corresponding to $A/m = 60\ m^2/kg$. As can be expected, the analytical prediction for the optimal control is most accurate in the case $e_{00} = 0$; the maximum discrepancy of about one degree is due to higher-order (in ϵ_s) effects. It is interesting to note that initially the solar radiation force points slightly inwards from the velocity vector and its magnitude is smaller than that for the case where the force is aligned with the velocity. This is true for both the numerical and the first-order analytical

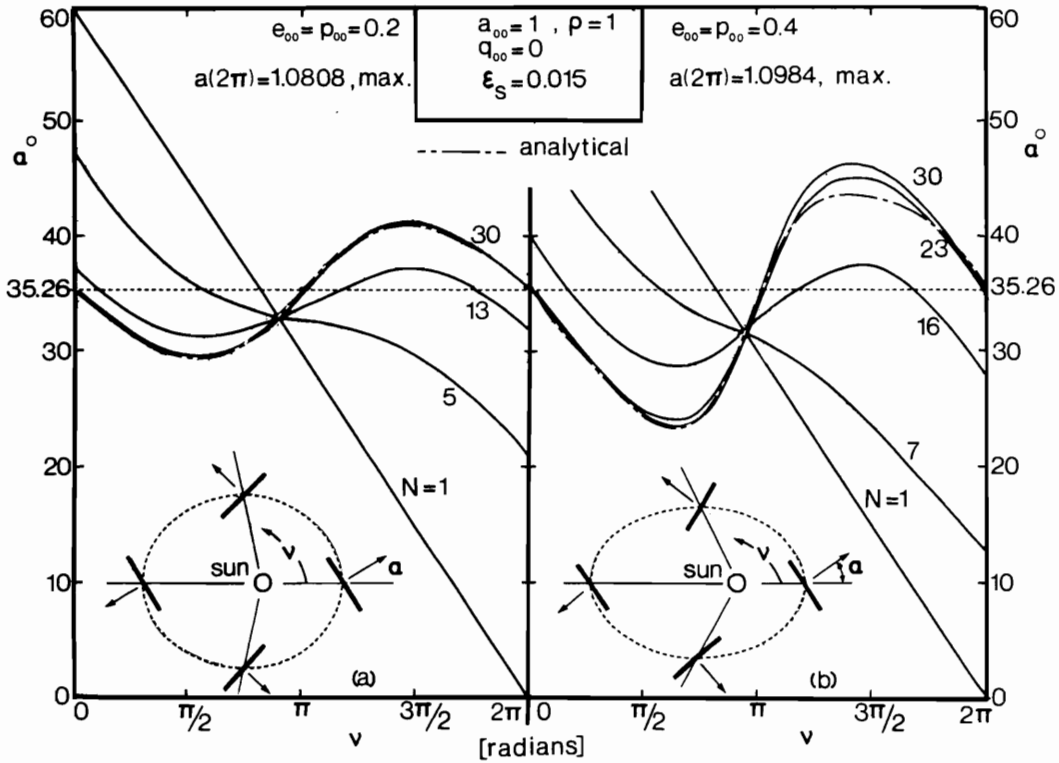

FIG. 3. Comparison of Analytical and Numerical Optimal Strategies for $\epsilon_S = 0.015$: (a) $e_{00} = 0.2$; (b) $e_{00} = 0.4$

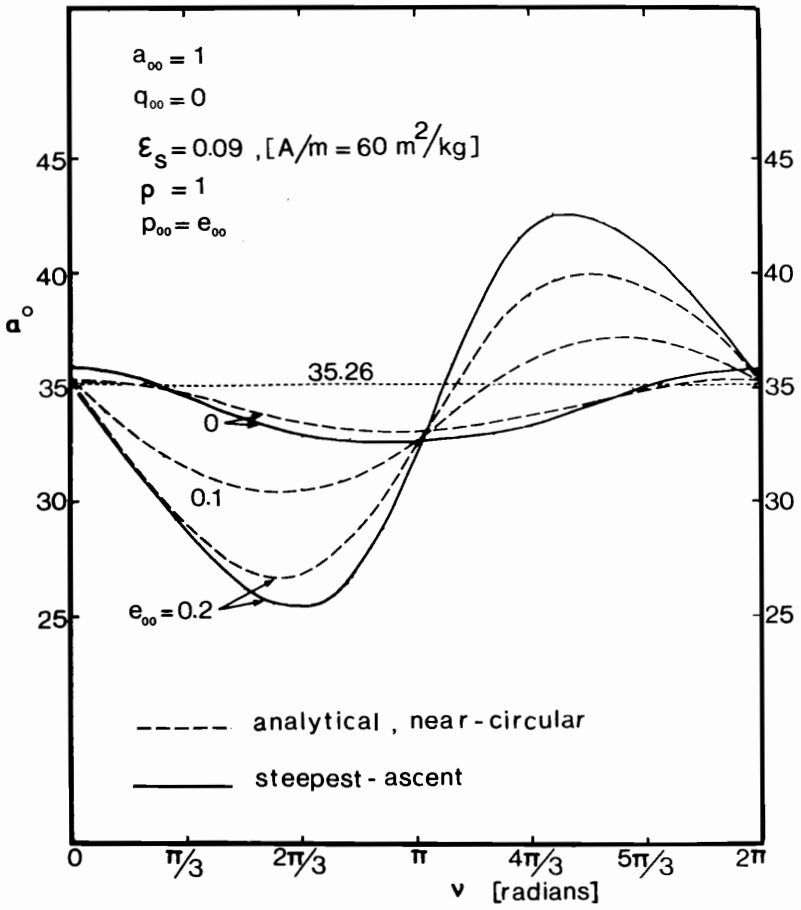

FIG. 4. Optimal Sail Setting for $\epsilon_s = 0.09$, $\omega_{00} = 0$ and a Few Values of e_{00}

results, although the effect is less pronounced in the latter case. This apparent waste of energy is more than recouped during the middle phase of the orbit when the spacecraft is closer to the sun and the force is larger. In this phase, the direction of the force is kept outward from the velocity vector, thus providing an additional boost to its magnitude. In the final phase the force tends to align itself with the velocity. The osculating ellipses corresponding to the resulting trajectory show that the eccentricity increases from 0 to a maximum of about 0.2 near $\nu = 190°$ and decreases to about 0.02 with the position of the perigee at about 70° in the end, $\nu = 2\pi$. The analytical result for $e_{00} = 0.2$ shows a maximum error of about 2.5 degrees as compared to the steepest-ascent solution. Figure 5 shows the optimal steering programs for three different starting points in the same initial orbit of eccentricity $e_{00} = 0.2$ ($\epsilon_s = 0.15$), obtained by the steepest-ascent iteration routine.

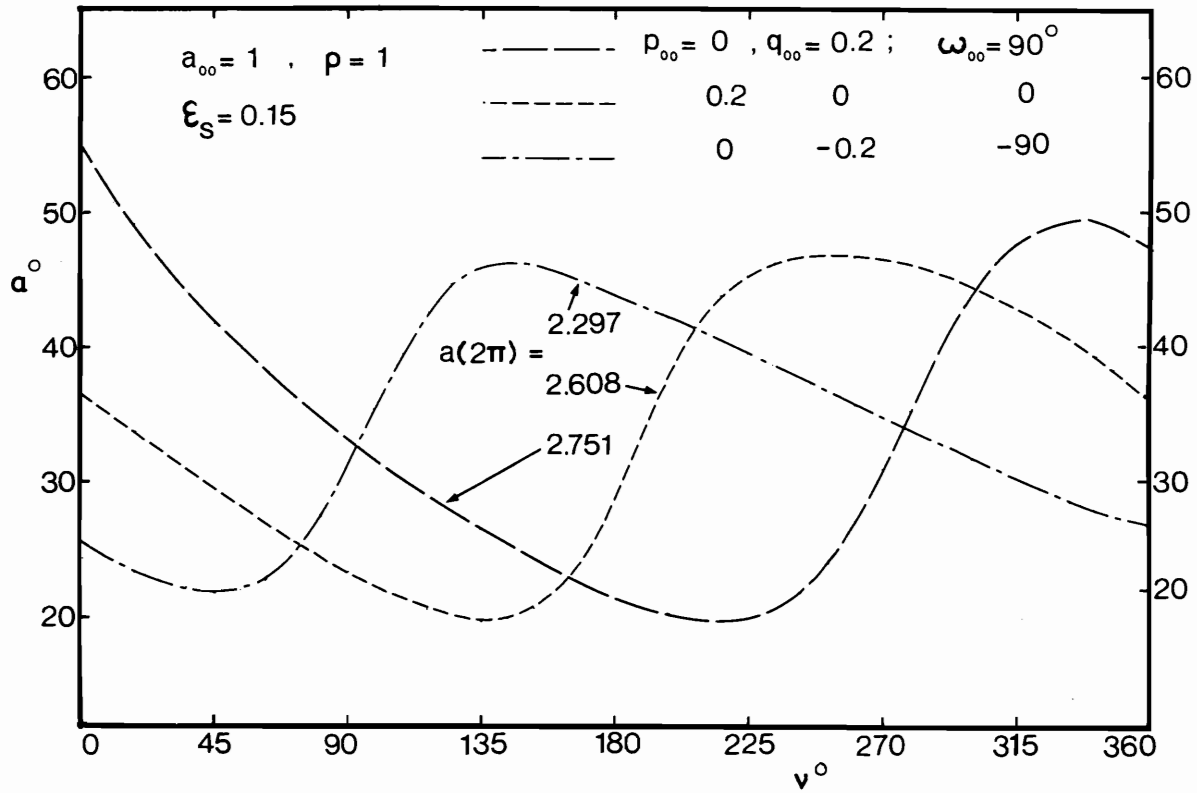

FIG. 5. Optimal Control Programs for $\epsilon_s = 0.15$, $e_{oo} = 0.2$ and a Few Values of ω_{oo}

It is seen that the nature of the control strategy as well as the resulting final value $a(2\pi)$ vary considerably with the position of the starting point.

It is interesting to compare the effectiveness of the optimal strategies with that of other near-optimal control programs, in particular the constant sail setting $\alpha = \arcsin(3^{-1/2}) = 35.26^\circ$. The latter control is expected to be a very effective strategy for small ϵ_s and small e_{00} since it generates the maximum component of the force along the velocity for an unperturbed circular orbit. Table I gives a comparative overview of the response $a(2\pi)$ for a few values of ϵ_s and e_{00} (ω_{00} is taken zero). Although the results seem to be close in most cases, it must be emphasized that a difference of one digit in the fourth decimal place represents a physical distance of about 15,000 km. On the other hand, it is evident that $\alpha(\nu) = 35.26^\circ$ is a very effective control strategy even for eccentricities as high as 0.4. It should be mentioned that the results in Table I are derived numerically, since the analytical prediction for the response under the optimal control, Eq. (25), yields useful values for $\epsilon(2\pi)$ only for small ϵ_s and e_{00} and is not capable of providing accuracy beyond three significant digits in the most favorable case, while being in error by as much as 0.3 in the most severe situation of Table I.

The actual trajectory resulting from the optimal strategy for $\epsilon_s = 0.09$ is depicted in Fig. 6. It is seen that Mars' orbit is intercepted at about $\nu = 135^\circ$ after approximately one year. Also the inward trajectory crossing Venus' orbit is shown. These trajectories are obtained from the steepest-ascent results. It may be mentioned that the leading term in the analytical solution of the optimal strategy for inward trajectories is equal to but opposite in sign compared to the one for the outward ones. The first-order (in ϵ_s) terms, however, are different and can be readily evaluated by taking $\lambda_0 = -1$ rather than $+1$. These conclusions are substantiated by the numerical results.

Finally, the optimal sail settings leading to the maximum increase in angular momentum for a few values of initial eccentricity and solar parameter are shown in

TABLE I. Response $a(2\pi)$ for Optimal Control Strategy and for $\alpha = \arcsin(3^{-1/2})$

e_{00}/ϵ_s	0.015	0.09	0.15
0	1.0761*	1.590	2.280
	1.0760	1.587	2.258
0.2	1.0808	1.668	2.608
	1.0796	1.640	2.454
0.4	1.0984	1.962	4.314
	1.0922	1.819	3.202

*The upper values correspond to the optimal response while the lower ones represent the results for $\alpha = 35.26^\circ$.

Fig. 7. The approximate analytical solution for the present case is likely to be more accurate than the ones before (previous section) due to the fact that $\alpha_1(\nu)$ is obtained for general e_{00}, Eq. (30), leaving only the errors caused by higher-order (in ϵ_s) terms. It may be noted that the resulting optimal control for $e_{00} = 0$ corresponds identically (up to first-order) to the one which maximizes $a(2\pi)$, Fig. 4. Compared to the optimal strategy for maximization of $a(2\pi)$, the present control programs are closer to the 35.26°line, representing the zeroth-order approximation of the optimal control for circular as well as elliptic orbits.

Concluding Remarks

Important aspects of the analysis and conclusions based on them can be summarized as follows:

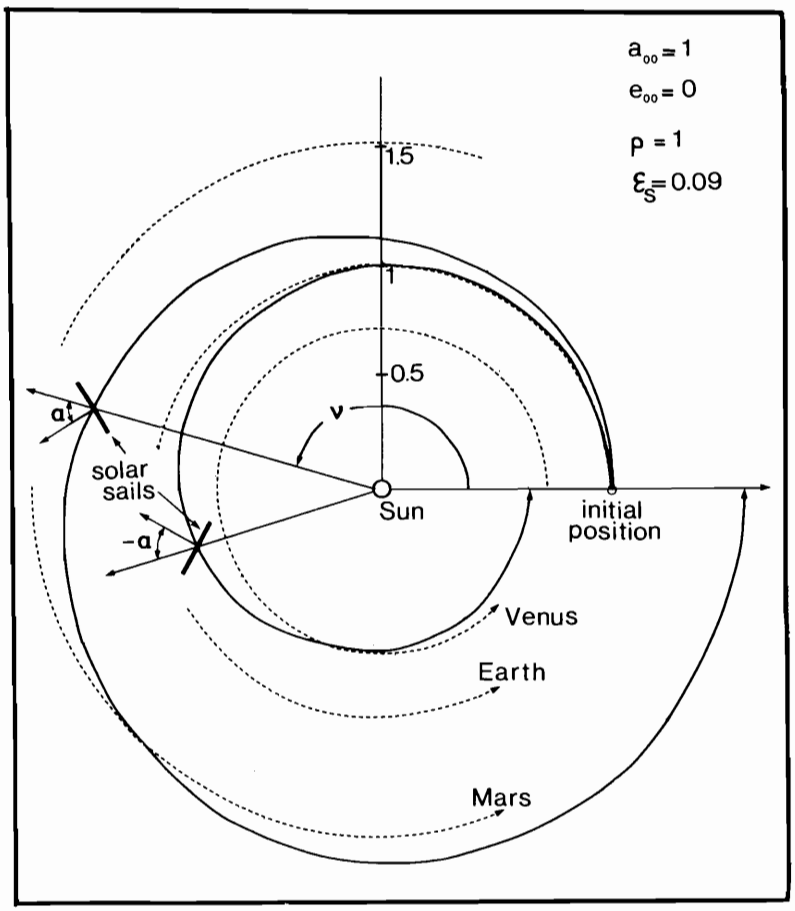

FIG. 6. Actual Trajectory under Optimal Sail Setting, Showing Interception with Mars' and Venus' Orbits

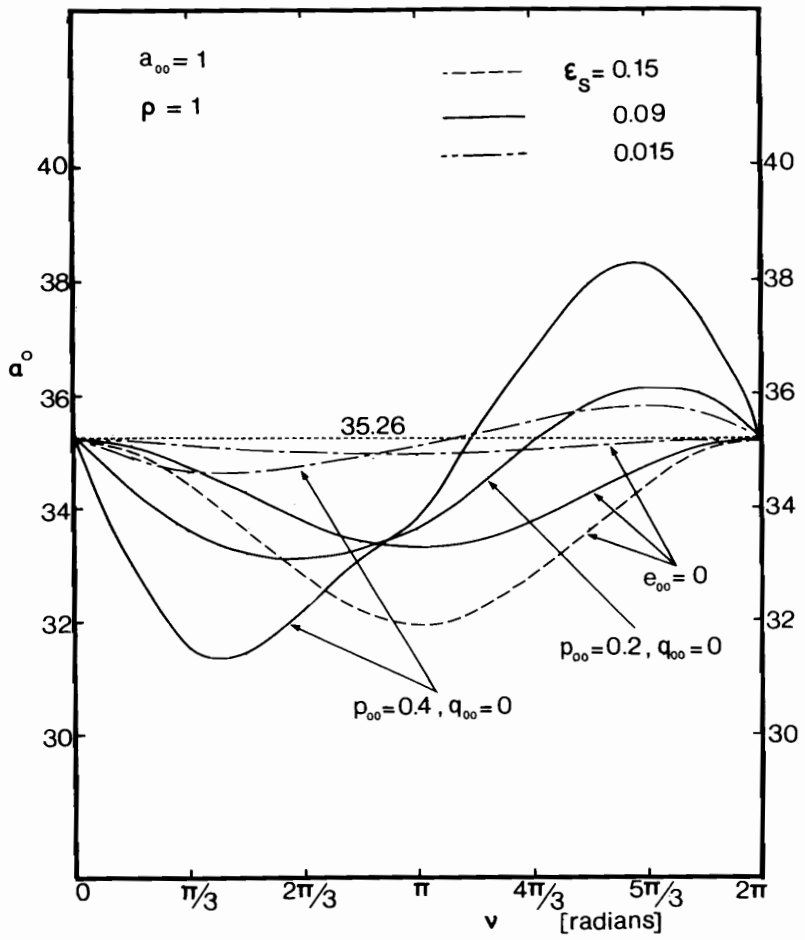

FIG. 7. Control Strategies for Maximization of Angular Momentum

(i) Analytical approximate solutions for the time-dependent optimal sail setting maximizing the total energy (major axis) or the angular momentum (latus rectum) after one revolution are obtained from Pontryagin's Maximum Principle by means of a straightforward perturbation expansion of the state and adjoint variables.

(ii) The validity of the approximate solution is assessed by means of a numerical iteration procedure based upon the steepest-ascent method. In general, the accuracy of the analytical solution decreases with increasing ϵ_s and eccentricity. For values of ϵ_s as high as 0.1 and e up to 0.2, the maximum deviation in control angle is less than 3° (which is comparable to the expected error in maneuvering the sail).

(iii) It is found that the optimal strategy as well as the response may vary considerably depending on the starting point in the orbit.

(iv) Effectiveness of the optimal sail setting is compared with that of a near-optimal constant sail orientation showing a growing divergence in responses for increasing values of ϵ_s and e_{00}.

(v) The optimal steering program for maximizing angular momentum stays relatively close to the 35.26° line and coincides with the optimal sail setting for maximizing $a(2\pi)$ (in first-order) when $e_{00} = 0$.

References

[1] GARWIN, R. L. "Solar Sailing–A Practical Method of Propulsion Within the Solar System," *Jet Propulsion,* Vol. 28, No. 3, March 1958, pp. 188-190.

[2] TSU, T. C. "Interplanetary Travel by Solar Sail," *Journal of the American Rocket Society,* Vol. 29, No. 6, 1959, pp. 422-427.

[3] KIEFER, J. W. "Feasibility Considerations for a Sail Powered Multi-Mission Solar Probe," *Proceedings of the XV-th International Astronautical Congress of the I.A.F.,* Warszawa, 1964, Vol. I, Editor-in-chief: M. Lunc, Gauthier-Villars, Paris, 1965, pp. 383-416.

[4] LONDON, H. S. "Some Exact Solutions of the Equations of Motion of a Solar Sail with Constant Sail Setting," *Journal of the American Rocket Society,* Vol. 30, No. 2, 1960, pp. 198-200.

[5] MODI, V. J. and VAN DER HA, J. C. "Three-Dimensional Solar Sail Trajectories," Presented at the *Solar Sail Small Symposium,* Jet Propulsion Laboratory, Pasadena, California, April 20-21, 1977.

[6] LEBEDEV, V. N. "Some Problems of the Optimal Transfer Theory," *Proceedings of the XIV-th International Astronautical Congress of the I.A.F.,* Paris, 1963, Vol. IV, Editors: E. Brun and I. Hersey, Gauthier-Villars, Paris 1965, pp. 389-408.

[7] ZHUKOV, A. N. and LEBEDEV, V. N. "Variational Problem of Transfer Between Heliocentric Circular Orbits by Means of a Solar Sail," *Cosmic Research,* Vol. 2, No. 1, January-February 1964, pp. 41-44.

[8] MOISEYEV, N. N. and LEBEDEV, V. N. "Problems of Theory of Vehicles Optimal Controls," *Proceedings of the XVI-th International Astronautical Congress,* Athens, 1965, Guidance and Control, Editor-in-chief: M. Lunc, Gauthier-Villars-Dunod, Paris, 1966, pp. 301-320.

[9] LEE, E. B. and MARKUS, L. *Foundations of Optimal Control Theory,* John Wiley and Sons, New York, 1967.

[10] PONTRYAGIN, L. S., BOLTYANSKII, V. G., GAMKRELIDZE, R. V., and MISHCHENKO, E. F. *The Mathematical Theory of Optimal Processes,* Interscience, New York, 1962.

[11] BRYSON, A. E. and DENHAM, W. F. "A Steepest-Ascent Method for Solving Optimum Programming Problems," *Transactions of the ASME, Series E, Journal of Applied Mechanics,* Vol. 29, No. 2, June 1962, pp. 247-257.