Gradient Theory of Optimal Flight Paths

HENRY J. KELLEY[1]

Grumman Aircraft Engineering Corp.
Bethpage, N. Y.

An analytical development of flight performance optimization according to the method of gradients or "method of steepest descent" is presented. Construction of a minimizing sequence of flight paths by a stepwise process of descent along the local gradient direction is described as a computational scheme. Numerical application of the technique is illustrated in a simple example of orbital transfer via solar sail propulsion. Successive approximations to minimum time planar flight paths from Earth's orbit to the orbit of Mars are presented for cases corresponding to free and fixed boundary conditions on terminal velocity components.

THE PAST decade has seen considerable progress in techniques for the determination of flight paths which are optimal in the sense of various performance criteria. Treatments have employed almost exclusively the classical "indirect" method of the calculus of variations which is based on the reduction of variational problems to differential equations. A number of works on this subject are listed in (1 through 9).[2] An excellent bibliography is presented in the survey paper of (10).

Although many interesting results have been forthcoming from analytical solutions of the Euler-Lagrange differential equations governing optimal flight, the idealizing assumptions usually invoked limit their applicability in practical situations. Under more realistic assumptions, a numerical attack on these equations is required, and in this approach a serious difficulty may arise in the satisfaction of two-point boundary conditions. [See, for example, (11, 12 and 13).] This difficulty becomes a limiting factor where the order of the differential equations governing the basic system is four or higher.

Attention is directed in the present work to one of the direct methods of the calculus of variations, namely the method of gradients or "method of steepest descent," which offers circumvention of the two-point boundary value difficulty. The method also possesses the attractive feature of simultaneous optimization with respect to configuration parameters.

The notion of descent along the gradient direction was originally introduced by Hadamard in connection with mathematical existence proofs (14). Only in recent years has it found practical application to multivariable minimum problems of ordinary calculus (15) and to solution of systems of algebraic equations (16, 17) and integral equations (18).

The main idea of the present treatment stems from material presented by Prof. R. Courant in a 1941 address to the American Mathematical Society (19). An application of the gradient method to fixed end-point variational problems has been given by Stein (20). Our extension of the gradient idea to the case which includes differential equations as subsidiary conditions is heuristic in character.

Problem Formulation

For present purposes it will be assumed that the system of differential equations to be satisfied along the flight path is given in first-order form

$$\dot{x}_m = g_m(x_1, \ldots, x_n, y, t) \qquad m = 1, \ldots, n \qquad [1]$$

These equations relate velocities and positions, forces and accelerations, mass and flow of propellants and coolants, and the like. x_m may be termed problem variables, and y the control variable. Differentiation with respect to the independent variable, time t, is denoted by a superscribed dot.

An important class of problems is that in which the performance quantity to be minimized is expressed as a function of the final values of the variables x_m and t

$$P = P(x_{1f}, \ldots, x_{nf}, t_f) \qquad [2]$$

At a specified initial time t_0 as many as n boundary conditions on the x_m may be stipulated. Since an entire function $y(t)$ is at our disposal, we may reasonably consider problems in which numerous conditions are imposed upon the x_m at various subsequent t values. In the following we will restrict attention to conditions imposed at the terminal point of the flight path. Among the $n + 1$ quantities consisting of the n final values of the x_m plus the final time t_f, no more than n relations may be specified in order that the value of P not be predetermined.

Neighboring Solutions—Variations

We now assume that a solution of Equations [1] is available which satisfies the boundary conditions but which does not minimize P. Denoting the solution by $x_m = \bar{x}_m(t)$, $y = \bar{y}(t)$, we examine behavior in the neighborhood of this solution by setting $x_m = \bar{x}_m + \delta x_m$, $y = \bar{y} + \delta y$ and linearizing

$$\delta \dot{x}_m = \sum_{j=1}^{n} \frac{\partial g_m}{\partial x_j} \delta x_j + \frac{\partial g_m}{\partial y} \delta y \qquad m = 1, \ldots, n \qquad [3]$$

The partials of g_m are evaluated along $x_m = \bar{x}_m$, $y = \bar{y}$ and are therefore known functions of the independent variable t. The functions δx_m and δy are the *variations* of x_m and y in the neighborhood of $\bar{x}_m$, $\bar{y}$.

A formal solution of Equations [3] may be written in the form

$$\delta x_m = \sum_{p=1}^{n} \delta x_{p0} \xi_{mp}(t) + \int_{t_0}^{t} \mu_m(\tau, t - \tau) \delta y(\tau) d\tau \qquad m = 1, \ldots, n \qquad [4]$$

where the first member represents solution of the homogeneous system of equations and the second a superposition of control variable effects. The functions μ_m are Green's functions or in-

Presented at the ARS Semi-Annual Meeting, May 9–12, 1960, Los Angeles, Calif.

[1] Section Leader, Systems Research. Member ARS.

[2] Numbers in parentheses indicate References at end of paper.

fluence functions; $\mu_m(\tau, t - \tau)$ may be thought of as the solution for δx_m corresponding to δy a unit impulse (Dirac delta function) introduced at time τ (21).

Since interest centers on final values of x_m, we evaluate the expressions [4] at $t = t_f$; however, to cover the possibility of a variable end point, $t_f = \bar{t}_f + \delta t_f$, a first-order correction term must be included

$$\delta x_{mf} = \sum_{p=1}^{n} \delta x_{p0}\xi_{mp}(\bar{t}_f) + \int_{t_0}^{\bar{t}_f} \mu_m(\tau, \bar{t}_f - \tau)\delta y(\tau)d\tau + \dot{\bar{x}}_{mf}\delta t_f$$
$$= \sum_{p=1}^{n} \delta x_{p0}\xi_{mp}(\bar{t}_f) + \int_{t_0}^{\bar{t}_f} \mu_m(\tau, \bar{t}_f - \tau)\delta y(\tau)d\tau + \bar{g}_{mf}\delta t_f \qquad m = 1, \ldots, n \quad [5]$$

where $\bar{g}_{mf} = g_m(\bar{x}_{1f}, \ldots, \bar{x}_{nf}, \bar{y}_f, \bar{t}_f)$.

Computation of the Functions μ_m

Since computation of the functions $\mu_m(\tau, t - \tau)$ over a complete range of both arguments is unnecessary, only their evaluation at $t = \bar{t}_f$ required for subsequent calculations, it is reasonable to seek a means for performing the special computation which avoids the labor of the more general one. The following development relates the functions $\mu_m(\tau, \bar{t}_f - \tau)$ to solutions of a system of equations adjoint to the system [3] through an application of Green's theorem. The scheme employed is due to Bliss, as reported by Goodman and Lance (22).

We rewrite Equations [3] employing a subscript notation suitable to our immediate purpose

$$\delta \dot{x}_i = \sum_{j=1}^{n} \frac{\partial g_i}{\partial x_j}\delta x_j + \frac{\partial g_i}{\partial y}\delta y \qquad i = 1, \ldots, n \quad [6]$$

and write the system of equations adjoint to this system

$$\dot{\lambda}_i = -\sum_{j=1}^{n} \frac{\partial g_j}{\partial x_i}\lambda_j \qquad i = 1, \ldots, n \quad [7]$$

which is obtained by transposing the matrix of coefficients and changing the sign.

The solutions of the two systems are related by

$$\frac{d}{dt}\sum_{i=1}^{n} \lambda_i \delta x_i = \sum_{i=1}^{n} \lambda_i \frac{\partial g_i}{\partial y}\delta y \quad [8]$$

After integration of both left and right members between definite limits t_0 and $\bar{t}_f$, we find

$$\sum_{i=1}^{n} \lambda_i(\bar{t}_f)\delta x_i(\bar{t}_f) - \sum_{i=1}^{n} \lambda_i(t_0)\delta x_i(t_0) = \int_{t_0}^{\bar{t}_f} \sum_{i=1}^{n} \lambda_i \frac{\partial g_i}{\partial y}\delta y\, dt \quad [9]$$

This is the one-dimensional form of Green's theorem (22).

We now consider numerical solutions of the adjoint system with all boundary values specified at $t = \bar{t}_f$. To the special solutions corresponding to

$$\lambda_i(\bar{t}_f) = 0 \qquad i \neq m$$
$$\lambda_i(\bar{t}_f) = 1 \qquad i = m \quad [10]$$

we assign the symbols $\lambda_i^{(m)}(t)$. In this fashion n expressions for the values of the $\delta x_m(\bar{t}_f)$ are obtained from [9]

$$\delta x_m(\bar{t}_f) = \sum_{i=1}^{n} \lambda_i^{(m)}(t_0)\delta x_i(t_0) + \int_{t_0}^{\bar{t}_f} \sum_{i=1}^{n} \lambda_i^{(m)} \frac{\partial g_i}{\partial y}\delta y\, dt \qquad m = 1, \ldots, n \quad [11]$$

By comparison with Equations [5] it may be seen that the desired relation between the $\mu_m(\tau, \bar{t}_f - \tau)$ and the $\lambda_i^{(m)}$ is the following one

$$\mu_m(\tau, \bar{t}_f - \tau) = \sum_{i=1}^{n} \lambda_i^{(m)} \frac{\partial g_i}{\partial y} \qquad m = 1, \ldots, n \quad [12]$$

and, also, that the $\xi_{mp}(\bar{t}_f)$ of Equations [5] is equal to $\lambda_p^{(m)}(t_0)$.

In the preceding development the choice of symbols λ for the variables of the adjoint system is deliberate, for Equations [7] are precisely those governing the Lagrange multiplier functions of the "indirect" theory. We note the important distinction, however, that the coefficients of [7] employed in the "indirect" theory are evaluated along a minimal solution of Equations [1], whereas in gradient computations they correspond to nonminimal paths.

Descent Parameter

Following Courant (19), we now introduce a parameter σ as a second independent variable, and seek a functional dependence of the control variable $y(t, \sigma)$ on this parameter such that the derivative of the performance quantity $dP/d\sigma$ is negative. In fact, within the restrictions imposed by the boundary conditions and the system equations, we shall attempt to make the slope of descent $dP/d\sigma$ "as steep as possible."

To enable operation within the restrictions just mentioned, we break down the control variable y as follows

$$y(t, \sigma) = \phi(t, \sigma) + \sum_{q=1}^{r} a_q(\sigma)f_q(t) \quad [13]$$

Here the $f_q(t)$ are a set of known linearly independent functions of t. Our intention is that the coefficients a_q of the second member of [13] be sacrificed to the fulfillment of boundary conditions, the number r being chosen appropriately for this purpose. This will leave the function ϕ free for the minimization of P.

We now take the derivatives of various quantities with respect to σ and evaluate them at $\sigma = \bar{\sigma}$, corresponding to the nonminimal solution $x_m(t, \bar{\sigma}) = \bar{x}_m$, $y(t, \bar{\sigma}) = \bar{y}$ introduced in the preceding section. For $\sigma - \bar{\sigma} = \Delta\sigma$ small, the variations appearing in Equations [5] may be identified as

$$x_{m_0} = \bar{x}_{m_0} + \delta x_{m_0} = \bar{x}_{m_0} + (dx_{m_0}/d\sigma)\Delta\sigma \qquad m = 1, \ldots, n \quad [14]$$

$$x_{mf} = \bar{x}_{mf} + \delta x_{mf} = \bar{x}_{mf} + (dx_{mf}/d\sigma)\Delta\sigma \qquad m = 1, \ldots, n \quad [15]$$

$$t_f = \bar{t}_f + \delta t_f = \bar{t}_f + (dt_f/d\sigma)\Delta\sigma \quad [16]$$

$$y = \bar{y} + \delta y = \bar{y} + (\partial y/\partial\sigma)\Delta\sigma$$
$$= \bar{\phi} + \sum_{q=1}^{r} \bar{a}_q f_q(t) + \left[\frac{\partial\phi}{\partial\sigma} + \sum_{q=}^{r} \frac{da_q}{d\sigma} f_q(t)\right]\Delta\sigma \quad [17]$$

Equations [5] then take the form

$$\frac{\partial x_{mf}}{\partial\sigma} = \sum_{p=1}^{n} \frac{dx_{p0}}{d\sigma}\xi_{mp}(\bar{t}_f) + \int_{t_0}^{\bar{t}_f} \mu_m(\tau, \bar{t}_f - \tau)\frac{\partial\phi}{\partial\sigma}d\tau +$$
$$\sum_{q=1}^{r} \frac{da_q}{d\sigma}\int_{t_0}^{\bar{t}_f} \mu_m(\tau, \bar{t}_f - \tau)f_q(\tau)d\tau + \bar{g}_{mf}\frac{dt_f}{d\sigma} \qquad m = 1, \ldots, n \quad [18]$$

Boundary Conditions

We consider boundary conditions of the separated type, i.e., equations relating either initial values or final values. Boundary values may be variable on a surface, typified by

$$\mathfrak{Z}(x_{uf}, x_{vf}, t_f) = 0 \quad [19]$$

in which case the following linear combination of derivatives must vanish

$$\frac{d\mathfrak{Z}}{d\sigma} = \frac{\partial\mathfrak{Z}}{\partial x_u}\frac{dx_{uf}}{d\sigma} + \frac{\partial\mathfrak{Z}}{\partial x_v}\frac{dx_{vf}}{d\sigma} + \frac{\partial\mathfrak{Z}}{\partial t_f}\frac{dt_f}{d\sigma} = 0 \quad [20]$$

In this expression the partial derivatives of $\mathfrak{J}$ are evaluated at $x_{uf} = \bar{x}_{uf}$, $x_{vf} = \bar{x}_{vf}$ and $t_f = \bar{t}_f$.

In the case of fixed boundary conditions of the form $x_{uf} = \bar{x}_{uf} =$ constant, $t_f = \bar{t}_f =$ constant, the relations to be satisfied take the simple form

$$\frac{dx_{uf}}{d\sigma} = 0 \quad \text{or} \quad \frac{dt_f}{d\sigma} = 0 \qquad [21]$$

As many as n final conditions, fixed or of the form [19] may be specified, as mentioned earlier. Where fewer than this number are specified, we speak of "free" or "open" boundary values.

At the initial point similar freedom of choice may exist among the n initial values of the x_m, the only difference here being that t_0 will usually be fixed.

Gradient of P

Upon inspection it may be noted that there are n equations of the form [18] and at most $2n$ equations for boundary values, making a possible total of $3n$ equations. These relate $2n + 1$ derivatives of the x_{m_0}, x_{mf} and t_f, the r derivatives of the a_q and integrals containing $\partial\phi/\partial\sigma$. We accordingly choose the number r of the a_q as $r = n - 1 - s$, where s is the number of open boundary conditions. Thus the system of equations will be determinate for arbitrary $\partial\phi/\partial\sigma$.

These equations may be arranged as a linear simultaneous system

$$AZ = B \qquad [22]$$

where Z has $3n - s$ elements consisting of the derivatives with respect to σ of x_{m_0}, x_{mf}, t_f and a_q. The matrix A is square and contains among its elements the quantities ξ_{mp}, $\bar{g}_{mf}$, the $\mathfrak{J}$ partials, and integrals of products $\mu_m f_q$, all of which are known. The column B includes integrals containing $\partial\phi/\partial\sigma$ in their integrands. The solution

$$Z = A^{-1}B \qquad [23]$$

may be obtained through matrix inversion or equivalent processes. The matrix A expresses the relationship between small shifts in boundary values and small adjustments in the control function (through the a_q) to deal with them. Hence in normal circumstances A will be nonsingular.

The derivative with respect to σ of the quantity P to be minimized

$$\frac{dP}{d\sigma} = \frac{\partial P}{\partial x_{1f}}\frac{dx_{1f}}{d\sigma} + \ldots + \frac{\partial P}{\partial x_{nf}}\frac{dx_{nf}}{d\sigma} + \frac{\partial P}{\partial t_f}\frac{dt_f}{d\sigma} \qquad [24]$$

may now be expressed in terms of known quantities and integrals containing $\partial\phi/\partial\sigma$ according to the solution [23]. It will take the form of a linear combination of the integrals

$$\frac{dP}{d\sigma} = \sum_{m=1}^{n} C_m \int_{t_0}^{\bar{t}_f} \mu_m(\tau, \bar{t}_f - \tau)\frac{\partial\phi}{\partial\sigma}\, d\tau \qquad [25]$$

which may be expressed as a single integral

$$\frac{dP}{d\sigma} = \int_{t_0}^{\bar{t}_f} \left[\sum_{m=1}^{n} C_m\mu_m(\tau, \bar{t}_f - \tau)\right]\frac{\partial\phi}{\partial\sigma}\, d\tau \qquad [26]$$

By analogy with a characteristic property of a vector gradient, we are now prepared to identify the gradient of P with respect to the function ϕ. [See p. 222 of the Courant-Hilbert English edition, (23), also (19).] If ϕ were a vector possessing a finite number of components ϕ_i, $i = 1, \ldots, j$, the derivative of P with respect to σ

$$\frac{dP}{d\sigma} = \sum_{i=1}^{j}\frac{\partial P}{\partial\phi_i}\frac{d\phi_i}{d\sigma} \qquad [27]$$

could be expressed as

$$\frac{dP}{d\sigma} = \text{grad } P \cdot \frac{d\phi}{d\sigma} \qquad [28]$$

If $\phi(t, \sigma)$ is a continuous function, and P a functional of ϕ, as in the case of present interest

$$\frac{dP}{d\sigma} = \int [P]_\phi \frac{\partial\phi}{\partial\sigma}\, d\tau \qquad [29]$$

and the function occurring in product with $\partial\phi/\partial\sigma$, which we denote $[P]_\phi$, may be regarded as the gradient of P by extended definition.

Thus for the problem at hand it is evident from Equation [26] that

$$[P]_\phi = \sum_{m=1}^{n} C_m\mu_m(\tau, \bar{t}_f - \tau) \qquad [30]$$

is the gradient of P.

Were the solution $x_m = \bar{x}_m$, $y = \bar{y}$ such as to minimize P, contrary to our assumption, the gradient $[P]_\phi$ would vanish. If our development were to parallel the classical or "indirect" approach, the construction of a solution would be sought from the vanishing of $[P]_\phi$ for which P is stationary and possibly a minimum. Thus $[P]_\phi$ is in some sense an Euler expression and the equation $[P]_\phi = 0$ and the relations [23] have an equivalence to the Euler-Lagrange equations and transversality conditions of the "indirect" theory.

Descent Process

Returning momentarily to the elementary geometric concept of a vector gradient, we regard $P(\phi_1, \ldots, \phi_j)$ as a surface, and starting from $P(\phi_1, \ldots, \phi_j)$ we move a point along this surface so that P and ϕ_i become functions of a time parameter σ. Then the velocity of ascent or descent along a line on the surface is as given by Equations [27 and 28]. We now choose the line along which the descent is as steep as possible, characterized by

$$\frac{d\phi_i}{d\sigma} = k\frac{\partial P}{\partial\phi_i} \qquad [31]$$

where k positive corresponds to ascent and k negative to descent.

It is clear that in this continuous process, wherein the point moves according to the system of ordinary differential equations [31], the process will for $\sigma \to \infty$ approach a position for which grad $P = 0$, if P is bounded below.

This elementary idea may be generalized to the present variational problem according to the extended interpretation of the gradient of P, Equation [29]. We set

$$\frac{\partial\phi}{\partial\sigma} = k[P]_\phi = -[P]_\phi \qquad [32]$$

and starting from the nonminimal solution $\sigma = \bar{\sigma}$, $\phi = \bar{\phi}$, $x_m = \bar{x}_m$, $a_q = \bar{a}_q$, re-evaluate these quantities on a continuous basis as the parameter σ increases.

Thus the continuous version of descent along the gradient requires numerical treatment of a partial differential equation for $\phi(t, \sigma)$ with determination of $\phi(t, \infty)$ the ultimate objective.

Stepwise Version

As an alternative to the continuous procedure given by Equation [32], we may elect to proceed stepwise, correcting a set of approximations to the solution $[P]_\phi = 0$ by corrections proportional to the negative of the gradient

$$\phi^{(i+1)} = \phi^{(i)} - [P]_\phi\Delta\sigma \qquad [33]$$

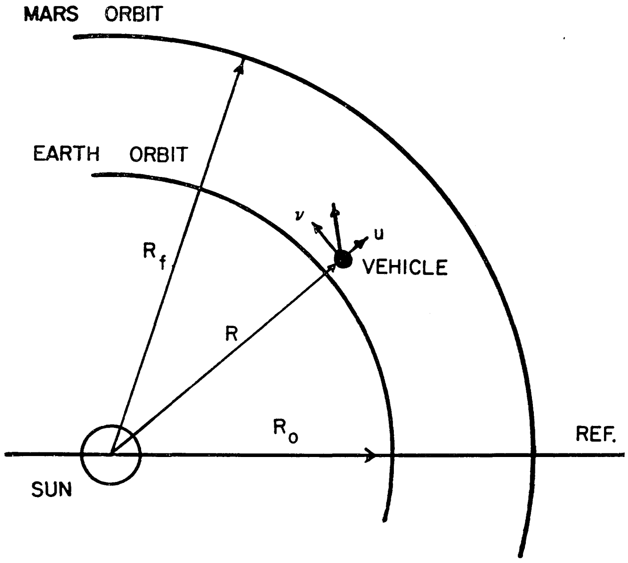

Fig. 1 Orbital transfer schematic

After choice of step size $\Delta\sigma$ and evaluation of $\phi^{(i+1)}$ and the $a_q^{(i+1)}$, the functions $x_m^{(i+1)}$ would be obtained from numerical integration of the basic system Equations [1] and $P^{(i+1)}$ determined from the values at $t_f^{(i+1)}$.

Since the determination of the gradient $[P]_\phi$ is expensive in terms of volume of numerical computations, it would seem desirable to exploit each calculation of local gradient direction of the utmost, taking $\Delta\sigma$ as large as possible. A procedure employed in some applications (17) is to follow the local gradient direction until the function P reaches a minimum. Such a procedure could be implemented by numerical integration of Equations [1] for a number of values of $\Delta\sigma$ and selection of the value of $\Delta\sigma$ for minimum P.

A possible pitfall of such a scheme of operation is that the boundary conditions, having been satisfied only in linearized version (Eq. [20]), may long since have been violated before minimum P is reached (24). Thus the choice of $\Delta\sigma$ involves a compromise which must be decided in the particular applition at hand. If the minimum P rule is employed for choice of $\Delta\sigma$, the boundary conditions must be restored by a correction cycle designed to recover from the departures.

After such restoration has been accomplished, for example via the coefficients a_q, the solution so obtained takes on the role of $\bar{x}_m$, $\bar{y}$ for the computation of a new local gradient direction.

Convergence

The convergence of the descent process has been investigated by Stein for a case not complicated by subsidiary conditions (20); his conclusion was that the process will converge if the functional whose minimum is sought is bounded below. In the process described in the preceding sections, the question of convergence is intimately related to the success of the technique for maintaining or correcting the boundary values at the terminal point.

The possibility of correcting small departures in terminal values of the x_m through small changes in the coefficients a_q can be shown to hinge on the nonvanishing of the determinant of the matrix A of Equation [22]. In certain cases a tendency of this determinant to become small may be observed as a minimum of P is approached. Where the difficulty is acute, the behavior may be likened to breakdown of a first-order differential correction scheme for guidance along the trajectory.

This type of behavior is a close relative of the conjugate end-point phenomenon of classical variational theory. If an extremal through point 1 has on it a contact point 2 with the envelope of a family of extremals through 1, then point 2 is said to be conjugate to point 1. The term *conjugate end points* refers to specification of boundary values at two such points as 1 and 2. The reader is advised to consult the text of Bliss (25) for the analytical basis of conjugate point theory.

The relevant test of the classical theory, *the necessary condition of Jacobi*, employs a determinant as a criterion. The requisite analysis amounts to investigation of *optimal* differential corrections of terminal values. Thus a matrix which is the optimal correction analog of the matrix A is set up, and the criterion of the Jacobi test is the nonvanishing of the determinant of this matrix at points along the extremal *up to* the terminal point of interest. Vanishing *at* the terminal point indicates that the end points are conjugates, and the necessary condition is fulfilled in borderline fashion. Computational procedures for applying the Jacobi test are of some interest; however, the acquisition of a "test specimen" solution requires that any convergence difficulty encountered in the descent process first be overcome.

Thus a possibility of convergence difficulty arises in conjugate end-point cases if the functions f_q selected happen to resemble the optimal correction functions of the indirect theory. There are certainly other possibilities for unfortunate choice of correction functions. Computational experience will be required to establish guidelines on this matter.

Configuration Parameters

In many practical engineering applications, optimal performance is sought not only in terms of flight path selection but also in terms of parameters influencing vehicle configuration. We have, for clarity, avoided complicating the preceding analytical work by such considerations. It is an easy matter, however, to add terms to Equations [5] of the form $(\partial x_{mf}/\partial e_i)\delta e_i$ and to carry them through along with constraints relating e_i, thus forming a basis for simultaneous optimization of configuration and flight path. The slopes $\partial x_{mf}/\partial e_i$ are probably best determined by numerical integration of Equations [1] for small changes in e_i.

Solar Sailing Example

For the purpose of exploring the computational aspect of the gradient optimization technique, we have chosen a planar case of transfer between planetary orbits by means of the interesting solar sailing scheme. The potential capabilities of solar sail propulsion have been investigated in the papers of Garwin (26), Cotter (27), Tsu (28) and London (29). This problem has the simplicity appropriate to an exploration of method, yet sufficient complexity to render analytical solution quite difficult unless drastic simplifications are introduced.

The equations of motion and kinematic relations are given in a notation nearly the same as that of Tsu (28). With reference to the schematic of Fig. 1, these are as follows:

Radial acceleration

$$\dot{u} = g_1 = \frac{v^2}{R} - A_0\left(\frac{R_0}{R}\right)^2 + \alpha\left(\frac{R_0}{R}\right)^2 |\cos^3\theta| \qquad [34]$$

Circumferential acceleration

$$\dot{v} = g_2 = -\frac{uv}{R} - \alpha(R_0/R)^2 \sin\theta\cos^2\theta \qquad [35]$$

Radial velocity

$$\dot{R} = g_3 = u \qquad [36]$$

Circumferential angular velocity

$$\dot{\psi} = v/R \qquad [37]$$

Since the heliocentric angle ψ does not appear in the first three equations, and will not appear in the statements of boundary conditions to be considered, Equation [37] may be ignored for purposes of gradient optimization. This amounts to the assumption that terminal matching of the heliocentric angles of vehicle and "target" planet is accomplished by selection of launch time.

Seeking minimum-time transfer, we identify the functional P as

$$P = t_f \tag{38}$$

The functions u, v, R are the variables x_m of the theoretical development, and the sail angle θ appears in the role of the control variable y.

As initial conditions we specify velocity components u, v and radius R corresponding to motion in Earth's orbit approximated as a circle

$$t_0 = 0 \tag{39}$$

$$u(0) = u_0 = u_E = 0 \tag{40}$$

$$v(0) = v_0 = v_E \tag{41}$$

$$R(0) = R_0 = R_E \tag{42}$$

We consider terminal conditions corresponding to arrival at the orbit of the planet Mars (also taken as a circle) with prescribed velocity components

$$u(t_f) = u_f \tag{43}$$

$$v(t_f) = v_f \tag{44}$$

$$R(t_f) = R_f = R_M \tag{45}$$

For fixed boundary values of u, v and R, the equations corresponding to Equations [5] of the preceding theoretical development are

$$\delta u_f = \int_0^{\bar{t}_f} \mu_1 \delta\theta \, d\tau + \bar{g}_{1f}\delta t_f = 0 \tag{46}$$

$$\delta v_f = \int_0^{\bar{t}_f} \mu_2 \delta\theta \, d\tau + \bar{g}_{2f}\delta t_f = 0 \tag{47}$$

$$\delta R_f = \int_0^{\bar{t}_f} \mu_3 \delta\theta \, d\tau + \bar{g}_{3f}\delta t_f = 0 \tag{48}$$

In this case the number of functions f_q and coefficients a_q required is

$$r = n - 1 - s = 3 - 1 - 0 = 2 \tag{49}$$

We select the functions f_q as

$$f_1(t) = t \tag{50}$$

$$f_2(t) = t^2 \tag{51}$$

and the control function $\theta(t)$ is broken down as

$$\theta = \phi + a_1 t + a_2 t^2 \tag{52}$$

The system of equations corresponding to Equation [22] simplifies to

$$\left(\int_0^{\bar{t}_f} \tau\mu_1 d\tau\right)\frac{da_1}{d\sigma} + \left(\int_0^{\bar{t}_f} \tau^2\mu_1 d\tau\right)\frac{da_2}{d\sigma} + \bar{g}_{1f}\frac{dt_f}{d\sigma} = -\int_0^{\bar{t}_f} \mu_1 \frac{\partial\phi}{\partial\sigma}\, d\tau \tag{53}$$

$$\left(\int_0^{\bar{t}_f} \tau\mu_2 d\tau\right)\frac{da_1}{d\sigma} + \left(\int_0^{\bar{t}_f} \tau^2\mu_2 d\tau\right)\frac{da_2}{d\sigma} + \bar{g}_{2f}\frac{dt_f}{d\sigma} = -\int_0^{\bar{t}_f} \mu_2 \frac{\partial\phi}{\partial\sigma}\, d\tau \tag{54}$$

$$\left(\int_0^{\bar{t}_f} \tau\mu_3 d\tau\right)\frac{da_1}{d\sigma} + \left(\int_0^{\bar{t}_f} \tau^2\mu_3 d\tau\right)\frac{da_2}{d\sigma} + \bar{g}_{3f}\frac{dt_f}{d\sigma} = -\int_0^{\bar{t}_f} \mu_3 \frac{\partial\phi}{\partial\sigma}\, d\tau \tag{55}$$

This 3×3 case may conveniently be inverted analytically. There seems little point in listing the inverse elements here, however.

$$\begin{bmatrix} \dfrac{da_1}{d\sigma} \\ \dfrac{da_2}{d\sigma} \\ \dfrac{dt_f}{d\sigma} \end{bmatrix} = -\begin{bmatrix} & & \\ & C & \\ & & \end{bmatrix}\begin{bmatrix} \int_0^{\bar{t}_f} \mu_1 \dfrac{\partial\phi}{\partial\sigma}\, d\tau \\ \int_0^{\bar{t}_f} \mu_2 \dfrac{\partial\phi}{\partial\sigma}\, d\tau \\ \int_0^{\bar{t}_f} \mu_3 \dfrac{\partial\phi}{\partial\sigma}\, d\tau \end{bmatrix} \tag{56}$$

The slope of descent $dt_f/d\sigma$ is given by

$$\frac{dt_f}{d\sigma} = -\int_0^{\bar{t}_f} (C_{31}\mu_1 + C_{32}\mu_2 + C_{33}\mu_3)\frac{\partial\phi}{\partial\sigma}\, d\tau \tag{57}$$

and the gradient of P by

$$[P]_\phi = -(C_{31}\mu_1 + C_{32}\mu_2 + C_{33}\mu_3) \tag{58}$$

Accordingly we set

$$\partial\phi/\partial\sigma = -[P]_\phi = C_{31}\mu_1 + C_{32}\mu_2 + C_{33}\mu_3 \tag{59}$$

and proceed with stepwise descent as in Equation [33].

A particularly suitable case for a first illustration of computational technique is the one in which terminal velocity components are unspecified—"free" boundary conditions. Here Equations [46 and 47] may be deleted and

$$r = n - 1 - s = 3 - 1 - 2 = 0 \tag{60}$$

so that no functions f_q are needed. Hence

$$\theta = \phi$$

and

$$\frac{\partial\phi}{\partial\sigma} = \frac{\mu_3}{\bar{g}_{3f}} \tag{61}$$

$$\frac{dt_f}{d\sigma} = -\frac{1}{\bar{g}_{3f}^{\,2}}\int_0^{\bar{t}_f} \mu_3^2\, d\tau \tag{62}$$

in this case.

Computations have employed numerical values of the various constants from Tsu's paper, with

$$\alpha = 0.1 \text{ cm/sec}^2 = 3.28 \times 10^{-3} \text{ fps}^2$$

This value corresponds to about $10^{-4}\, g$ thrust acceleration developed by the sail when oriented broadside to the sun ($\theta = 0$) at Earth's orbit radius, or about 17 per cent of the sun's gravitational attraction.

Results of descent computations for the case of "open" terminal velocity components are shown in Figs. 2 through 4. The control program of the original flight path (Figs. 2 and 3), chosen arbitrarily, was far from optimal in that the radial velocity component at crossing of Mars' orbit was small. The greatest reduction in flight time—more than half of the original—is seen to be obtained in the course of the first descent (Fig. 4). In three descents minimum flight time has been attained for practical purposes, although small changes in the detailed structure of the control program are still in evidence.

Results for the case of terminal velocity components matched to the target planet

$$u_f = u_M \qquad v_f = v_M$$

are presented in Figs. 5 through 8. The tendency of the terminal values to depart from the prescribed values is shown in Fig. 5. These were restored via an iterative correction process employing increments in the coefficients a_1 and a_2 of Equation [52]. Typically, two or three iteration cycles were required to correct each point. Descent curves are shown in Fig. 6. The approach to the minimum-time solution is depicted in Figs. 7 and 8.

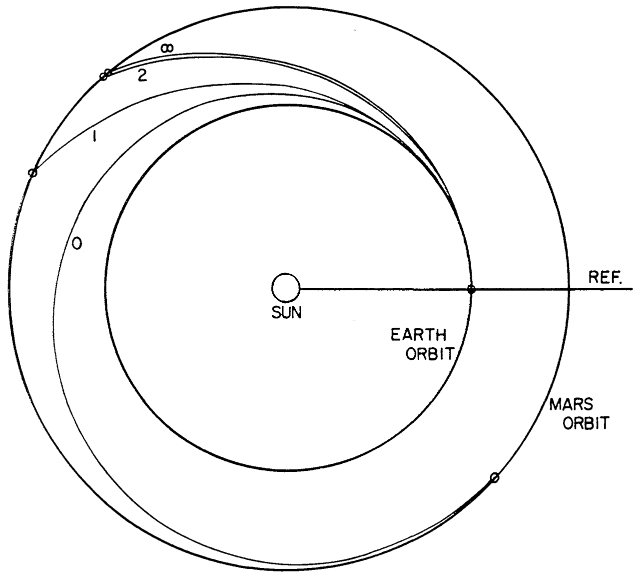

Fig. 2 Successive approximations to optimal transfer path, terminal velocity components open

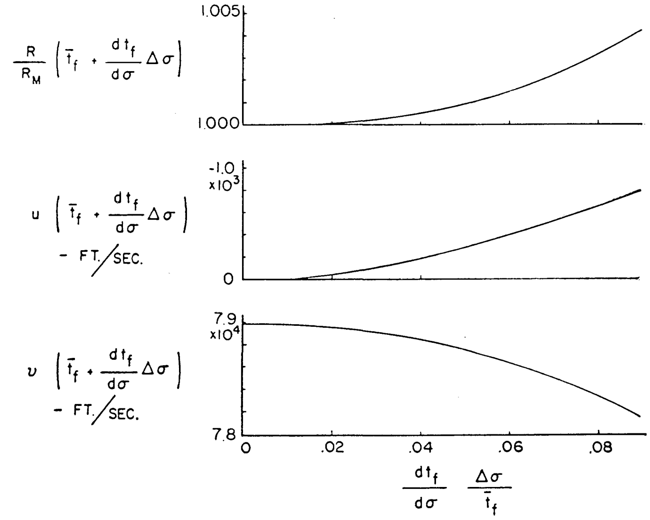

Fig. 5 Departure of terminal values, "matched" terminal velocity components

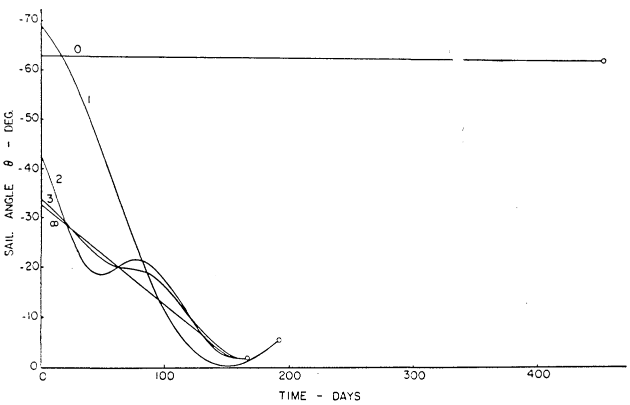

Fig. 3 Successive approximations to optimal sail angle program, terminal velocity components open

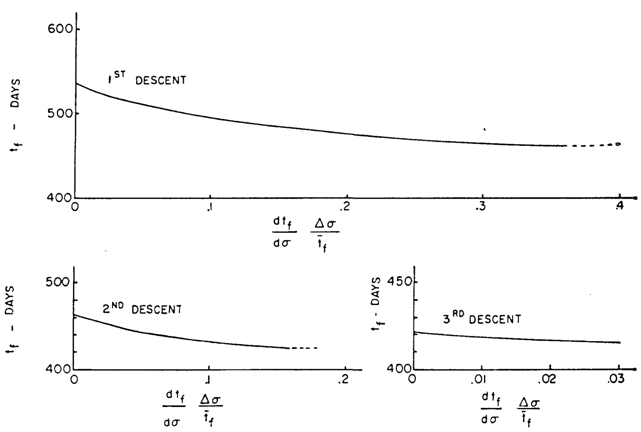

Fig. 6 Descent curves—solar sail transfer, "matched" terminal velocity components

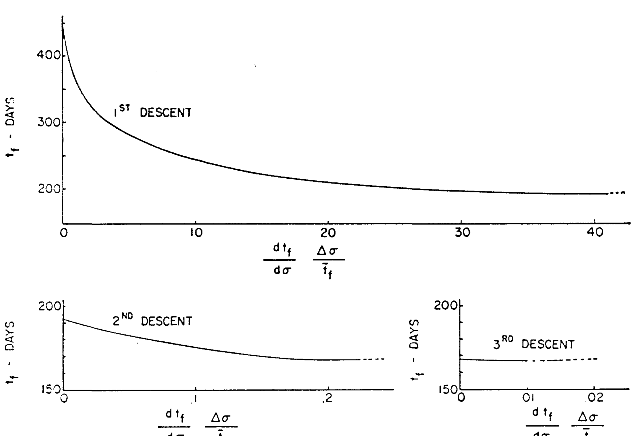

Fig. 4 Descent curves—solar sail transfer, terminal velocity components open

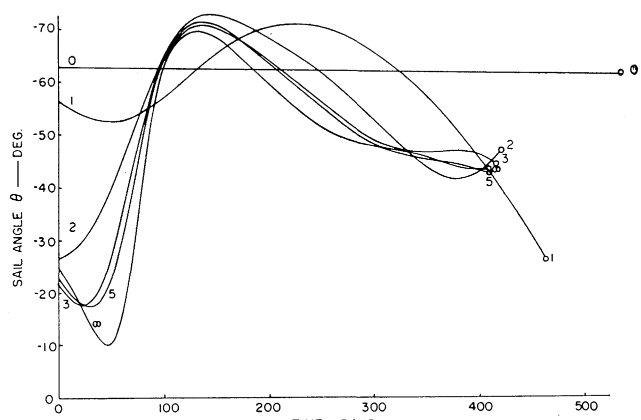

Fig. 7 Successive approximations to optimal sail angle program, "matched" terminal velocity components

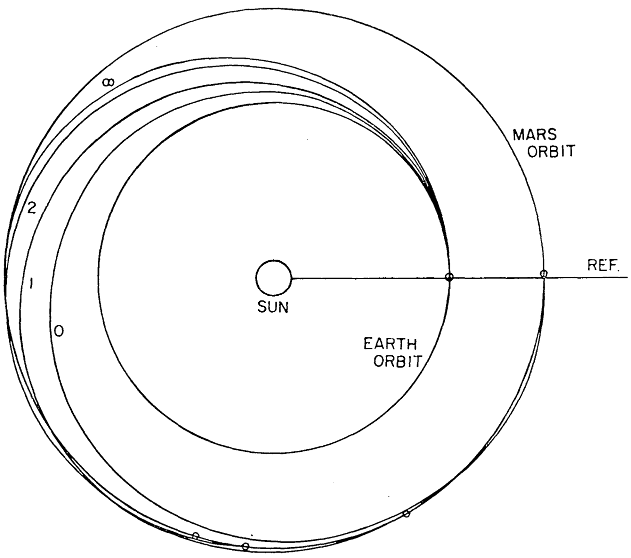

Fig. 8 Successive approximations to optimal transfer path, "matched" terminal velocity components

The first attempt at computations for this "matched velocity" case employed functions f_q constant and linear with time. This met with near zero determinant difficulty of the type discussed earlier. The combination of linear and square-law corrections indicated in the foregoing was successfully used to avoid this difficulty.

Concluding Remarks

Attention has been confined in the preceding development to the main ideas of the gradient technique. The extension to cases involving several control variables offers no particular difficulty. The limited computational experience reported here suggests that the gradient technique may be a useful one in applications, and particularly in those presenting difficulty when the classical "indirect" approach—numerical integration of the Euler-Lagrange equations—is used. Apart from the feature of surmounting the two-point boundary value difficulties of such cases, the gradient method may be particularly appropriate in distinguishing between minimal solutions of the Euler-Lagrange equations and those which are merely stationary.

Acknowledgments

The writer is pleased to acknowledge the contributions of Messrs. William P. O'Dwyer and H. Gardner Moyer of Grumman's Computation Facility in handling the computational phase of this study on the IBM 704, and of Mrs. Agnes Zevens of the Systems Research Section in checking and preparing the numerical results for publication.

Nomenclature

x_m	= problem variables
y	= control variable
g_m	= functions of x_m and y appearing in basic system equations
t	= time
P	= function of final values of x_m to be minimized
δx_m, δy	= variations of x_m and y
ξ_{mp}	= functions appearing in the solution of Equations [3]
μ_m	= Green's functions in the solution of Equations [3]
τ	= variable of integration
λ	= variables of the adjoint system, Equation [7]
σ	= descent parameter
ϕ	= auxiliary control variable, Equation [13]
f_q	= known linearly independent functions of t
a_q	= coefficients of the functions f_q
$\Delta\sigma$	= increment in σ
$\mathfrak{J}$	= function relating boundary values of x_m and t
A	= matrix of coefficients, Equation [22]
Z	= column matrix having $dx_{m_0}/d\sigma$, $dx_{m_f}/d\sigma$, $dt_f/d\sigma$ and $da_q/d\sigma$ as elements, Equation [22]
B	= column matrix containing integrals of $\mu_m(\partial\phi/\partial\sigma)$ products, Equation [22]
C_m	= coefficients of the linear combination of μ_m, Equation [26]
$[P]_\phi$	= gradient of P, Equation [30]
k	= proportionality constant of descent
e_i	= configuration parameters
u	= radial velocity component (see Fig. 1)
v	= circumferential velocity component (see Fig. 1)
R	= radial distance to sun (see Fig. 1)
ψ	= heliocentric angle
A_0	= acceleration due to sun's gravitational attraction at Earth's radial distance
α	= thrust acceleration of radially oriented solar sail at Earth's radial distance
θ	= sail angle measured from radial orientation

Subscripts

f	= denotes final value
0	= denotes initial value
m, p, i, j / q, r, u, v	= general indexes
E	= Earth
M	= Mar
$(\dot{\ })$	= denotes time derivative
$(\bar{\ })$	= denotes nonminimal solution
(i)	= denotes value at ith step
(m)	= denotes special functional notation, explained by Equation [10]

References

1 Hestenes, M. R., "A General Problem in the Calculus of Variations with Applications to Paths of Least Time," Rand RM-100, March 1950.

2 Garfinkel, B., "Minimal Problems in Airplane Performance," *Quart. Appl. Math.*, vol. 9, no. 2, 1951.

3 Cicala, P., "Le Evoluzioni Ottime di un Aereo," Atti della Accademia delle Scienze di Torino, Classe di Scienze Matematiche, Fisiche e Naturali, vol. 89, 1954–1955. (Translated as NASA Technical Translation TTF-4, Oct. 1959.)

4 Miele, A., "An Extension of the Theory of the Optimum Burning Program for the Level Flight of a Rocket-Powered Aircraft," Purdue University School of Aeronautical Engineering, Rep. no. A-56-1, Technical Note no. AFOSR-TN-56-302, June 1956; *also J. Aeron. Sci.*, vol. 24, no. 12, Dec. 1957.

5 Breakwell, J. V., "The Optimization of Trajectories," North American Aviation, Inc., Rep. AL-2706, Aug. 1957; *also J. Soc. Ind. Appl. Math.*, vol. 7, no. 2, June 1959.

6 Okhotsimskii, D. E. and Eneev, T. M., "Some Variation Problems Connected with the Launching of Artificial Satellites of the Earth," *Uspekhi Fizicheskikh Nauk*, Sept. 1957. (Translated in *J. Brit. Interplanet. Soc.*, vol. 16, no. 5, Jan.-Feb. 1958.)

7 Miele, A., "General Variational Theory of the Flight Paths of Rocket-Powered Aircraft, Missiles and Satellite Carriers," Purdue University School of Aeronautical Engineering, Rep. no. A-58-2, Technical Note no. AFOSR-TN-58-246, Jan. 1958; *also Astronautica Acta*, vol. IV, fasc. IV, Dec. 1958.

8 Kelley, H. J., "An Investigation by Variational Methods of Flight Paths for Optimum Performance," Sc.D. thesis, Guggenheim School of Aeronautics, New York University, May 1958. (Published in part as "An Investigation of Optimal Zoom Climb Techniques," *J. Aero/Space Sci.*, vol. 26, no. 12, Dec. 1959.)

9 Leitmann, G., "On a Class of Variational Problems in Rocket Flight," Lockheed Missile and Space Systems Division Rep. 5067, September 1958; *also J. Aero/Space Sci.*, vol. 26, no. 3, Sept. 1959.

10 Leitmann, G., "The Optimization of Rocket Trajectories—A Survey," in "Progress in the Astronautical Sciences," S. F. Singer, Ed., to appear in 1961.

11 Mengel, A. S., "Optimum Trajectories," Proceedings of Project Cyclone Symposium 1 on REAC Techniques, Reeves Instrument Corp. and USN Special Devices Center, N.Y., March 15–16, 1951.

12 Irving, J. H. and Blum, E. K., "Comparative Performance of Ballistic and Low-Thrust Vehicles for Flight to Mars," in "Vistas in Astronautics," vol. II, Pergamon Press, N. Y., 1959, pp. 191–218.

13 Faulders, C. R., "Low-Thrust Steering Program for Minimum Time Transfer Between Planetary Orbits," SAE National Aeronautic Meeting, Los Angeles, Calif., Sept. 29–Oct. 4, 1958.

14 Hadamard, J., "Mémoire sur le Problèm d'Analyse Relatif à l'Équilibre des Plaques Élastiques Encastrées," Mémoires presentés par divers savants éstrangers à l'Académie des Sciences de l'Institut de France (2), vol.

33, no. 4, 1908.
15 Curry, H. B., "The Method of Steepest Descent for Non-Linear Minimization Problems," *Quart. Appl. Math.*, vol. II, no. 3, Oct. 1944.
16 Hestenes, M. R. and Stiefel, E., "Method of Conjugate Gradients for Solving Linear Systems," *Nat. Bur. Standards, J. Research*, vol. 49, 1952, pp. 409–436.
17 Forsythe, A. I. and Forsythe, G. E., "Punched-Card Experiments with Accelerated Gradient Methods for Linear Equations," Nat. Bur. Standards, Applied Mathematics Series 39, Sept. 1954, pp. 55–69.
18 Müller, C., "A New Method for Solving Fredholm Integral Equations," New York University Institute of Mathematical Sciences, Division of Electromagnetic Research Rep. no. BR-15, 1955.
19 Courant, R., "Variational Methods for the Solution of Problems of Equilibrium and Vibrations," *Bull. Amer. Math. Soc.*, vol. 49, no. 1, Jan. 1943.
20 Stein, M. L., "On Methods for Obtaining Solutions of Fixed End-Point Problems in the Calculus of Variations," *Nat. Bur. Standards, J. Research*, vol. 50, no. 5, May 1953.
21 Friedman, B., "Principles and Techniques of Applied Mathematics," John Wiley and Sons, Inc., N. Y., 1956.
22 Goodman, T. R. and Lance, G. N., "The Numerical Integration of Two-Point Boundary Value Problems," *Mathematical Tables and Other Aids to Computation*, vol. 10, no. 54, April 1956.
23 Courant, R. and Hilbert, D., "Methoden der Mathematischen Physik," Julius Springer, Berlin, vol. 1, 1924, 1931. (Translated and revised edition of vol. 1, Interscience Publishers, N. Y., 1953.)
24 Tompkins, C. B., "Methods of Steep Descent," in "Modern Mathematics for the Engineer," E. F. Beckenbach, Ed., McGraw-Hill Book Co., Inc., N. Y., 1956, chap. 18.
25 Bliss, G. A., "Lectures on the Calculus of Variations," University of Chicago Press, Chicago, 1946.
26 Garwin, R. L., "Solar Sailing—A Practical Method of Propulsion Within the Solar System," JET PROPULSION, vol. 28, no. 3, March 1958.
27 Cotter, T. P., "Solar Sailing," Sandia Research Colloquium, April 1959.
28 Tsu, T. C., "Interplanetary Travel by Solar Sail," ARS JOURNAL, vol. 29, no. 6, June 1959.
29 London, H. S., "Some Exact Solutions of the Equations of Motion of a Solar Sail With Constant Sail Setting," ARS JOURNAL, vol. 30, no. 2, Feb. 1960.

Optimum Thrust Programming of Electrically Powered Rocket Vehicles in a Gravitational Field

C. R. FAULDERS[1]
North American Aviation, Inc.
Downey, Calif.

The general problem of optimum thrust programming of an electrically powered rocket under the condition of constant jet power is considered. The thrust vector is assumed to be parallel to the instantaneous velocity vector at all times. The various optimization problems possible under these restrictions are shown to be equivalent to maximization of the change in velocity for specified propellant mass and arbitrary range, maximization of range for specified propellant mass and arbitrary change in velocity, and maximization of the change in velocity for specified range and specified propellant mass. The calculus of variations is employed to obtain analytical expressions for the thrust acceleration program for the foregoing problems with a constant gradient of the tangential component of gravitational force. Limiting values of this gradient for which the tangential component of gravity can be assumed constant in the derivation of optimum thrust programs are determined.

ELECTRICALLY powered rocket engines, such as the ion rocket and the plasma rocket, are characterized by a power source that is separate from the propellant. For a fixed power setting, therefore, the rate of propellant expenditure and the exhaust velocity can be varied over wide ranges, and a variety of thrust programs can be achieved. Electrical rockets are presently limited to very low thrust levels of the order of 10^{-3} or 10^{-4} Earth g's per unit mass of the complete space vehicle. For this reason, an entire mission would generally be carried out under power, with operating times measured in days or weeks.

The various possible requirements for optimum thrust programs are summarized in Table 1, assuming that the total time of powered flight is a specified parameter, and that the thrust is always in the tangential direction. The change in vehicle velocity is indicated by ΔV, the change in position, or range, by Δs, and the propellant mass by m_p.

Table 1 Optimum thrust program requirements (total time specified)

Case	ΔV	Δs	m_p
1	maximum	arbitrary	specified
2	specified	arbitrary	minimum
3	arbitrary	maximum	specified
4	arbitrary	specified	minimum
5	maximum	specified	specified
6	specified	maximum	specified
7	specified	specified	minimum

Received Dec. 8, 1959.
[1] Research Specialist, Aero-Space Laboratories, Missile Division.