NASA CONTRACTOR REPORT

NASA CR-828

ANALYSIS AND DESIGN OF SPACE VEHICLE FLIGHT CONTROL SYSTEMS

VOLUME IX - OPTIMIZATION METHODS

by Arthur L. Greensite

Prepared by
GENERAL DYNAMICS CORPORATION
San Diego, Calif.
for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • JULY 1967

NASA CR-828

ANALYSIS AND DESIGN OF SPACE VEHICLE FLIGHT CONTROL SYSTEMS

VOLUME IX - OPTIMIZATION METHODS

By Arthur L. Greensite

Distribution of this report is provided in the interest of information exchange. Responsibility for the contents resides in the author or organization that prepared it.

Issued by Originator as GDC-DDE66-026

Prepared under Contract No. NAS 8-11494 by
GENERAL DYNAMICS CONVAIR
A DIVISION OF GENERAL DYNAMICS CORPORATION
San Diego, Calif.

for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION

For sale by the Clearinghouse for Federal Scientific and Technical Information
Springfield, Virginia 22151 – CFSTI price $3.00

FOREWORD

This report was prepared under NASA Contract NAS 8-11494 and is one of a series intended to illustrate methods used for the design and analysis of space vehicle flight control systems. Below is a complete list of the reports in the series:

Volume I	Short Period Dynamics
Volume II	Trajectory Equations
Volume III	Linear Systems
Volume IV	Nonlinear Systems
Volume V	Sensitivity Theory
Volume VI	Stochastic Effects
Volume VII	Attitude Control During Launch
Volume VIII	Rendezvous and Docking
Volume IX	Optimization Methods
Volume X	Man in the Loop
Volume XI	Component Dynamics
Volume XII	Attitude Control in Space
Volume XIII	Adaptive Control
Volume XIV	Load Relief
Volume XV	Elastic Body Equations
Volume XVI	Abort

The work was conducted under the direction of Clyde D. Baker, Billy G. Davis and Fred W. Swift, Aero-Astro Dynamics Laboratory, George C. Marshall Space Flight Center. The General Dynamics Convair program was conducted under the direction of Arthur L. Greensite.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

1. STATEMENT OF THE PROBLEM

The usual performance criteria for control systems are expressed in terms of undamped natural frequency, relative damping factor, gain and phase margin, steady-state error, etc. At a higher level of sophistication, one is often faced with the need to optimize the control system design from the point of view of minimum fuel expenditure, minimum time, maximum payload, or similar requirements. Optimal use of available resources is an implicit goal of every engineering design. The perennial demands for increased performance in aerospace control systems have not only revitalized the classical mathematics that provides the medium and foundation for optimality, but have also stimulated further research to improve available methods.

Certain problems in this area are well defined and can be expressed in a mathematical format well suited to the application of the classical variational calculus. Among these are the classical problems of programming rocket thrust to achieve maximum altitude(20) and attain minimum fuel transfer between circular orbits.(21) But a multitude of related optimality criteria for launch or orbital trajectories, such as maximum range, minimum fuel, and minimum time, lead to progressively more difficult analytical formulations and strain the capabilities of the classical variational methods. References 20 through 40 are only a representative list of studies in optimal trajectories.

While optimization criteria can generally be clearly defined for space trajectories, they are not so clearly delineated for control systems dealing with "short-period" dynamics. Here stability and constraints on the state variables are generally the paramount considerations. It is often difficult to relate natural mathematical optimality criteria with meaningful physical parameters. Thus, for example, it is often expedient to use some quadratic function of state or control variables as a mathematical measure of performance. However, it is not always possible to justify this choice by purely physical means. Indeed, any linear feedback system is optimal in the sense that it minimizes the integral of a quadratic function of state and control variables whose weighting factors are appropriately chosen. Nevertheless, recent studies have indicated that meaningful criteria can be established in terms of specified closed-loop response(14) or model-following.(13)

The basic theories for control system and trajectory optimization are identical; the difference is in fundamental time constants, which will generally differ by several orders of magnitude. It is therefore unnecessary to distinguish between the two in a basic exposition. In defining the scope of this monograph, we are guided first by the general theme of this series, which is the presentation of established and proven methods that have been shown to be useful in the design of aerospace control systems. Second, it is intended to reflect the current state of technical development. Subject to the limitations of space and mathematical sophistication expected of the reader, fundamental results will therefore be presented concisely and (hopefully) clearly, with a minimum of analytical abstractions. In addition to pedagogic examples, applications to realistic aerospace control systems will be discussed. Limitations and subtleties in the basic theory will be emphasized, especially as they affect basic design.

The discussion of the general analytical tools of optimal control theory (variational calculus, gradient methods, maximum principle, dynamic programming) will be supplemented by examples of application and by recent extensions and generalizations. The monograph is intended to summarize the present status of optimal control theory for aerospace control systems and to facilitate the study of the numerous problems in this area, as delineated in the extensive list of references.

2. STATE OF THE ART

An intelligent discussion of optimal control systems depends on defining an explicit measure of performance quality. Often this can be done only qualitatively. For example, a launch vehicle autopilot or satellite attitude control system may be required merely to exhibit "good" dynamic response characteristics, such as frequency, relative damping, and low sensitivity to extraneous disturbances. This is essentially a problem in conventional design. If further demands are made on the system — such as performing a stipulated mission in minimum time, or with minimum fuel or energy expenditure — a more sophisticated analysis is required.

There are essentially two aspects to the problem of optimal control. One is the proper formulation of the problem (which is not always trivial); the other is the development or application of an appropriate tool to solve the problem.

The first is more an art than a science, since the formulation of a mathematical model for a real system must be a compromise between the needs for analytical tractability and for including all significant dynamic features. The results of an analysis must therefore be interpreted with due regard for these factors.

This monograph is concerned primarily with the analytic tools for solving optimization problems. The natural vehicle for this exposition is the classical calculus of variations. Broadly speaking, the problem is formulated as follows. Assume a system whose dynamic properties are described by a set of ordinary differential equations. Some type of initial and terminal conditions are prescribed. One or more degrees of freedom are available in the form of a control function that must be programmed such that a stipulated function of the control and state variables is minimized (or maximized).

In this fashion, we have the classical problem of Mayer†, the solution of which is well established. Some qualifications are in order, however. First, this "solution" is often in the form of a system of nonlinear differential equations with two-point boundary conditions. The computational difficulties associated with this are far from trivial and often overwhelming. Second, the classical formulation includes no constraints on the state or control variables. This severely limits its usefulness, since, in the real world, bounds on energy and magnitude of a control force are hard realities. In some instances, a state variable must be contained within prescribed bounds — such as limiting an angle of attack to ensure structural integrity.

As a result of these limitations in the classical techniques, several new approaches to the problem have been initiated in the last 10 years. These have led to the development of dynamic programming, the gradient method, the maximum principle, and the

† Also called the problem of Lagrange or Bolza, when expressed in slightly different fashion.

generalized Newton-Raphson technique. With this new body of theory, the range of problems that can be resolved has been considerably enlarged. But as usual, few panaceas are available.

Dynamic programming, which exhibits the greatest indifference to the presence of "pathological" functions, is limited by computer storage capacity.

The gradient method, one of the more powerful techniques, generally yields only local extrema, primarily because small deviations from a reference condition are analyzed.

The maximum principle, in which is included the effect of control constraints, is a more elegant formulation of the classical Weierstrauss condition in variational calculus. It is, however, completely identical to the classical methods when the inequality constraints are included via the Valentine condition.[99]

The generalized Newton-Raphson technique is a powerful new tool for solving nonlinear systems with two-point boundary conditions. It has been remarkably successful in solving problems that strained the capability of conventional methods. This, of course, has served to invigorate the classical approach, which is attractive because of its analytical elegance. The weakness of the method is that only sufficient conditions for convergence are known. Even these are hard to apply. It is known, however, that the method works in cases where the sufficiency condition is not satisfied. (In other words, this sufficient condition may not be necessary.) Consequently, a certain amount of "cut-and-try" is necessary in practice.

For any given problem, therefore, one may select one of a multitude of superficially unrelated methods to achieve optimization. Generally, the form of the problem or the type of results desired will indicate the method to be used. A fuller understanding of the virtues and limitations of each method requires recognition that all these methods are implicitly (if not explicitly) related. We have therefore not derived any of these results in the conventional manner but have shown (Sec. 3.1.5) that each is a consequence of one general formulation.

The examples of application of the theory are an attempt to balance simplicity of exposition with practical realism. These examples, together with a sampling of the literature cited in the references, provide a fair indication of the current state of the art.

3. RECOMMENDED PROCEDURES

The basic problem of optimal control can be stated generally as follows. Assume a dynamic system described by some appropriate set of differential equations. A parameter or set of parameters (controls) is to be programmed in such a way that a prescribed measure of performance takes on an extremal (maximum or minimum) value. Usually, there are prescribed values for the initial and/or final state of the system. Furthermore, there may be constraints, dictated by physical considerations, on the control or state variables. Taking account of all these factors, we seek to determine the control function form that ensures that the measure of performance is maximized, or minimized, as the case may be.

There are many aspects to this problem depending on the complexity of the system, the form of the constraints, and the type of performance criterion adopted. Furthermore a solution may be characterized as: open-loop, wherein the control is obtained as a function of time and therefore of the initial conditions; or closed-loop, in which case the control is a function of the current state of the system.

All these considerations directly affect the mathematical complexity and the ease with which a solution may be obtained. This section presents the essential features of standard optimization techniques, beginning with the simplest problems expressed in the classical variational format. This is followed by a discussion of the classical theory limitations that motivated some of the modern techniques, such as the maximum principle, gradient methods, and dynamic programming.

Typical applications and standard solutions are then described. The concluding sections deal with recent theoretical developments, together with typical aerospace applications.

3.1 MATHEMATICAL CONCEPTS

The mathematical features of the basic optimization techniques are described in Sections 3.1.1 through 3.1.4. In general, the treatment is concise but explicit; the primary aim is to facilitate application of the theory to practical engineering problems. Elaborate motivations and mathematical abstractions are therefore avoided. Illustrative examples are used liberally to enhance comprehension of the basic ideas.

3.1.1 Variational Calculus

For the solution of optimization problems of a more general character than is possible via the elementary calculus, the classical tool is the calculus of variations. Most problems of interest to the aerospace controls engineer can be formulated in terms of the so-called Mayer problem(134), which includes a variety of related problems as special cases. It is therefore of extreme generality and will accordingly be described in some detail.

The Mayer problem is usually stated as follows: there are n functions, $x_k(t)$, of an independent variable, t, that satisfy the differential constraints

$$\varphi_j (\dot{x}, x, t) = 0 \qquad j = 1, 2, \cdots\cdot, p\,(< n) \tag{1}$$

$$x(t) \equiv n \text{ vector}$$

subject to boundary conditions of the type

$$\psi_r \left(x_i, x_f, t_i, t_f\right) = 0 \qquad r = 1, 2, \cdots\cdot, q\;(\leqq 2n + 2) \tag{2}$$

where t_i and t_f are the initial and final times respectively, and†

$$x_i \equiv x\,(t_i)$$

$$x_f \equiv x\,(t_f)$$

$$(\dot{\ }) \equiv \frac{d}{dt}\,(\ )$$

It is required to choose the functions, $x_k\,(t)$, such that the quantity

$$J = \Big[G\,(x, t)\Big]_i^f$$

$$\equiv G\,(x, t)\big|_{t=t_f} - G\,(x, t)\big|_{t=t_i} \tag{3}$$

is minimized subject to the constraints (1) and (2).

†The subscripts i and f are used exclusively to denote initial and final value respectively, not the components of a vector. Symbols other than i and f will be used to denote vector components.

The problem thus formulated is more general than is superficially apparent. For example, in the problem of Lagrange, the function to be minimized is

$$J = \int_{t_i}^{t_f} L\ (\dot{x}, x, t)\, dt \qquad (4)$$

If we define an auxiliary variable by

$$x_{n+1}(t) = \int_{t_i}^{t} L\ (\dot{x}, x, t)\, dt \qquad (5)$$

with

$$x_{n+1}(t_i) = 0 \qquad (6)$$

then we reduce to a Mayer-type problem such that

$$J = x_{n+1}(t_f) \qquad (7)$$

with the additional constraint

$$\varphi_{n+1} \equiv \dot{x}_{n+1}(t) - L\ (\dot{x}, x, t) = 0 \qquad (8)$$

In the problem of Bolza, the function to be minimized is

$$J = \Big[G\ (x, t)\Big]_i^f + \int_{t_i}^{t_f} L\ (\dot{x}, x, t)\, dt \qquad (9)$$

We again reduce this to a Mayer-type problem by defining an auxiliary variable in the manner shown above.

Inequality Constraints

If there exist inequality constraints of the form[†]

$$K_1 \lessgtr \dot{x}_k \lessgtr K_2 \qquad (10)$$

[†] The discussion also applies if the inequality constraint is on x_k instead of $\dot{x}_k$.

then by introduction of a new variable, $x_{n+1}(t)$, defined by(99)

$$\left(\dot{x}_k - K_1\right)\left(K_2 - \dot{x}_k\right) - x_{n+1}^2 = 0 \tag{11}$$

the problem is put in the Mayer format, where J, defined by Eq. (3), is minimized subject to the added differential constraint

$$\varphi_{n+1} \equiv \left(\dot{x}_k - K_1\right)\left(K_2 - \dot{x}_k\right) - x_{n+1}^2 = 0 \tag{12}$$

The Mayer format is therefore extremely general and finds wide application in aerospace controls problems. A solution is obtained in the following way. Form the augmented function

$$F = \sum_{j=1}^{p} \lambda_j \varphi_j \tag{13}$$

where the λ_j are time-dependent functions called the Lagrange multipliers. A necessary condition for the minimization of the criterion function defined by Eq. (3) is that the Euler Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}_k}\right) - \frac{\partial F}{\partial x_k} = 0 \qquad k = 1, 2, \cdots, n \tag{14}$$

be satisfied.

The system of equations (1) and (14) is subject to (2n + 2) boundary conditions. Of these, q are supplied by Eq. (2), while the remaining ones are determined from the transversality condition(134)

$$\left[\delta G + \left(F - \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k\right) \delta t + \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \delta x_k\right]_i^f = 0 \tag{15}$$

where $\delta(\ )$ is the variation operator. In general,

$$\delta G = \sum_{k=1}^{n} \frac{\partial G}{\partial x_k} \delta x_k + \frac{\partial G}{\partial t} \delta t \tag{16}$$

which means that Eq. (15) can be expressed as

$$\left[\left(\frac{\partial G}{\partial t} + F - \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k\right)\delta t + \sum_{k=1}^{n} \left(\frac{\partial G}{\partial x_k} + \frac{\partial F}{\partial \dot{x}_k}\right)\delta x_k\right]_i^f = 0 \tag{17}$$

This expression splits up into (2n + 2) subconditions of the form

$$\left.\left(\frac{\partial G}{\partial t} + F - \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k\right)\right|_{t=t_i} \delta t_i = 0 \tag{18}$$

$$\left.\left(\frac{\partial G}{\partial t} + F - \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k\right)\right|_{t=t_f} \delta t_f = 0 \tag{19}$$

$$\left.\left(\frac{\partial G}{\partial x_k} + \frac{\partial F}{\partial \dot{x}_k}\right)\right|_{t=t_i} \delta x_{ki} = 0 \tag{20}$$

$$k = 1, 2, \cdots, n$$

$$\left.\left(\frac{\partial G}{\partial x_k} + \frac{\partial F}{\partial \dot{x}_k}\right)\right|_{t=t_f} \delta x_{kf} = 0 \tag{21}$$

$$k = 1, 2, \cdots, n$$

If the boundary condition for a particular variable, say x_{kf}, is not prescribed, then δx_{kf} is arbitrary, which means that

$$\left.\left(\frac{\partial G}{\partial x_k} + \frac{\partial F}{\partial \dot{x}_k}\right)\right|_{t=t_f} = 0 \tag{22}$$

Notice that boundary values cannot be assigned to nonderivated variables. These are, in fact, a mathematical consequence of the constraining equations (1) and the Euler Lagrange equations (14).

If the augmented function, F, does not contain t explicitly, then it can be shown that†

† Ref. 134, p. 205.

$$-F + \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k = C \equiv \text{constant} \tag{23}$$

This means that Eq. (17) simplifies to

$$\left[\left(\frac{\partial G}{\partial t} - C\right)\delta t + \sum_{k=1}^{n} \left(\frac{\partial G}{\partial x_k} + \frac{\partial F}{\partial \dot{x}_k}\right)\delta x_k\right]_i^f = 0 \tag{24}$$

and the set of equations (18) through (21) is simplified accordingly.

The solution of certain variational problems is characterized by the fact that one or more of the derivatives $\dot{x}_k$ are discontinuous. For example, in a rocket vehicle, the acceleration is discontinuous at the point where thrust is terminated or staging occurs. In this case, a mathematical criterion is needed to join the portions of the extremal arc† at points of discontinuity. This is supplied by the Erdmann-Weierstrass corner conditions(134)

$$\left(\frac{\partial F}{\partial \dot{x}_k}\right)_- = \left(\frac{\partial F}{\partial \dot{x}_k}\right)_+ \tag{25}$$

$$\left(-F + \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k\right)_- = \left(-F + \sum_{k=1}^{n} \frac{\partial F}{\partial \dot{x}_k} \dot{x}_k\right)_+ \tag{26}$$

where the negative and positive signs denote conditions immediately before and after a point of discontinuity.

For the special case in which F does not depend explicitly on t, the last relation reduces to

$$(C)_- = (C)_+ \tag{27}$$

by virtue of (23). In other words, the integration constant, C, has the same value for all subarcs composing the extremal arc.

Satisfying the Euler Lagrange equations, (14), merely ensures that function J, defined by Eq. (3), is either a maximum or a minimum. More precisely, it is a local extremum; in other words, it may be only one of many extremal functions. This

† The extremal arc is that curve in n space characterized by x_1 (t), ····, x_n (t), which extremizes J.

corresponds roughly to the case of a curve in elementary calculus whose derivative vanishes at several points. The fact that the slope is zero gives no information as to whether the point on the curve is a maximum or minimum or whether it is indeed the global maximum or minimum. In the variational calculus, the necessary and/or sufficient conditions for type of extrema are not easily applied. Perhaps the most useful for engineering purposes is the Legendre-Clebsch condition,[134] which states that the relation

$$\sum_{k=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 F}{\partial \dot{x}_k \, \partial \dot{x}_j} \delta \dot{x}_k \delta \dot{x}_j \gtrless 0 \tag{28}$$

must be satisfied if J is a minimum. If the derivative, $\dot{x}_r$, of function x_r does not appear explicitly in F, then in (28), we replace $\dot{x}_r$ with x_r.

The above condition is merely necessary (not sufficient) to ensure that J is a minimum. Furthermore, only a local minimum is ensured. Fortunately, in most cases of engineering interest, the type of extremal arc obtained is determined from physical or numerical considerations, and only rarely does an ambiguity exist between local and global extrema.

The methods discussed above are illustrated in the following examples.

Example 1, The Sounding Rocket: This is one of the earliest aerospace problems treated by variational methods.[5,20,150] In the usual formulation, we seek the thrust-time history such that a maximum height is attained for a given propellant weight. Alternatively, we may seek the minimum propellant weight to attain a specfied altitude. The relevant differential equations are

$$T - D = m(\dot{V} + g) \tag{29}$$

$$\dot{h} = V \tag{30}$$

$$T = \beta \mu \tag{31}$$

$$\beta = -\dot{m} \tag{32}$$

Here

$T \equiv$ thrust

$D \equiv D(V, h) \equiv$ drag

$m \equiv$ instantaneous mass

$h \equiv$ altitude

$V \equiv$ velocity

$g \equiv$ gravity acceleration

$\mu \equiv$ exit velocity of burned propellant

$\beta \equiv$ propellant mass flow

Physical considerations require that the mass flow satisfy the inequality

$$0 \mathrel{\overline{<}} \beta \mathrel{\overline{<}} \beta_M \tag{33}$$

where β_M is a given constant.

The problem is easily formulated as a Mayer type as follows. The differential constraints are

$$\varphi_1 \equiv \dot{h} - V = 0 \tag{34}$$

$$\varphi_2 \equiv \dot{V} + g + \frac{(D - \beta\mu)}{m} = 0 \tag{35}$$

$$\varphi_3 \equiv \dot{m} + \beta = 0 \tag{36}$$

$$\varphi_4 \equiv \beta\left(\beta_M - \beta\right) - \alpha^2 = 0 \tag{37}$$

The last relation, which is in the form of Eq. (12), accounts for the inequality constraint (33). No special physical significance is attached to the variable, α.

The problem variables are identified as follows.

$x_1 \equiv h$

$x_2 \equiv V$

$x_3 \equiv m$

$x_4 \equiv \beta$

$x_5 \equiv \alpha$

The augmented function is therefore

$$F = \lambda_1 (\dot{h} - V) + \lambda_2 \left(\dot{V} + g + \frac{D - \beta\mu}{m}\right) + \lambda_3 (\dot{m} + \beta) + \lambda_4 \left[\beta\left(\beta_M - \beta\right) - \alpha^2\right] \tag{38}$$

and the Euler Lagrange equations become

$$\dot{\lambda}_1 = \frac{\lambda_2}{m} \frac{\partial D}{\partial h} \tag{39}$$

$$\dot{\lambda}_2 = -\lambda_1 + \frac{\lambda_2}{m} \frac{\partial D}{\partial V} \tag{40}$$

$$\dot{\lambda}_3 = \frac{(\beta\mu - D)\lambda_2}{m^2} \tag{41}$$

$$0 = -\left(\frac{\mu}{m}\lambda_2 - \lambda_3\right) + \lambda_4 \left(\beta_M - 2\beta\right) \tag{42}$$

$$0 = \alpha\lambda_4 \tag{43}$$

Since F does not contain t explicitly, we have, from Eq. (23), combined with the Euler Lagrange Eqs., (42) and (43),

$$\lambda_1 V - \left(g + \frac{D}{m}\right)\lambda_2 + \left(\frac{\mu}{m}\lambda_2 - \lambda_3\right)\beta = C \tag{44}$$

where C is an integration constant.

It now remains to formulate the criterion function, J, and the boundary conditions, ψ_r. If we seek to maximize the altitude attained, subject to the boundary conditions

$$\left.\begin{aligned} h_i &= 0 & t_i &= 0 \\ V_i &= 0 & V_f &= 0 \\ m_i &= m_0 & m_f &= m_p \end{aligned}\right\} \tag{45}$$

then J becomes[†]

$$J = -h_f \tag{46}$$

Since t_f and h_f are not prescribed, the transversality condition, (17), yields

$$\left(\frac{\partial G}{\partial t} - C\right)\Bigg|_{t=t_f} \delta t_f = 0 \tag{47}$$

$$\left(\frac{\partial G}{\partial h} + \frac{\partial F}{\partial h}\right)\Bigg|_{t=t_f} \delta h_f = 0 \tag{48}$$

The first of these relations implies that

$$C = 0 \tag{49}$$

since G (= -h) is independent of t and δ t_f is arbitrary. Eq. (48) leads to[‡]

$$-1 + \lambda_{1f} = 0 \tag{50}$$

According to Eq. (43), an extremal arc is composed of subarcs along which $\alpha = 0$ or $\lambda_4 = 0$. The first of these implies that $\beta = 0$ or $\beta = \beta_M$ by virtue of (37). If we write

$$K = \frac{\mu}{m} \lambda_2 - \lambda_3 \tag{51}$$

then the condition $\lambda_4 = 0$ and Eq. (42) leads to

$$K = 0 \text{ and } \dot{K} = 0 \tag{52}$$

Taking the time derivative of (51) and combining with the Euler Lagrange equations, we obtain

$$m(D + mg)\dot{K} = \beta\left(\mu \frac{\partial D}{\partial V} + D\right) K + \lambda_1 \left[(V - \mu) D + \left(V \frac{\partial D}{\partial V} - mg\right)\mu\right] \tag{53}$$

[†] Obviously, the problem of minimizing ($-h_f$) is equivalent to that of maximizing h_f.

[‡] The notation λ_{1f} stands for λ_1 (t_f).

after making use of Eq. (44).

Consequently, along a variable thrust subarc, the relation

$$(V - \mu) D + \left(V \frac{\partial D}{\partial V} - mg\right)\mu = 0 \tag{54}$$

must be satisfied. This therefore determines the variable thrust parameter, $\beta = -\dot{m}$. In order to ascertain the physical significance of this equation, consider a drag function given by

$$D = c_1 V^2, \; c_1 = \text{positive constant} \tag{55}$$

Combining this with Eq. (54), we find

$$m = \frac{c_1 V^2}{g} \left(\frac{V}{\mu} + 1\right)$$

This shows that a decrease in mass is accompanied by a decrease in velocity; in other words, the acceleration is negative. Consequently, the variable thrust arc is such that the vehicle is flown with thrust always less than drag during this phase.

It now remains to determine how the subarcs for zero, maximum, and variable thrust are connected such that the optimality conditions are satisfied and in a manner that is consistent with the given boundary conditions. To do this, we make use of the Weierstrass Erdmann corner conditions and the Legendre Clebsch conditions. Eqs. (25) and (26) show that at each corner point

$$\begin{aligned} (\lambda_1)_- &= (\lambda_1)_+ \\ (\lambda_2)_- &= (\lambda_2)_+ \\ (\lambda_3)_- &= (\lambda_3)_+ \\ (C)_- &= (C)_+ \end{aligned} \tag{56}$$

That is, the above parameters are continuous at the corner points of the extremal arc.

Now since C = 0 and since V and m are also continuous, Eq. (44) shows that β may be discontinuous provided that

$$(K)_- = (K)_+ = 0 \tag{57}$$

which implies that

$$(\dot{K})_{-} = (\dot{K})_{+} = 0 \tag{58}$$

Furthermore, the Legendre Clebsch condition (28) leads to

$$-2\lambda_4 \left[(\delta\,\alpha)^2 + (\delta\,\beta)^2\right] \geq 0$$

Therefore, when $\lambda_4 \neq 0$, we have by virture of (42)

$$-\frac{2K}{\beta_M - 2\beta}\left[(\delta\,\alpha)^2 + (\delta\,\beta)^2\right] \geq 0$$

or, equivalently,

$$-\frac{K}{\beta_M - 2\beta} \geq 0 \tag{59}$$

It was noted earlier that $\lambda_4 \neq 0$ implies that $\beta = 0$ or $\beta = \beta_M$. Consequently, the above inequality shows that

$$\begin{aligned} &\beta = \beta_M \quad \text{when } K > 0 \\ &\beta = 0 \qquad \text{when } K < 0 \\ &0 < \beta < \beta_M \quad \text{when } K = 0 \end{aligned} \tag{60}$$

The last relation follows from the fact that for $\lambda_4 = 0$, $\alpha \neq 0$ which means that β may assume an intermediate value as shown by Eq. (37). The quantity β (= $-\dot{m}$) is in fact determined from Eq. (54).

It now appears that the solution is completely determinate. The six equations, (34) through (36) and (39) through (41) together with the six boundary conditions

$$\begin{array}{ll} h_i = 0 & \lambda_{1f} = 1 \text{ from Eq. (50)} \\ V_i = 0 & V_f = 0 \\ m_i = 0 & m_f = m_p \end{array}$$

may be integrated to yield m, V, and h as well as the Lagrange multipliers. Quantity K, defined by Eq. (51), is calculated in the course of the solution and determines the thrust switching points in accordance with Eq. (60).

It is tempting to consider the problem as now solved. However, from a practical point of view, several factors disturb this state of euphoria. First of all, the difficulties associated with solving a system of nonlinear differential equations subject to two-point boundary conditions are too well known to be belabored here. In the present case, one must guess the initial values of λ_1, λ_2, and λ_3 in Eqs. (39) through (41) and hope that the final boundary conditions are satisfied. Various iterative procedures are available for adjusting the initial guesses to ensure that one converges to the stipulated final values. This is not always possible, since poor guesses lead to numerical instabilities, with the result that the engineer is often led to seeking other ways of making a living. Nevertheless, the practical importance of this problem has stimulated extensive research for improved methods. Some promising approaches are discussed in Appendix A.

In the present problem, it is not necessary to solve the equations of motion in order to determine the general features of the optimal thrust program. It does require, however, that certain simplifications be introduced. Specifically, we assume that the drag is given by Eq. (55). Then, since D is independent of h, Eq. (39) reduces to $\dot{\lambda}_1 = 0$, which means that $\lambda_1 = 1$ by virtue of Eq. (50). As a result, Eqs. (53) and (54) may be written as

$$m\left(c_1 V^2 + m g\right)\dot{K} = c_1 \beta V (2\mu + V) K + Q \tag{61}$$

$$Q \equiv c_1 V^2\left(\frac{V}{\mu} + 1\right) - m g = 0 \tag{62}$$

We may plot the variable thrust condition, (62), in the m–V plane as shown in Fig. 1. The initial and final points on the trajectory are here designated by ① and ⑬ respectively. It is obvious that an extremal arc begins with maximum thrust from point ① (corresponding to the initial conditions $m_i = m_0$, and $V_i = 0$). The rest of the figure shows various possibilities of using maximum thrust, coasting, and variable thrust to arrive at the terminal point, ⑬, which is obviously approached via a coasting arc (constant mass). We will show that assuming the quadratic drag law given by Eq. (55) leads to a unique optimal thrust program consisting of the three subarcs shown in Fig. 2. Consider first the possible path ① - ② - ③ - ⑤ shown in Fig. 1. Along the maximum thrust arc, ① - ②, we must have $K > 0$, while along the coasting arc, ② - ③, $K < 0$ as required by the optimal switching conditions, (60).

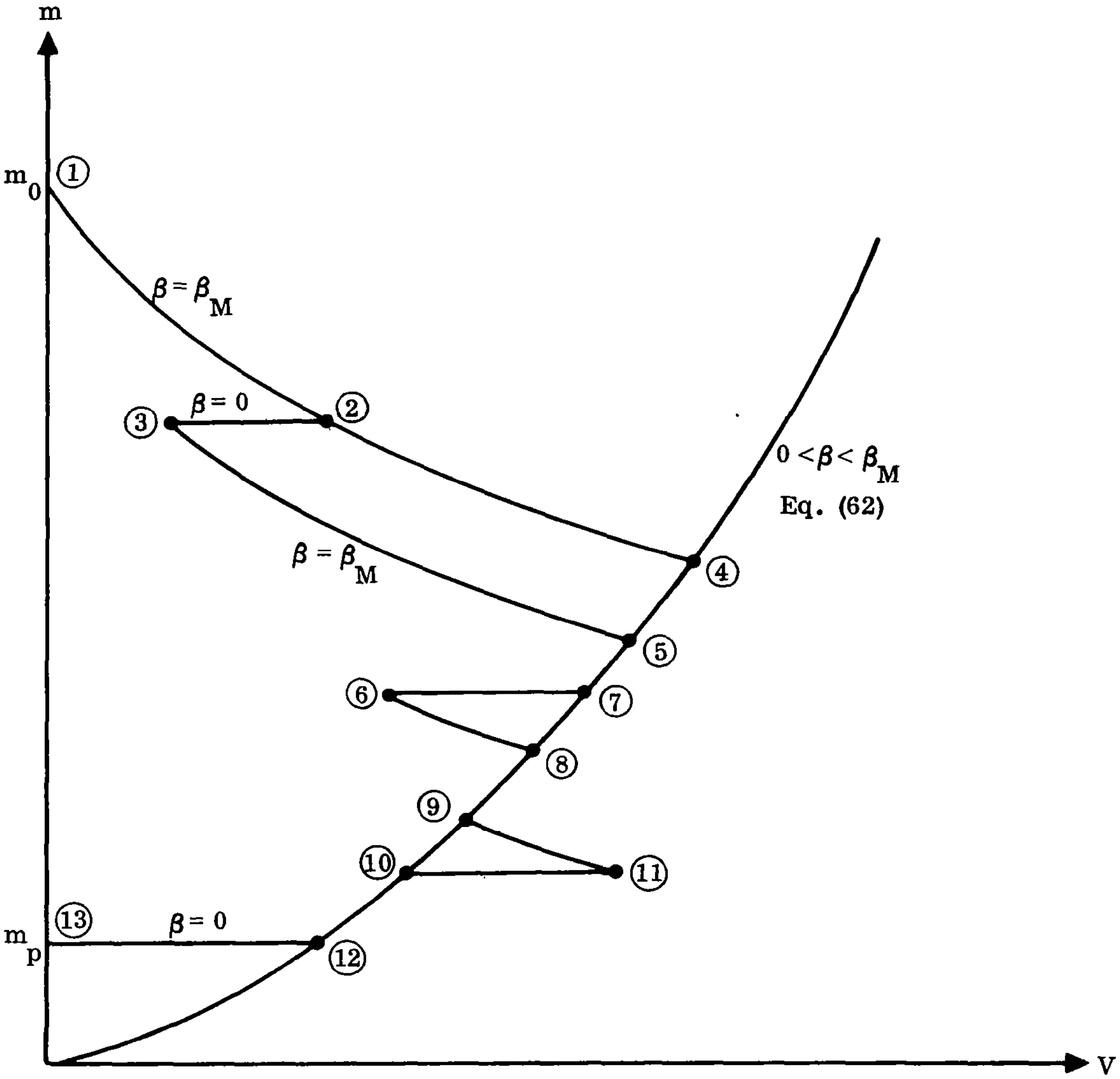


Figure 1. Extremal Thrust Arcs in m-V Plane

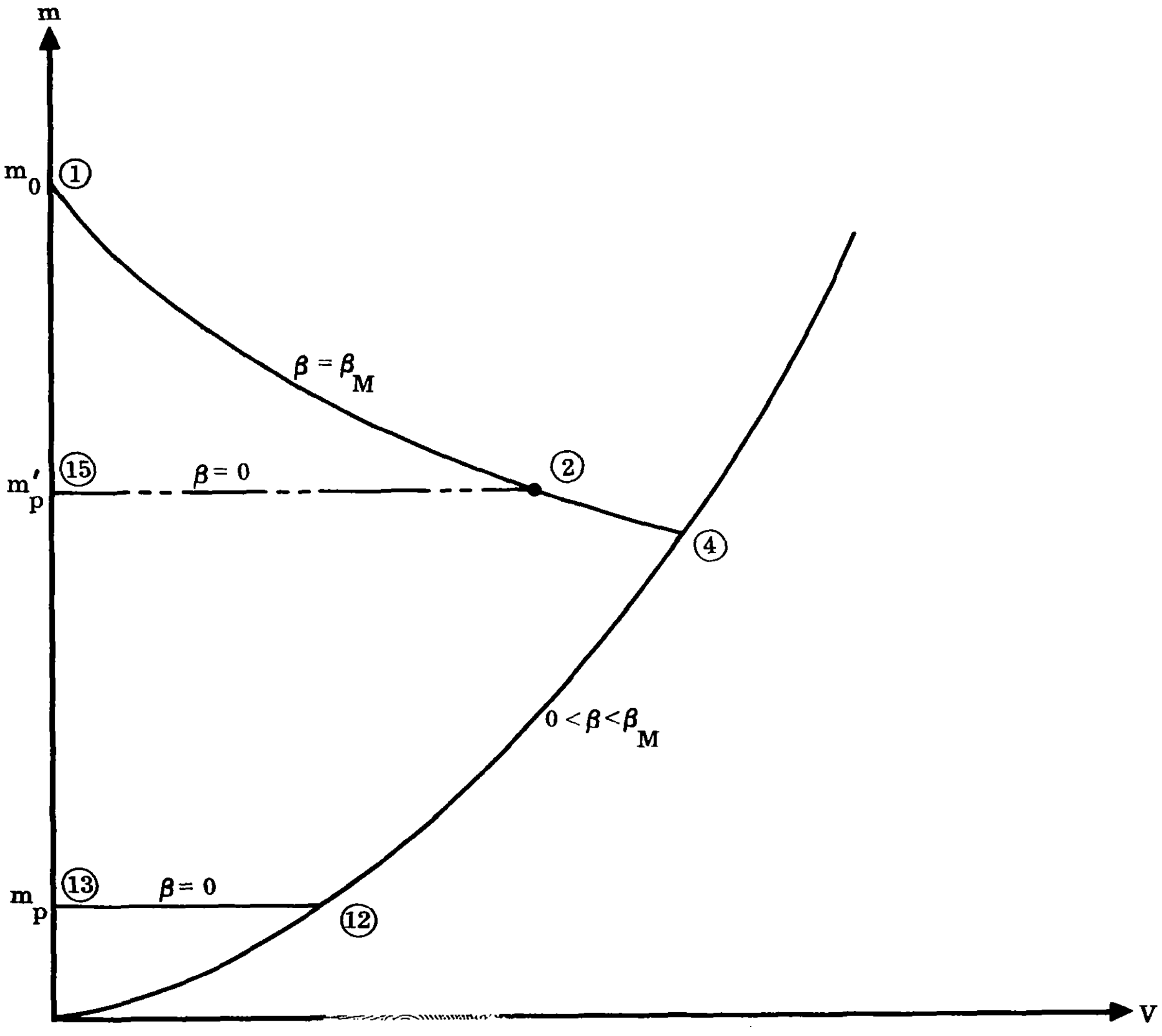


Figure 2. Unique Thrust Program for Quadratic Drag Law

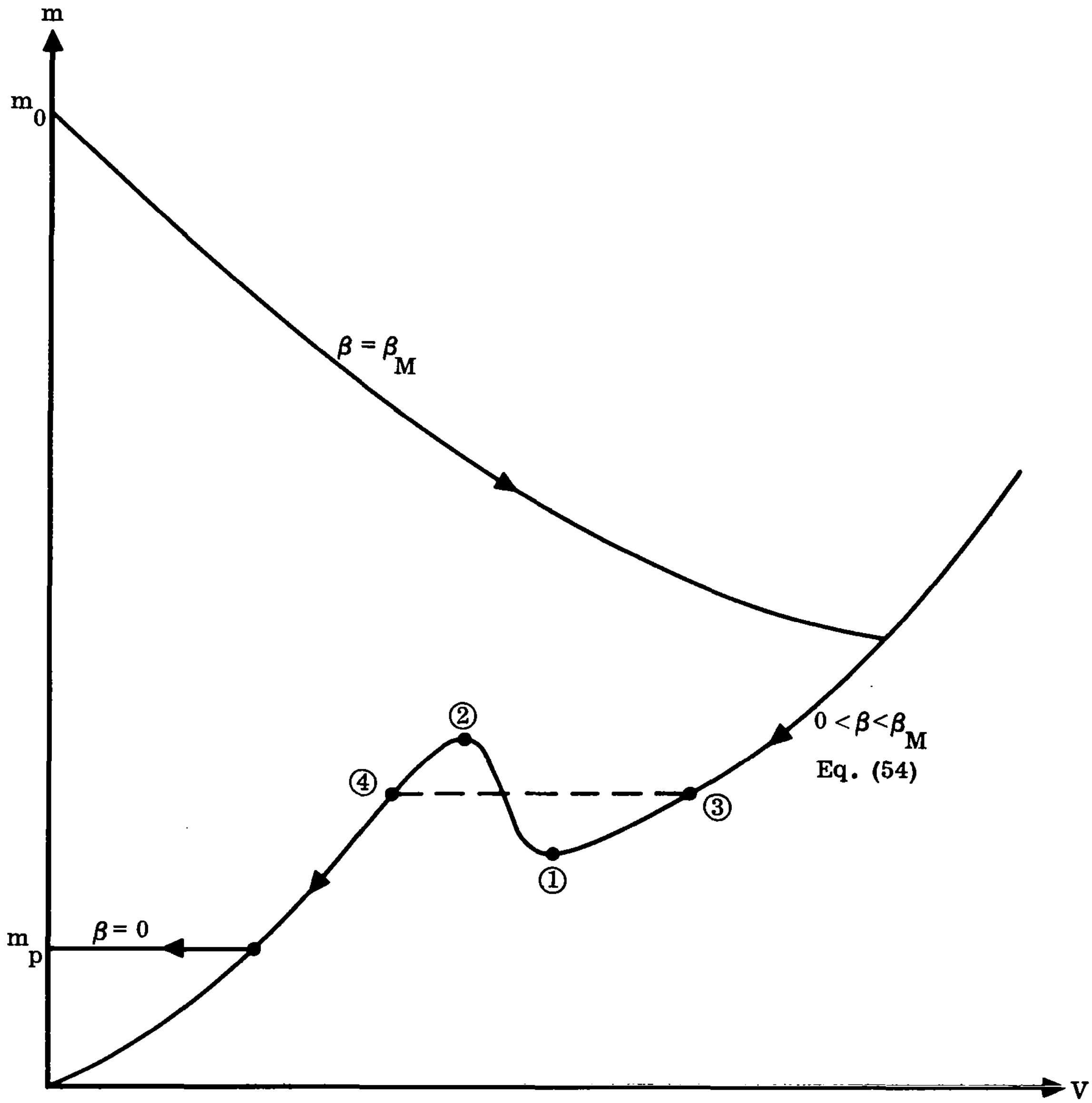


Figure 3. Possible Ambiguity in Thrust Program for General Drag Law

But since Eq. (57) shows that K is continuous at a switching point, it follows that

at point ②: $K = 0,\ \dot{K} < 0$

at point ③: $K = 0,\ \dot{K} > 0$

In this case, Eq. (61) indicates that $Q < 0$ at point ② and $Q > 0$ at point ③. Expressed in terms of Eq. (62)

$$c_1 V_2^2 \left(\frac{V_2}{\mu} + 1\right) < m_2 g \qquad \text{at point ②}$$

$$c_1 V_3 \left(\frac{V_3}{\mu} + 1\right) > m_2 g \qquad \text{at point ③}$$

According to Fig. 1, $V_3 < V_2$, so that the above two conditions are incompatible. By similar arguments, we eliminate the subarcs ⑦ - ⑥ - ⑧ and ⑨ - ⑪ - ⑩. The optimal path must therefore be as shown in Fig. 2, composed of a maximum, followed by variable thrust, and then a coasting arc.

If the final mass is sufficiently large, then there will be only a path of maximum thrust followed by coasting, as shown by path ① - ② - ⑮ in Fig. 2.

It may be shown that the solution is not completely determinate in all cases. Consider, for example, the situation shown in Fig. 3. The peak and trough in the variable thrust curve, Eq. (54), may be due to the peculiarities of the drag function near Mach 1. A variable thrust along ① - ② is not admissible, since this corresponds to increasing mass. Consequently, there must be some intermediate coasting path, ③ - ④. However, the precise location of this path cannot be determined from the present theory.

We have analyzed this example in some detail, since it highlights both the virtues and the limitations of the classical variational calculus. The theory is well adapted for solving relatively simple (textbook) problems but suffers from severe limitations of a computational nature when applied to moderately realistic practical problems. This motivated the development of the so-called direct methods (gradients, dynamic programming), which generally exhibit marked computational advantages over the variational approach. Before turning to these, we will first discuss the Pontryagin maximum principle, in which the classical variational calculus is highly systematized, and which is therefore much simpler to apply in practice.

3.1.2 Maximum Principle

A wide class of optimum control problems can be formulated as follows. Given the system

$$\dot{x} = f(x, u, t) \qquad (63)$$

where x is an n-dimensional state vector and u is an m-dimensional control vector. The components of the latter are subject to constraints of the form

$$\nu_j \lesssim u_j \lesssim \mu_j \qquad j = 1, 2, \cdots, m \tag{64}$$

where the ν_j and μ_j are known constants. It is required to determine u (t) in the interval, $t_i \lesssim t \lesssim t_f$, such that the criterion function

$$J = c^T x_f \tag{65}$$

is a minimum (maximum). Here c is a constant n vector that is known, and the notation x_f means x (t_f). In other words, we seek to minimize (maximize) a prescribed linear combination of the final value of the state vector components.

As noted in Sec. 3.1.1, many types of criterion function can be reduced to form (65). Thus, if it is required to minimize (maximize) the integral

$$\int_{t_i}^{t_f} L\,(x, u, t)\, dt \tag{66}$$

we define a new state variable by

$$x_{n+1}\,(t) = \int_{t_i}^{t} L\,(x, u, t)\, dt \tag{67}$$

with

$$x_{n+1}\,(t_i) = 0 \tag{68}$$

and add the equation

$$\dot{x}_{n+1} = L\,(x, u, t) \tag{69}$$

to system (63). With these modifications, the problem is reduced to that of minimizing (maximizing) the quantity

$$J = x_{n+1}\,(t_f) \tag{70}$$

Furthermore, the problem of minimizing (maximizing) some function of the final state vector, $\zeta\ (x_f)$, reduces to the same format if we define an additional state variable

$$x_{n+1} = \zeta\ (x) \tag{71}$$

with †

$$x_{n+1}\ (t_i) = \zeta\ (x_i) \tag{72}$$

and add to the system, (63), the equation

$$\dot{x}_{n+1} = \sum_{j=1}^{n} f_j \frac{\partial \zeta}{\partial x_j} \tag{73}$$

The new problem is then reduced to that of minimizing (maximizing) the quantity

$$J = x_{n+1}\ (t_f) \tag{74}$$

In order to avoid any possible ambiguity due to the notation adopted, we emphasize that the subscripts i and f are used exclusively to denote initial and final values respectively. They are not to be interpreted as components of a vector. Vector components will be designated by subscripts j, k, etc. Thus

$$x_i \equiv \begin{bmatrix} x_1\ (t_i) \\ x_2\ (t_i) \\ \vdots \\ x_n\ (t_i) \end{bmatrix}$$

$x_j \equiv j^{th}$ component of the vector x

$x_{jf} \equiv x_j\ (t_f)$, etc.

One frequently encounters the so-called "minimum-time" problem, in which it is required to transfer the system from a given initial to a given final state in minimum

† The symbol x_i means $x\ (t_i)$.

time. Mathematically, this requirement is expressed as that of minimizing the integral

$$\int_{t_i}^{t} dt \tag{75}$$

By introducing a new state variable such that

$$\dot{x}_{n+1} = 1 \,, \quad x_{n+1}(t_i) = 0 \tag{76}$$

$$x_{n+1} = \int_{t_i}^{t} dt \tag{77}$$

the problem is reduced to one of minimizing the final value of x_{n+1}.

In what follows, it will be assumed that all necessary transformations have been accomplished; we will deal exclusively with the set of equations (63) through (65).

We now define the components of a vector, λ, as follows†.

$$\dot{\lambda}_j = - \sum_{k=1}^{n} \lambda_k \frac{\partial f_k}{\partial x_j} \tag{78}$$

$$j = 1, 2, \cdots, n$$

We also define the function

$$H = \sum_{j=1}^{n} \lambda_j f_j \equiv \lambda^T f \tag{79}$$

One sometimes refers to λ as the costate vector; the function, H, is known as the hamiltonian.

The maximum principle of L. S. Pontryagin(4) is then stated as follows:

"In the system described by Eq. (63), if there exists a control vector, u, which satisfies the constraints, (64), while minimizing (maximizing) the criterion function, (65), then this control vector necessarily maximizes (minimizes) the hamiltonian (79)."

†Our notation anticipates the interpretation of this quantity as the Lagrange multiplier.

Making use of the definition of the hamiltonian, Eq. (79), we may write Eqs. (63) and (78) as

$$\dot{x}_j = \frac{\partial H}{\partial \lambda_j} \qquad j = 1, 2, \cdots, n \tag{80}$$

$$\dot{\lambda}_j = - \frac{\partial H}{\partial x_j} \qquad j = 1, 2, \cdots, n \tag{81}$$

These are identical in form to the Hamilton canonical equations in analytical mechanics (which accounts for the name accorded the function H).

In application, the extremal control vector, u, is obtained as a function of λ after the minimizing (maximizing) operation on H is performed. A completely determinate solution is then obtained by integrating the 2n equations (80) and (81). This, in turn, requires that 2n boundary conditions be specified. However, there are only n boundary conditions immediately apparent, namely,

$$x_j (t_i) = x_{ji} \qquad i = 1, 2, \cdots, n \tag{82}$$

The other boundary conditions depend on the constraints imposed on the final value of the state variables and the form of the criterion function, (65). We distinguish several cases.

I. Final Time Fixed; No Constraints on Final State Variables

In this case, the remaining n boundary values are

$$\lambda_j (t_f) = - c_j \qquad j = 1, 2, \cdots, n \tag{83}$$

where the c_j are the components of the vector c in Eq. (65)

II. Final Time Fixed; Constraints Imposed on Final State Variables

One type of constraint imposed on the final state vector is that it lie within a prescribed region of the state space; viz.,

$$\psi \left[x (t_f) \right] \lessgtr 0 \tag{84}$$

In this case, if the function ψ is differentiable, we write

$$b_j = \frac{\partial \psi}{\partial x_j}\bigg|_{t = t_f} \tag{85}$$

and the boundary values for λ are given by

$$\lambda_j(t_f) = - c_j - \eta\, b_j \left[x(t_f)\right] \qquad j = 1, 2, \cdots, n \tag{86}$$

$$\psi\left[x(t_f)\right] = 0 \tag{87}$$

where η is a constant. There are (n+1) equations in the set (86), (87) for the (n+1) unknowns $\lambda_j(t_f)$, $j = 1, 2, \cdots, n$, and the constant η. Thus one of the above equations may be used to eliminate η, which in combination with the initial values (82) yields the required 2n boundary conditions for the set (80), (81).

In the special case where q (< n) components of $x_j(t_f)$ are prescribed exactly,

$$x_j(t_f) = x_{jf} \qquad j = 1, 2, \cdots, q\ (< n) \tag{88}$$

the remaining boundary values are given by

$$\lambda_k(t_f) = - c_k \qquad k = (q+1), \cdots, n \tag{89}$$

III. Final Time Open

Since there is an additional degree of freedom in the system, an additional equation is required to make the solution determinate. If t_f does not appear in the criterion function (65), then

$$H(t_f) = \sum_{j=1}^{n} \lambda_j f_j \bigg|_{t=t_f} = 0 \tag{90}$$

is the additional relation required. If t_f does appear in J, then the variable x_{n+1} is introduced such that

$$\dot{x}_{n+1} = 1$$

$$x_{n+1}(t_i) = 0$$

The final value of x_{n+1} replaces t_f in the criterion function, and the added relation takes the form

$$H(t_f) = \sum_{j=1}^{n+1} \lambda_j f_j \Big|_{t=t_f} = 0 \tag{91}$$

The statement of the maximum principle, together with the boundary conditions as specified above, contains all the information necessary to solve the optimal control problem. There is, however, one possible ambiguity and this is discussed next.

Singular Solutions

For a certain class of problems, the control, u, is a scalar which enters the hamiltonian in a linear manner; viz.†

$$H = I(x, \lambda) + u\ K(x, \lambda) \tag{92}$$

where I and K are scalar functions that are independent of u. Here it has been assumed that the explicit dependence of the system equations or criterion function on time, t, has been removed by defining an additional state variable x_{n+1} such that

$$\dot{x}_{n+1} = 1 \quad , \qquad x_{n+1}(0) = 0$$

It is possible that the switching function, K(x, λ), is identically zero over some finite time interval. In this case, the hamiltonian H ceases to be an explicit function of the control variable u, and the maximum principle is unable to provide information about the desired optimal control. The usual procedure of selecting u so as to maximize H breaks down. In some instances it is possible to show that the condition K(x,λ) = 0 violates the constraints of the system (see the latter part of Example 2). Thus no further analysis is necessary; i.e., the solution is not singular.

†Actually, the discussion that follows is applicable to any function of u as long as the form of Eq. (92) remains.

In what follows, we will investigate the nature of singular solutions based on the work of Johnson and Gibson[176]. It may be shown that[†]

$$H^* = I + u^* K = \max_{u} (I + u K) = 0 \tag{93}$$

$$0 \leq t \leq t_f$$

Therefore the condition

$$I (\lambda, x) = K (\lambda, x) = 0 \tag{94}$$

corresponds to the case of singular control. The problem is simplified materially if the above condition reduces to a relation which is independent of λ. To do this, we may use the equations

$$\begin{aligned} I &= \dot{I} = \ddot{I} = \cdots\cdot = 0 \\ K &= \dot{K} = \ddot{K} = \cdots\cdot = 0 \end{aligned} \tag{95}$$

which follow directly from (94). If we now substitute[‡]

$$\dot{x}_j = \frac{\partial H \left(x, \lambda, u_s^*\right)}{\partial \lambda_j} \tag{96}$$

$$\dot{\lambda}_j = - \frac{\partial H \left(x, \lambda, u_s^*\right)}{\partial x_j} \tag{97}$$

into (95), then it may happen that the optimal control in the singular region reduces to a function involving the state variables only; viz.,

$$u_s^* = u_s^* (x)$$

A result of this type is obtained in Example 2. This is not always possible, especially for high-order systems. However, when a reduction of this type can be achieved, the optimal control in the singular region is available as a function of the current state of the system.

[†] Here, and in the remainder of the monograph, the asterisk superscript will be used to denote the optimal value.

[‡] The subscript s on u^* is used to denote the optimal control in the singular region.

For a more detailed discussion of the problem, the reader is referred to Johnson and Gibson.(176)

The application of the maximum principle to problems of optimal control may be summarized as follows.

a. Form the hamiltonian, Eq. (79).

b. Derive the optimal control, u*(t), by choosing that u(t) which (subject to the control constraints) maximizes the hamiltonian.

c. Investigate possible singular solutions.

d. Determine the optimal trajectory by integrating the 2n equations (80) and (81) using the optimal control u*(t) and the appropriate boundary conditions.

It is instructive to compare the necessary conditions for an optimum, Eq. (78), with the Euler-Lagrange equations (14). If we write

$$\varphi_j \equiv \dot{x}_j - f_j = 0 \tag{98}$$

then

$$F = \sum_j \lambda_j \left(\dot{x}_j - f_j\right) \tag{99}$$

whereupon Eqs. (14) become

$$\dot{\lambda}_k = -\sum_j \lambda_j \frac{\partial f_j}{\partial x_k} \tag{100}$$

Comparing with (78), we see that the choice of notation is indeed justified.

It should be emphasized that the maximum principle provides only a set of <u>necessar</u> conditions, much the same way as the Euler Lagrange equations and the Legendre Clebsc condition provide only necessary conditions in the variational calculus. Sufficient conditions are hard to come by (except in the case of linear systems). Nevertheless, this is usually a mathematical luxury. In practical situations, physical considerations will generally confirm uniqueness and sufficiency.

We should note also that the maximum principle yields a solution in the form of a set of (in general) nonlinear equations with two-point boundary conditions. Thus there exist the same computational difficulties as with the variational calculus.

In the final analysis, it must be granted that the maximum principle solves no problem that cannot also be solved by variational methods. Its appeal is mainly in the elegance of its format and the ease with which it can be applied. This, however, is a powerful argument in its favor for purposes of practical application.

The results obtained thus far yield an optimal control that is a function only of time and the initial conditions. In other words, the control is open loop. For most control systems it is desirable to have a control which is a function of the current state of the system; i.e., closed loop control. The latter is generally hard to come by except in special cases. For linear systems with a quadratic performance function, it is possible to obtain an optimal closed-loop control by deriving the Hamilton-Jacobi equation of the system. This is done in Sec. 3.2.1.

Example 2, The Sounding Rocket: The problem formulated in Example 1 will here be solved via the maximum principle. We have

$$\dot{x}_1 = x_2 \equiv f_1 \tag{101}$$

$$\dot{x}_2 = -g + \frac{(\beta\mu - D)}{x_3} \equiv f_2 \tag{102}$$

$$\dot{x}_3 = -\beta \equiv f_3 \tag{103}$$

The boundary conditions are

$$\begin{aligned} x_1(t_i) &= 0 & t_i &= 0 \\ x_2(t_i) &= 0 & x_2(t_f) &= 0 \\ x_3(t_i) &= 0 & x_3(t_f) &= m_p \end{aligned} \tag{104}$$

where $x_1 = h$, $x_2 = V$, $x_3 = m$

We seek to maximize the final altitude. Consequently, the criterion function to be minimized is

$$J = -h_f \equiv -x_{1f} \tag{105}$$

Eq. (89) then supplies the additional boundary condition

$$\lambda_1 (t_f) = 1 \tag{106}$$

We might just as easily have sought to maximize h_f, in which case we would have $\lambda_1 (t_f) = -1$. The form above has been adopted in order to permit ready comparison with the results of Example 1.

From Eqs. (78) or (81), we find

$$\dot{\lambda}_1 = \frac{1}{x_3} \frac{\partial D}{\partial x_1} \lambda_2 \tag{107}$$

$$\dot{\lambda}_2 = -\lambda_1 + \frac{1}{x_3} \frac{\partial D}{\partial x_2} \lambda_2 \tag{108}$$

$$\dot{\lambda}_3 = \frac{(\beta \mu - D)}{x_3^2} \lambda_2 \tag{109}$$

The hamiltonian is

$$\begin{aligned} H &= \lambda_1 x_2 + \lambda_2 \left[\frac{(\beta\mu - D)}{x_3} - g \right] - \lambda_3 \beta \\ &= \left[\lambda_1 x_2 - \lambda_2 \left(\frac{D}{x_3} + g \right) \right] + \beta \left[\frac{\mu \lambda_2}{x_3} - \lambda_3 \right] \end{aligned} \tag{110}$$

This function is <u>maximized</u> if

$$\begin{aligned} \beta &= \beta_M \qquad \text{when } K > 0 \\ \beta &= 0 \qquad \text{when } K < 0 \end{aligned} \tag{111}$$

where

$$K = \frac{\lambda_2 \mu}{x_3} - \lambda_3 \tag{112}$$

When K = 0, there is apparently no way to determine the optimal control since H is then independent of β. This difficulty may be resolved by employing the methods discussed in the section on "Singular Solutions." Using (95), the conditions corresponding to $K = \dot{K} = I = 0$ lead to the three equations,

$$\frac{\lambda_2 \mu}{x_3} - \lambda_3 = 0 \tag{113}$$

$$-x_3 \lambda_1 + \left(\frac{\partial D}{\partial x_2} + \frac{D}{\mu}\right)\lambda_2 = 0 \tag{114}$$

$$\lambda_1 x_2 - \lambda_2 \left(\frac{D}{x_3} + g\right) = 0 \tag{115}$$

Combining the last two of these leads to

$$D\left(x_2 - \mu\right) + \mu\left(x_2 \frac{\partial D}{\partial x_2} - x_3 g\right) = 0 \tag{116}$$

Since $\dot{x}_3 = -\beta$, this relation yields the value for the optimal control in the singular region. In other words, a variable thrust subarc (if it exists) must satisfy (116). This result is identical to that obtained in Example 1 by the variational calculus; i.e., Eq. (54).

The relation (90) for determining t_f is superfluous in the present case, since physical reasoning indicates that the final subarc is a coasting phase ($\beta = 0$), and the terminal condition is reached when $x_2(t) = 0$. However, Eq. (90) is satisfied for this value of t_f and provides a check on the solution.

We should now like to investigate the possibility of a variable thrust subarc in the absence of aerodynamics; i.e., D = 0. In this case, the condition (116) leads to $mg = 0$, an absurdity. Therefore, the optimal control in a vacuum is of the bang-bang type.

Example 3, Minimum Time Control of a Nonlinear Process: Consider the second-order nonlinear system described by

$$\dot{x}_1 = x_2 \equiv f_1 \tag{117}$$

$$\dot{x}_2 = -u_2\left(u_2 x_2 + \alpha u_1\right) \equiv f_2 \tag{118}$$

Here α is a positive constant, and the control functions, u_1 and u_2, satisfy the constraints

$$|u_1| \lessgtr 1 \tag{119}$$

$$0 < \gamma \lessgtr u_2 \lessgtr 1 \tag{120}$$

$$\gamma \equiv \text{prescribed constant}$$

It is required to determine the form of u_1 (t) and u_2 (t) such that the system is brought to the equilibrium position, $x_1 = x_2 = 0$, in minimum time from an arbitrary initial position

$$x_1 (t_i) = a$$

$$x_2 (t_i) = b$$

In other words, the function to be minimized is of the form (75). Consequently, we define

$$x_3 = \int_{t_i}^{t} dt$$

and

$$\dot{x}_3 = 1 \equiv f_3$$

$$x_3 (t_i) = 0$$

The problem now reduces to that of minimizing the quantity

$$J = x_3 (t_f)$$

The components of the costate vector, λ, are given by Eq. (78); viz.,

$$\dot{\lambda}_1 = 0 \tag{121}$$

$$\dot{\lambda}_2 = -\lambda_1 + u_2^2 \lambda_2 \tag{122}$$

$$\dot{\lambda}_3 = 0 \tag{123}$$

The hamiltonian is therefore expressed as

$$H = \lambda_1 x_2 - \lambda_2 u_2 \left(u_2 x_2 + \alpha u_1\right) + \lambda_3 \tag{124}$$

However, by Eq. (89),

$$\lambda_3 (t_f) = -1$$

This relation, in conjunction with (123) shows that

$$\lambda_3 (t) = -1$$

which means that (124) reduces to

$$H = \lambda_1 x_2 - \lambda_2 u_2 \left(u_2 x_2 + \alpha u_1\right) - 1 \tag{125}$$

It is immediately evident that H is maximized by[†]

$$u_1 = - \operatorname{sgn} \lambda_2 \tag{126}$$

If we write Eq. (125) in the form

$$H = -1 + \lambda_1 x_2 + \lambda_2 \left[\left(\frac{\alpha u_1}{2 x_2}\right)^2 - x_2 \left(u_2 + \frac{\alpha u_1}{2 x_2}\right)^2\right] \tag{127}$$

then we find that H is maximized if u_2 satisfies the following:

1. For $\lambda_2 < 0$ $(u_1 = 1)$

   a. If $x_2 > 0$
   then $u_2 = 1$

   b. If $x_2 < 0$, then
   $u_2 = 1$ if $|\rho| \geq 1$
   $= \rho$ if $\gamma \leq |\rho| \leq 1$
   $= \gamma$ if $|\rho| \leq \gamma$

[†] sgn y = 1 if y > 0
= -1 if y < 0

2. For $\lambda_2 > 0$ $(u_1 = -1)$

a. If $x_2 < 0$

then $u_2 = 1$

b. If $x_2 > 0$, then

$$\begin{aligned} u_2 &= 1 && \text{if } \rho \geq 1 \\ &= \rho && \text{if } \gamma \leq \rho \leq 1 \\ &= \gamma && \text{if } \rho \leq \gamma \end{aligned} \tag{128}$$

where

$$\rho = \frac{\alpha}{2\, x_2}$$

Conditions (126) and (128), together with the differential equations (117), (118), (121), and (122) and boundary values

$$\begin{aligned} x_1(t_i) &= a & \qquad x_1(t_f) &= 0 \\ x_2(t_i) &= b & \qquad x_2(t_f) &= 0 \end{aligned}$$

with the added constraint

$$\left[\lambda_1 x_2 - \lambda_2 u_2 \left(u_2 x_2 + \alpha u_1\right)\right]_{t=t_f} = 1$$

obtained from (91), are sufficient to determine the optimal u_1 and u_2.

For systems of moderately high order, the corresponding computational tasks are not at all trivial, even with computer assistance. In the present case, however, the phase-plane representation may be utilized to advantage in order to exhibit the salient features of the solution.

We observe parenthetically that the fundamental difficulty in starting a solution is that the sign of λ_2 is not known initially. However, the form of λ_1 and λ_2 is known, since Eqs. (121) and (122) are readily integrated; viz.,

$$\lambda_1 = \beta_1$$

$$\lambda_2 = e^{A}\left(\beta_2 - \int \beta_1 e^{-A}\, dt\right) \tag{129}$$

where

$$A = \int u_1^{\,2}\, dt$$

and β_1 and β_2 are constants of integration. The significant information available from an inspection of Eq. (129) is that λ_2 cannot change sign more than once.

Now for purposes of phase-plane representation, we express Eqs. (117) and (118) as

$$d\,x_1 = -\frac{x_2\, d\, x_2}{u_2\left(u_2\, x_2 + \alpha\, u_1\right)} \tag{130}$$

From condition (126), we know that $u_1^* = \mp 1$, while (128) shows that

$$u_2^* = 1 \text{ whenever } |x_2| \leq \frac{\alpha}{2}$$

regardless of the sign of λ_2. Consequently, in the phase plane, only two trajectories can enter the origin. One trajectory, L_1 (see Fig. 4), is obtained from the solution of Eq. (130) with $u_1^* = u_2^* = 1$. The other trajectory, L_1', is given by the solution of (130) with $u_1^* = -1$ and $u_2^* = 1$. We solve Eq. (130) by working backwards; that is, we take $x_1(0) = x_2(0) = 0$.

The resulting equations are

<u>For L_1</u>

$$\lambda_2 < 0,\ x_2 > 0$$

$$x_1 = -x_2 - \alpha \ln \left|\frac{\alpha}{x_2 + \alpha}\right|$$

<u>For L_1'</u>

$$\lambda_2 > 0,\ x_2 < 0$$

$$x_1 = -x_2 + \alpha \ln \left|\frac{\alpha}{\alpha - x_2}\right|$$

We now observe, for instance, that a trajectory cannot leave L_1 unless λ_2 changes sign. Assume this happens at point ①. Then u_1 switches sign and becomes -1, while u_2 takes the value γ, the trajectory continuing as shown by the path ① - ② - ③. At point ②, on line L_2, u_2 switches to

$$u_2^* = \frac{\alpha}{2\, x_2}$$

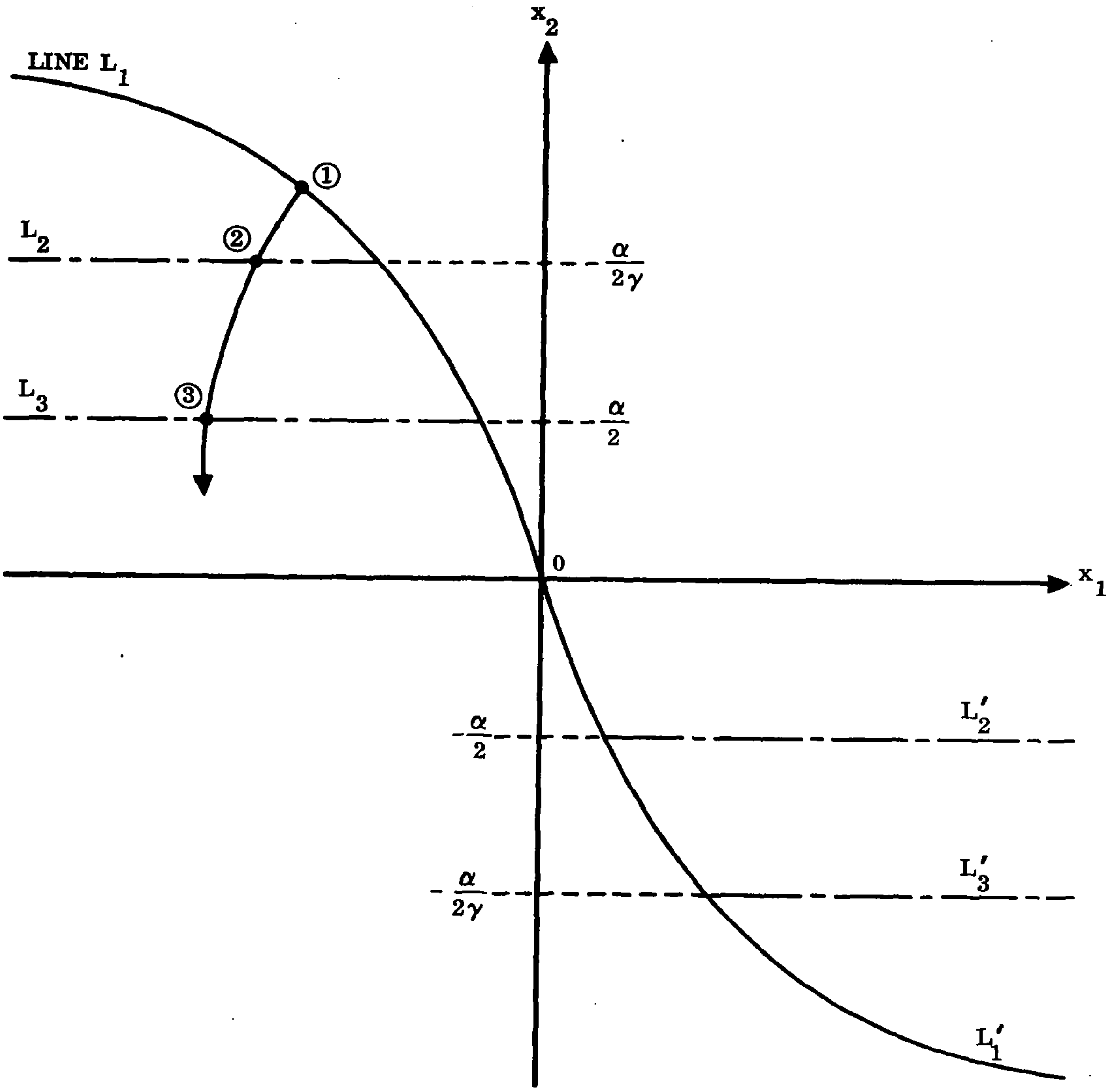


Figure 4. Optimum Trajectories in Phase Plane

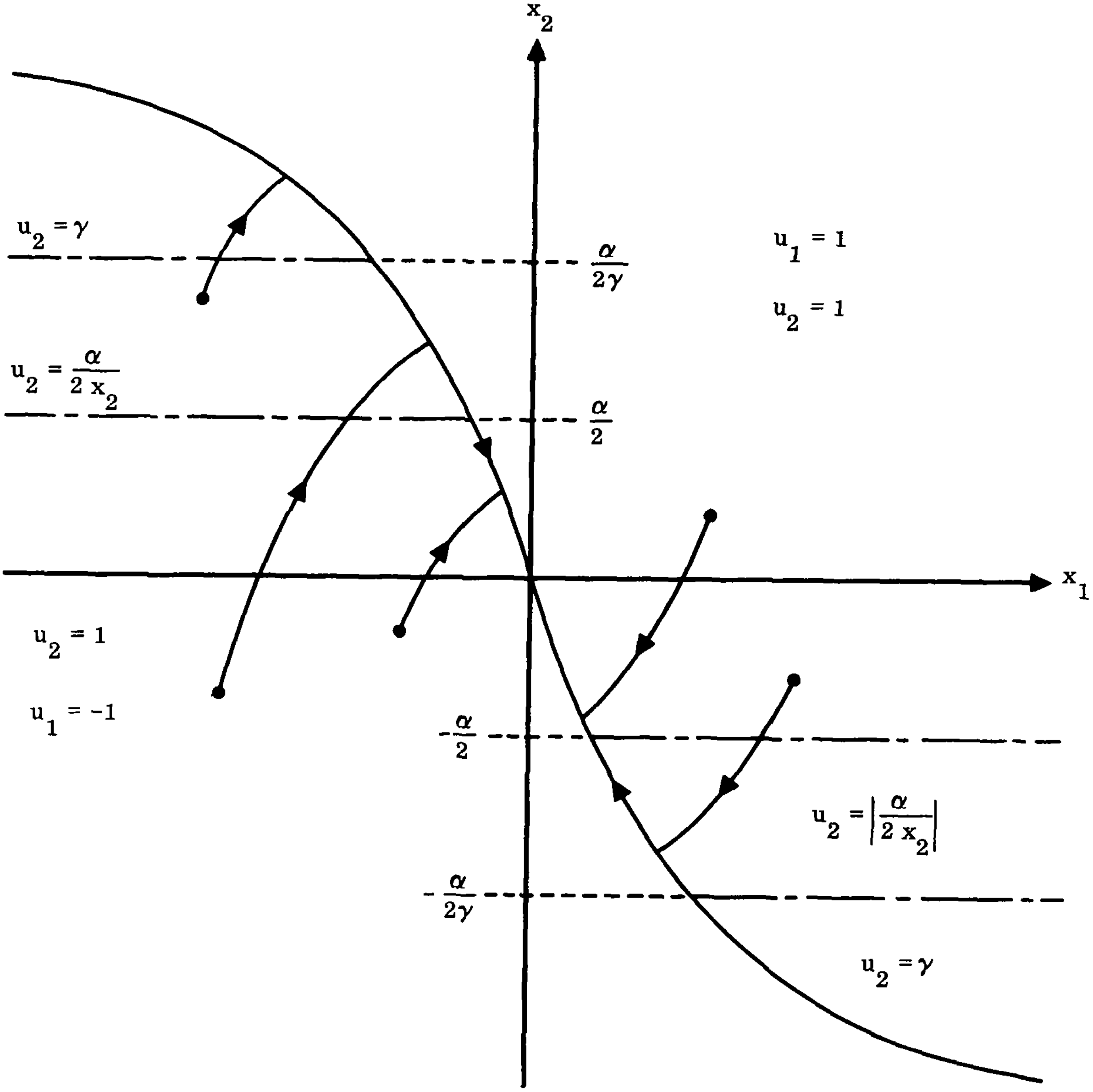


Figure 5. Optimal Control Regions in Phase Plane

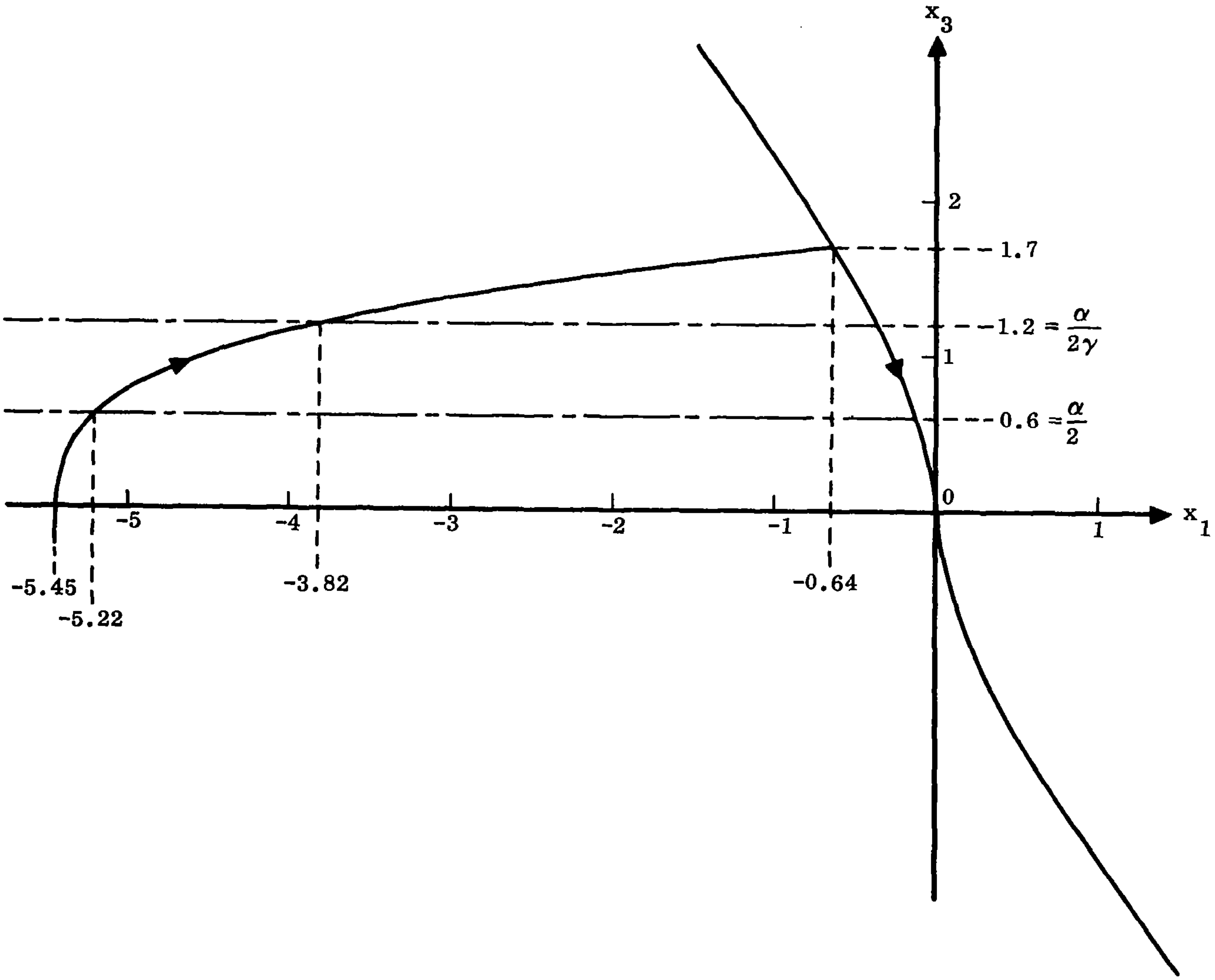


Figure 6. Optimal Phase Plane Trajectory for System of Example 3

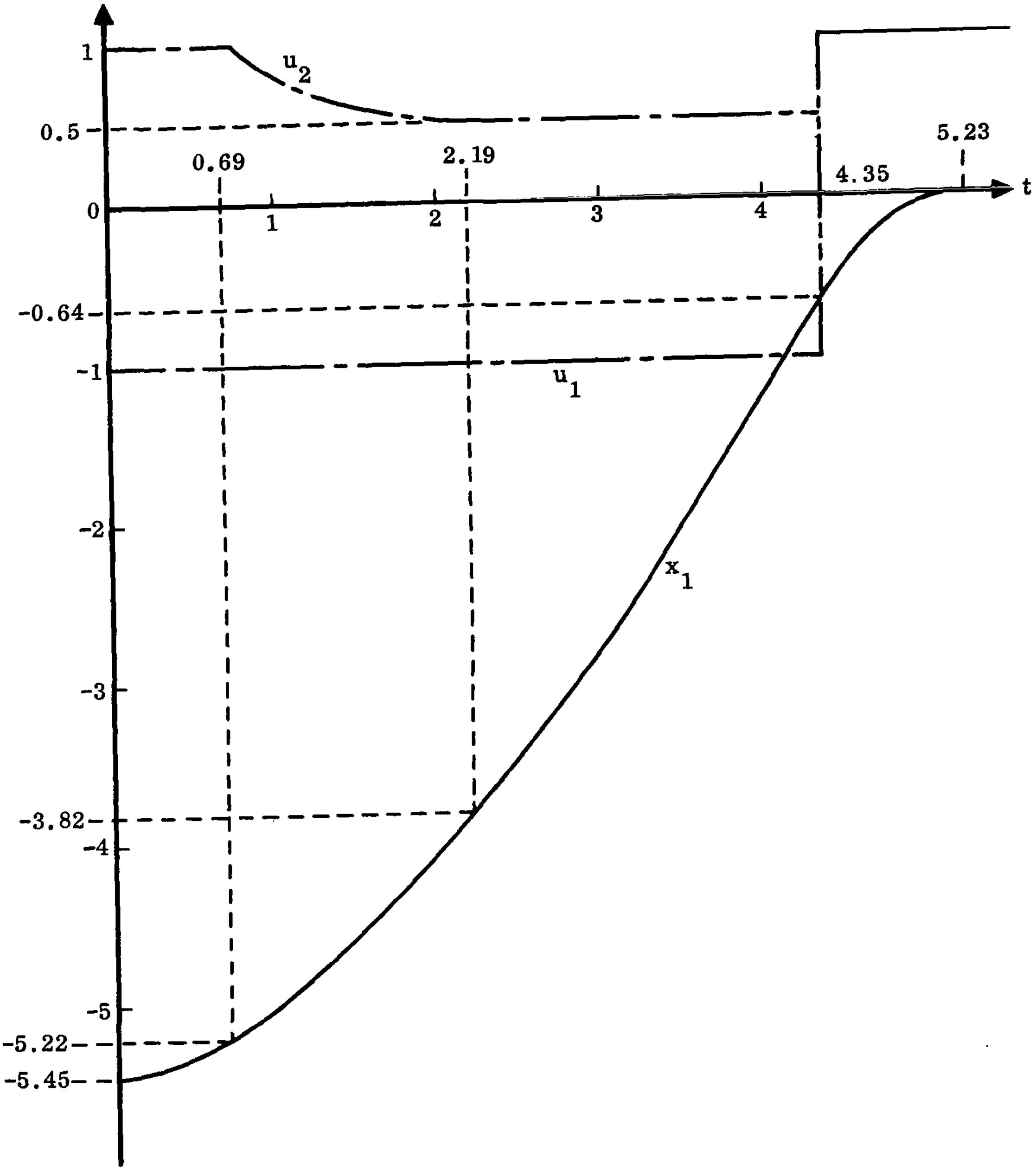


Figure 7. Transient Response for System of Example 3

satisfying this relation until the trajectory reaches point ③, after which $u_2^* = 1$.

Continuing the analysis along similar lines, we see that the phase plane is divided into six semi-infinite regions as shown in Fig. 5. In the half plane to the left of $L_1 - L_1'$, the control function, $u_1^* = -1$, while in the other half plane, $u_1^* = 1$. Since, as shown earlier, λ_2 can change signs only once, a trajectory, once it hits $L_1 - L_1'$, will remain on it until reaching the origin. Several typical trajectories from aribtrary initial points are sketched in Fig. 5.

A complete solution for the case where

$$x_1 (t_i) = -5.45, \quad \alpha = 1.2$$

$$x_2 (t_i) = 0 \quad , \quad \gamma = 0.5$$

is shown in Figs. 6 and 7.

3.1.3 Gradient Method

The earliest studies in optimal flight control systems employed the classical variational calculus, since this was the only tool available for the problem at hand. While many important and significant results were obtained, it became apparent that computational difficulties associated with the solution of nonlinear two-point boundary value problems (generated by the variational approach) severely limited the usefulness of the method. Since many problems of interest were of sufficient complexity to make the computational difficulties overwhelming, it was natural to seek new ways of solving them. In this respect, one of the more important concepts that emerged was the so-called "gradient method." The main virtue of this approach is that the numerical solution reduces to that of solving an initial value problem rather than a two-point boundary type. This enhances the computational aspects no end.

The main ideas of the method evolved in stages rather than appearing overnight in fully developed form. It appears that the fundamental concept is due to Courant.[151] A systematic development specifically oriented for aerospace applications is contained in the works of Kelley[75,55] and Bryson.[76,106] Various refinements and related techniques were developed by Ho[135], Knapp[139], Dreyfus[152], and Jazwinski.[153]

As in Sections 3.1.1 and 3.1.2, we will present only the main results reflecting the current state of the art. Proceeding from first principles in each area requires that a fairly elaborate groundwork be laid, which entails a task beyond the scope of this monograph. It will be shown later (Sec. 3.1.5) that the aforementioned results are quite readily derived from a more general point of view within the framework of the theory of dynamic programming.[16] The chief merit of this approach lies in its conceptual plausibility and intuitive appeal. The Euler Lagrange equations and the maximum principle are found to be necessary consequences of the principle of optimality.(16) However, in order to discuss the gradient method in this context, it is convenient to provide a motivational framework prior to presenting the main results.

The basic idea of the gradient method may be made pl_usible by considering an analogous problem in ordinary calculus.

Steepest Descent in Ordinary Calculus

Consider a scalar function of n variables

$$f = f(x) \tag{131}$$

where, as before, x is an n vector. It is required to determine the vector $x^{(0)}$ such that the function f is stationary; i.e., is either a maximum or a minimum.

A first-order change in f is given by

$$df = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} dx_j = (\nabla_x f)^T dx \tag{132}$$

where

$$\nabla_x f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} \tag{133}$$

which is the usual definition for the gradient of a function. It is known that the gradient is the directional derivative in the direction of greatest change. Therefore, for a prescribed increment, dx, the greatest change, df, is produced if dx is taken in the gradient direction, viz.,

$$dx = K \nabla_x f \tag{134}$$

where K is a constant that is chosen negative if a decrease in the value of f is desired and positive if an increase is desired.

With this choice of dx, the change in f becomes†

$$df = K (\nabla_x f)^2 \tag{135}$$

Note that since Eq. (132) is a linear approximation, dx must be "sufficiently small" to ensure that the linearity restrictions are not violated.

Suppose now that $x^{(1)}$ is a first guess for the minimizing vector. The function, (131), takes the value f $(x^{(1)})$. We now seek to apply a fixed increment, dx, to $x^{(1)}$ such that the corresponding decrease in f $(x^{(1)})$ is a maximum. As noted above, this is achieved by taking dx in the negative gradient direction; i.e., taking as an improved estimate

$$x^{(2)} = x^{(1)} - K \nabla_x f \tag{136}$$

K ≡ positive constant

The value of K must be chosen such that

$$f\left(x^{(2)}\right) < f\left(x^{(1)}\right) \tag{137}$$

If this relation is not satisfied, it implies that linearization was violated, and a smaller value of K must be used. The process is continued until

$$\left| f\left(x^{(r+1)}\right) - f\left(x^{(r)}\right) \right| \lessdot \epsilon \tag{137}$$

ϵ ≡ small preassigned positive constant

The vector $x^{(r)}$ is then the one that minimizes f (x) to the desired degree of accuracy. If $x^{(0)}$ is the exact minimizing vector, then

$$\nabla_x f \; x^{(0)} = 0 \tag{138}$$

It should be emphasized that this approach (like the variational one) ensures only a local extremum. In practice, global optimality is generally verified by physical considerations.

†For any vector, a, $a^2 \equiv a^T a$.

The nature of the gradient methdd suggests the picturesque analogy of descending a hill in some efficient manner. At any given point, one determines the steepest slope and then proceeds a short distance in that direction.

This procedure, in the context of ordinary calculus, deals with incrementing variables in order to optimize a function. In the variational calculus, one seeks to optimize the function of a function, and here we must vary a function rather than a variable. As a result, we are confronted with all sorts of disagreeable complications. Nevertheless, the basic principle still applies and yields a powerful tool for optimization problems in aerospace guidance and controls. It will be shown that the gradient format for variational problems is conceptually identical with that for the elementary calculus, except that Green's functions replace the partial derivatives and the vanishing of the gradient is intimately related to the conditions expressed by the Euler Lagrange equations. These ideas will be developed next.

Steepest Descent in Function Space

The essential features of the gradient method, as applied to variational problems, may be exhibited by investigating the following. Given the integral

$$I = \int_{t_i}^{t_f} L\left[\dot{x}(t), x(t), t\right] dt \tag{139}$$

it is required to choose the vector, x (t), such that I is minimized (or maximized). Note that there is an essential distinction between this and the previous problem. Whereas x was an n-dimensional variable before, x (t) is an n-dimensional function now.

Suppose now that t_i and t_f are fixed. By taking the variation[†] of I, we find

$$\delta I = \int_{t_i}^{t_f} \delta L(\dot{x}, x, t)\, dt$$

$$= \int_{t_i}^{t_f} \left\{ \left(\frac{\partial L}{\partial x}\right)^T \delta x + \left(\frac{\partial L}{\partial \dot{x}}\right)^T \delta \dot{x} \right\} dt$$

$$= \int_{t_i}^{t_f} \left\{ (\nabla_x L)^T \delta x + (\nabla_{\dot{x}} L)^T \delta \dot{x} \right\} dt \tag{140}$$

[†] See Appendix B.

Since the variation and differential operators are commutative,†

$$\frac{d}{dt}(\delta x) = \delta\left(\frac{dx}{dt}\right) \tag{141}$$

Eq. (140) becomes

$$\delta I = \int_{t_i}^{t_f} \left\{ (\nabla_x L)^T \delta x + \frac{d}{dt}\left[(\nabla_{\dot{x}} L)^T \delta x \right] - (\delta x)^T \frac{d}{dt}(\nabla_{\dot{x}} L) \right\} dt \tag{142}$$

However,

$$\int_{t_i}^{t_f} \frac{d}{dt}\left[(\nabla_{\dot{x}} L)^T \delta x \right] dt = (\nabla_{\dot{x}} L)^T \delta x \Big]_{t_i}^{t_f} = 0 \tag{143}$$

since $\delta x(t_f) = \delta x(t_i) = 0$ because t_i and t_f are both specified. Therefore, Eq. (142) reduces to

$$\delta I = \int_{t_i}^{t_f} (\delta x)^T \left\{ \nabla_x L - \frac{d}{dt}(\nabla_{\dot{x}} L) \right\} dt \tag{144}$$

Defining a generalized gradient by

$$\Big[\ \Big]_x \equiv \frac{d}{dt} \nabla_{\dot{x}} - \nabla_x \tag{145}$$

we may write

$$\delta I = \int_{t_i}^{t_f} (\delta x)^T \Big[L \Big]_x \, dt \quad . \tag{146}$$

†See Appendix B.

The variation of I is thus expressed as an inner product in function space[†], compared with the differential of f, Eq. (132), which is expressed as a conventional inner product of finite vectors. In addition, we note the correspondence

$$df \Longleftrightarrow \delta I$$

$$dx \Longleftrightarrow dx\,(t)$$

$$\nabla_x f \Longleftrightarrow \left[\frac{d}{dt}\nabla_{\dot{x}} - \nabla_x\right] L$$

The operator $[\ ]_x$ defined by Eq. (145) may thus be identified as the gradient of I in function space. Consequently, in analogy with (134), we choose

$$\delta x = K\left[L\right]_x \tag{147}$$

as the variation in x (t) that produces the greatest change in δI. Here, K is a constant that is negative if it is desired to decrease I, and positive otherwise. The gradient is evaluated along the nominal path, $x^{(1)}(t)$. The variation in I is therefore expressed as

$$\delta I = K\int_{t_i}^{t_f}\left\{\left[L\right]_x\right\}^2 dt \tag{148}$$

in complete analogy with Eq. (135).

We recall that a necessary condition for f to be an extremum is that the relation

$$\nabla_x f\,(x) = 0 \tag{149}$$

be satisfied; i.e., that the gradient vanish at the local extremum.

Invoking a fundamental theorem of the calculus of variations,[98] — namely that δ I must vanish if I is stationary — an inspection of Eq. (146) leads to the condition

$$\left[L\right]_x = 0 \tag{150}$$

since δ x is arbitrary between the end points, t_i and t_f. In other words, a necessary condition for the extremum in function space is that the generalized gradient vanish.

[†] See Appendix C.

In order to relate this condition to the necessary conditions derived via the conventional variational approach, we will solve the problem using the methods of Sec. 3.1.1. To this end we define

$$x_{n+1} = \int_{t_i}^{t} L(\dot{x}, x, t)\, dt \tag{151}$$

and pose the equivalent problem of minimizing

$$J = x_{n+1}(t_f) \tag{152}$$

subject to the constraint

$$\varphi \equiv \dot{x}_{n+1} - L(\dot{x}, x, t) = 0 \tag{153}$$

and initial condition

$$x_{n+1}(t_i) = 0 \tag{154}$$

Forming the augmented function†

$$F = \lambda\left(\dot{x}_{n+1} - L\right)$$

we find

$$\frac{\partial F}{\partial \dot{x}_j} = -\lambda \frac{\partial L}{\partial \dot{x}_j}$$

$$\frac{\partial F}{\partial x_j} = -\lambda \frac{\partial L}{\partial x_j} \qquad j = 1, 2, \cdots, n$$

while

$$\frac{\partial F}{\partial \dot{x}_{n+1}} = \lambda$$

$$\frac{\partial F}{\partial x_{n+1}} = 0$$

†See Eq. (13) et seq. Note that λ as used here is a scalar.

The Euler Lagrange equations are then

$$\frac{d}{dt}\left(\lambda \frac{\partial L}{\partial \dot{x}_j}\right) - \lambda \frac{\partial L}{\partial x_j} = 0 \qquad j = 1, 2, \cdots, n$$

$$\dot{\lambda} = 0$$

The last relation shows that λ is a constant, which means that the preceding equations become

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}_j}\right) - \frac{\partial L}{\partial x_j} = 0 \qquad j = 1, 2, \cdots, n \tag{155}$$

These are the Euler Lagrange equations for the problem at hand. Written as

$$\left(\frac{d}{dt} \nabla_{\dot{x}} - \nabla_x\right) L \equiv \left[L\right]_x = 0 \tag{156}$$

we find by comparison with (150) that the vanishing of the gradient in function space is another way of expressing the Euler Lagrange equations.

For purposes of numerical computation, the evaluation of the gradient employing the operator defined by Eq. (145) involves the calculation of higher-order derivatives. This is never an advisable numerical procedure, and other computational devices must be sought. It turns out that the use of Green's functions leads to an efficient algorithm for the numerical computation of the gradient. We turn to a discussion of the basic ideas involved in this approach.

Consider the system described by

$$\dot{x}_j = f_j (x, u, t) \qquad (j = 1, 2, \cdots, n) \tag{157}$$

where suitable boundary values are prescribed, and u is an m-dimensional control vector. It is required to determine u(t) such that the function

$$J = J\left[x (t_f)\right] \tag{158}$$

is an extremal. As noted in the previous sections, a wide variety of performance functions can be expressed in this format. Suppose now that one can select a control function, u(t), that satisfies the given boundary conditions for (157) but that does not necessarily optimize (158). We will refer to this control vector and the corresponding trajectory as the nominal u(t) and x(t) respectively. From this nominal trajectory, we may calculate a value for J.

Suppose now that we adopt a new control, u(t) + δ u(t), which represents a perturbed value of the nominal control. This will produce a perturbed trajectory, x(t) + δ x(t). We seek to determine the form of the perturbation that will produce the greatest improvement (increase or decrease, as the case may be) in the performance criterion, J. For δ u and δx sufficiently small, we have, to first order,

$$\dot{x}_j + \delta \dot{x}_j = f_j (x, u, t) + \delta f_j (x, u, t) \qquad (159)$$

$$(j = 1, 2, \cdots, n)$$

By way of the Taylor expansion,

$$\delta f_j (x, u, t) = \sum_{k=1}^{n} \frac{\partial f_j}{\partial x_k} \delta x_k + \sum_{\ell=1}^{m} \frac{\partial f_j}{\partial u_\ell} \delta u_\ell$$

and noting that†

$$\delta \dot{x}_j = \frac{d}{dt} (\delta x_j)$$

we have, from (159),

$$\frac{d}{dt} (\delta x_j) = \sum_{k=1}^{n} \frac{\partial f_j}{\partial x_k} \delta x_k + \sum_{\ell=1}^{m} \frac{\partial f_j}{\partial u_\ell} \delta u_\ell \qquad (160)$$

$$(j = 1, 2, \cdots\cdot, n)$$

This may be written in vector matrix form as

$$\frac{d}{dt} (\delta x) = A (t) \delta x + B (t) \delta u \qquad (161)$$

$$\delta x (t_i) = 0 \qquad (162)$$

†See Appendix B.

where

$$A(t) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1} & \frac{\partial f_n}{\partial x_2} & \cdots & \frac{\partial f_n}{\partial x_n} \end{bmatrix} \tag{163}$$

$$B(t) = \begin{bmatrix} \frac{\partial f_1}{\partial u_1} & \frac{\partial f_1}{\partial u_2} & \cdots & \frac{\partial f_1}{\partial u_m} \\ \frac{\partial f_2}{\partial u_1} & \frac{\partial f_2}{\partial u_2} & & \vdots \\ \vdots & \vdots & & \vdots \\ \frac{\partial f_n}{\partial u_1} & \frac{\partial f_n}{\partial u_2} & \cdots & \frac{\partial f_n}{\partial u_m} \end{bmatrix} \tag{164}$$

All partial derivatives are evaluated along the nominal trajectory.

Eq. (161) is a linear nonstationary system whose solution is given by(155)

$$\delta x(t_f) = \int_{t_i}^{t_f} \Phi(t_f, t)\, B(t)\, \delta u(t)\, dt \tag{165}$$

where $\Phi(t, t_k)$ is the state transition matrix for the system (161). It satisfies the relations

$$\delta x(t) = \Phi(t, t_k)\, b \; ; \quad \delta x(t_k) = b \tag{166}$$

$$\frac{d}{dt} \Phi(t, t_k) = A(t) \Phi(t, t_k) \tag{167}$$

$$\Phi(t_k, t_k) = I, \text{ the unit matrix} \tag{168}$$

Now

$$\delta J = J\left[x(t_f) + \delta x(t_f)\right] - J\left[x(t_f)\right]$$

which reduces to

$$\delta J = c^T \delta x(t_f) \tag{169}$$

where

$$c = \left(\frac{\partial J}{\partial x}\right)_{t=t_f} \equiv \left[\nabla_x J\right]_{t=t_f} \tag{170}$$

Substituting (165) in (169), we have

$$\delta J = \int_{t_i}^{t_f} (\delta u)^T B^T(t) \Phi^T(t_f, t) c \, dt \tag{171}$$

The evaluation of this expression is awkward, because the behavior of δJ depends on $\Phi(t_f, t)$ as a function of t, its <u>second</u> argument, while $\Phi(t_f, t)$ is normally related to current and past state as a function of its <u>first</u> argument; see Eq. (166). This predicament may be neatly resolved by employing the theory of adjoint equations. Specifically, we write†

$$\dot{\lambda} = -A(t) \lambda \tag{172}$$

†The notation anticipates the interpretation of λ as the vector Lagrange multiplier.

which is the vector matrix equation, adjoint to the system (161). If we take the initial conditions on this to be

$$\lambda\,(t_f) = c \tag{173}$$

where c is given by Eq. (170), then the solution of (172) is given by

$$\lambda\,(t) = \Psi\,(t,\, t_f)\, c \tag{174}$$

where $\Psi(t, t_k)$ is the state transition matrix for the system (172) and satisfies

$$\frac{d}{dt}\,\Psi\,(t,\, t_k) = -\,A(t)\,\Psi(t,\, t_k) \tag{175}$$

$$\Psi\,(t_k,\, t_k) = I, \text{ the unit matrix} \tag{176}$$

Notice that Eq. (174) is integrated "backwards" from the given final conditions. $\lambda(t)$ is therefore a known function (which, of course, depends implicitly on the reference trajectory).

Now it is known(155) from the theory of adjoint equations that

$$\Phi^T\,(t_f,\, t) = \Psi(t,\, t_f) \tag{177}$$

which means that (174) may be expressed as

$$\lambda(t) = \Phi^T\,(t_f,\, t)\, c \tag{178}$$

Substituting this in Eq. (171), we find

$$\delta\, J = \int_{t_i}^{t_f} (\delta u)^T\, B^T(t)\, \lambda\,(t)\, dt \tag{179}$$

If we define

$$H = f^T\, \lambda \tag{180}$$

then

$$\frac{\partial H}{\partial u} = \nabla_u\, H = B^T\, \lambda \tag{181}$$

Therefore, Eq. (179) becomes

$$\delta J = \int_{t_i}^{t_f} (\delta u)^T \left(\frac{\partial H}{\partial u}\right) dt \qquad . \tag{182}$$

This now has the form of an inner product in function space, and, in the light of the previous discussions, we interpret $(\partial H/\partial u)$ as the gradient in function space with respect to the control vector, u(t). Since the latter is arbitrary, the condition for the vanishing of δJ (which thereby ensures the J is an extremal) is that the gradient vanish; viz.,

$$\left(\frac{\partial H}{\partial u}\right) = 0 \tag{183}$$

It will be shown later (Sec. 3.1.5) that this condition is precisely the maximum principle discussed in the previous section. Thus H, as defined by Eq. (180), is indeed the hamiltonian of the system, and λ is the vector Lagrange multiplier. We note also that the quantities, $\partial H/\partial u_\ell$, are the Green's functions (sometimes called "influence functions") for the system; these indicate how small perturbations, δu, affect the performance function, J.

It now follows that a maximum change in δJ is obtained when

$$\delta u = K \nabla_u H \tag{184}$$

i.e., the change in δu is along its gradient.

K is taken negative if a decrease in J is sought, and positive otherwise. As a result,

$$\delta I = K \int_{t_i}^{t_f} (\nabla_u H)^2 dt \tag{185}$$

in the same manner as before.

In order to limit the variation δu, it is convenient to introduce the constraint

$$\int_{t_i}^{t_f} (\delta u)^2 dt = \epsilon^2 \tag{186}$$

where ϵ is a prescribed constant. Therefore, in order to extremize δJ subject to the constraint (186), we introduce an additional Lagrange multiplier, μ, and write

$$\delta J = \int_{t_i}^{t_f} (\delta u)^T \left(\frac{\partial H}{\partial u}\right) dt + \mu \left[\epsilon^2 - \int_{t_i}^{t_f} (\delta u)^2 \, dt\right]$$

$$= \int_{t_i}^{t_f} (\delta u)^T \left[\left(\frac{\partial H}{\partial u}\right) - \mu \, \delta u\right] dt + \mu \epsilon^2 \tag{187}$$

For an extremal δJ, the first variation must vanish; i.e.,

$$\delta (\delta J) = \int_{t_i}^{t_f} \left[\delta (\delta u)\right]^T \left[\left(\frac{\partial H}{\partial u}\right) - 2\mu \delta \, u\right] dt = 0$$

which leads to

$$\delta u = \frac{1}{2\mu} \left(\frac{\partial H}{\partial u}\right) \equiv \frac{1}{2\mu} \nabla_u H \tag{188}$$

where the Lagrange multiplier, μ , is evaluated from

$$\int_{t_i}^{t_f} \left[\frac{1}{2\mu} \left(\frac{\partial H}{\partial u}\right)\right]^2 dt = \epsilon^2 \tag{189}$$

The constant, μ , may be identified with constant K in Eq. (184).

In principle, the procedure is now straightforward. One selects a nominal u(t) that satisfies the boundary conditions for the system and that therefore yields a nominal trajectory, x(t), along with some value for the performance criterion, J. In order to decrease (or increase, as the case may be) this value of J, one uses a perturbed control function, u(t) + δu(t), where δ u(t) is selected in accordance with Eq. (184) or (188). One then determines a new trajectory and a new value for J that is improved (hopefully) over the preceding one. This procedure is continued until the improvements in J are less than some predetermined amount, or when $\delta J \rightarrow 0$.

In practice, a variety of complications preclude the formulation of a simple cookbook procedure that will always work. Thus far, these factors have been touched upon only lightly or not at all. Since a successful application of the gradient method depends upon a full understanding of the various subtleties involved and of how to resolve them in particular applications, we must consider these in a little greater detail.

The Problem of Constraints

Having provided a degree of motivation for the gradient technique, it remains to investigate the manner in which various realistic constraints are accounted for in the general approach. Three main types of constraint occur in physical applications.

1. Terminal

$$\psi_\alpha\left[x\,(t_f),\, t_f\right] = 0 \tag{190}$$

$$\alpha = 1,\, 2,\, \cdots,\, p$$

$$\psi_\beta\left[x\,(t_f),\, t_f\right] \leq 0 \tag{191}$$

$$\beta = p+1,\, p+2,\, \cdots,\, q$$

2. State Variable

$$g_k\,(x,\, u,\, t) \leq 0 \tag{192}$$

$$k = 1,\, 2,\, \cdots,\, r$$

3. Control

$$\nu_\ell \leq u_\ell \leq \mu_\ell \tag{193}$$

$$\ell = 1,\, 2,\, \cdots,\, m$$

The final time, t_f, is generally free (unspecified), and the initial conditions, $x_j(t_i)$, may be specified or left open to optimization.

The simplest type of constraint, as far as the gradient method is concerned, is the inequality constraint, (193), on the control variables. These are accounted for in the computational sequence as follows.[55] Assume that the nominal control vector $u^{nom}(t)$ satisfies (193) for all t.† The gradient technique then generates a new control

$$u^{new}(t) = u^{old}\,(t) + \delta u\,(t)$$

and it is observed whether at each instant of time, the relation

$$\nu_\ell \leq u_\ell^{old}\,(t_k) + \delta u_\ell\,(t_k) \leq \mu_\ell$$

$$(\ell = 1,\, 2,\, \cdots,\, m)$$

†Since the numerical procedure must of necessity deal with discrete instants of time, $u^{nom}(t)$ satisfies (193) at all discrete time instants involved in the computation.

is satisfied. If it is, we then use $u^{new}(t)$ to generate a new trajectory in the manner previously discussed. If not, we "trim" the control as follows.

$$u_\ell^{new}(t_k) = \mu_\ell \quad \text{if } u_\ell^{new}(t_k) > \mu_\ell$$

and

$$u_\ell^{new}(t_k) = \nu_\ell \quad \text{if } u_\ell^{new}(t_k) < \nu_\ell$$

To take account of the state variable constraint, (192), we define new state variables

$$x_{n+k}(t) = \int_{t_i}^{t} G_k(x, u, t)\,dt \qquad k = 1, 2, \cdots, r \tag{194}$$

with

$$\begin{aligned} G_k(x, u, t) &= 0 \quad \text{if } g_k \leqq 0 \\ &= g_k^2 \quad \text{if } g_k > 0 \end{aligned} \qquad k = 1, 2, \cdots, r \tag{195}$$

and add

$$\dot{x}_{n+k} = G_k(x, u, t) \qquad (k = 1, 2, \cdots, r) \tag{196}$$

to the other state equations, together with the initial conditions

$$x_{n+k}(t_i) = 0 \qquad k = 1, 2, \cdots, r \tag{197}$$

We note that the derivatives of G_k are everywhere continuous if the derivatives of g_k are continuous.(153)

As a result of this operation, the state variable inequality constraint, (192), is equivalent to

$$x_{n+k}(t_f) = 0 \qquad k = 1, 2, \cdots, r \tag{198}$$

which is in the form of the terminal equality constraint, (190).

Consequently, we will dismiss state variable inequality constraints from further consideration, assuming that they have been transformed to the equivalent constraint, (190).

The problems that now remain are the treatment of the terminal constraints, (190) and (191), the selection of appropriate step sizes and proportionality constants, and the finding of a nominal control vector that yields a trajectory satisfying prescribed terminal and initial values. These will be discussed within the context of the general solution, which is considered next.

The General Solution

Given the system

$$\dot{\underset{\cdot}{x}} = f(x, u, t) \tag{199}$$

$x \equiv$ n-dimensional state vector

$\underset{\cdot}{u} \equiv$ m-dimensional control vector

with the initial conditions

$$x_j(t_i) = a_j \qquad j = 1, 2, \cdots, s\ (\lesssim n) \tag{200}$$

where the (n−s) initial conditions, which are not specified, are left open to optimization. At the final time, t_f, the constraints

$$\psi_\alpha\left[x(t_f), t_f\right] = 0 \qquad \alpha = 1, 2, \cdots, p \tag{201}$$

$$\psi_\beta\left[x(t_f), t_f\right] \lesssim 0 \qquad \beta = p+1, p+2, \cdots, q \tag{202}$$

must be satisfied. In general, t_f is unspecified.

The components of the control vector must satisfy the inequality constraints

$$\nu_\ell \lessgtr u_\ell \lessgtr \mu_\ell \qquad \ell = 1, 2, \cdots, m \tag{203}$$

It is assumed that constraints on the state variable, as given by (192), have been transformed to type (190) by the introduction of additional state variables in the manner indicated previously.

We now seek to determine the control vector, u(t), that minimizes (maximizes) the performance index

$$J = J\left[x\,(t_f),\, t_f\right] \tag{204}$$

such that the constraints, (201) through (203), are satisfied, given the initial conditions, (200).

The problem thus formulated is of wide generality. Many realistic optimization problems in aerospace guidance and control can be expressed in this format.

In presenting the solution to the problem, we will limit the discussion to the main results. The complete derivations are quite lengthy, requiring many digressions that are beyond the scope of the present monograph. For a detailed treatment, the reader should consult the appropriate references.(3,42,55,75,76,103,135,153)

Suppose now that we arbitrarily select a nominal u(t) that satisfies constraint (203) and that thereby yields a nominal trajectory, x(t), satisfying the initial conditions (200) but not necessarily all the terminal constraints, (201) and (202).

The problem is now taken as one of minimizing the function†

$$P = K_0 J + \sum_{\gamma=1}^{q} K_\gamma \psi_\gamma^{\,2} \tag{205}$$

where

$K_0 = -1$ if J is to be maximized

$ = +1$ if J is to be minimized

†The values of J and ψ_γ are, of course, obtained from the nominal trajectory.

K_γ = 0 if the nominal trajectory satisfies the terminal constraint, ψ_γ, given by (201) or (202).

The basic idea here is that if the K_γ are large (compared with K_0), any control, u(t), that seriously violates terminal conditions (201) and (202) will give rise to large values of $K_\gamma \psi_\gamma^2$, with the result that the next iteration on u(t) will be biased in favor of satisfying the terminal constraints rather than optimizing J. Consequently, the more closely the terminal constraints are satisfied, the smaller the value of $K_\gamma \psi_\gamma^2$ (whatever the size of K_γ), with the result that subsequent iterations on u(t) become biased in favor of optimizing J. This notion, which is intuitively plausible, is also mathematically legitimate.[156, 157]

Now since the final time, t_f, is free, one of the $\psi_\gamma = 0$ may be singled out as a stopping condition — and is therefore not included in the summation term of Eq. (205). Call this stopping condition

$$\Omega\left[x\,(t_f),\ t_f\right] = 0 \tag{206}$$

At each pass, the integration of the trajectory equations is stopped when $\Omega = 0$.

Defining the hamiltonian as before,

$$H = f^T \lambda \tag{207}$$

where λ satisfies

$$\dot{\lambda} = -A(t)\,\lambda \tag{208}$$

with A(t) given by (163), we take, for the initial conditions on λ,

$$\lambda_j\,(t_f) = \left(\frac{\partial P}{\partial x_j}\right)_{t=t_f} - \frac{\dot{P}\,(t_f)}{\dot{\Omega}(t_f)} \left(\frac{\partial \Omega}{\partial x_j}\right)_{t=t_f} \tag{209}$$

Then, if the variation in the control vector is constrained by

$$\int_{t_i}^{t_f} \sum_{\ell=1}^{m} w_\ell\,(t) \left[\delta u_\ell\,(t)\right]^2 dt = \epsilon^2 \tag{210}$$

ϵ = prescribed constant

$w_j(t) > 0$, a weighting factor

we take, as a new control,

$$u^{new}(t) = u^{old}(t) + \delta u\,(t) \tag{211}$$

where $u^{old}(t)$ represents the original nominal control and

$$\delta u_{\ell}(t) = \frac{1}{2\,\mu\, w_{\ell}(t)} \frac{\partial H}{\partial u_{\ell}} \tag{212}$$

$$\ell = 1, 2, \cdots\cdot, m$$

where μ is an additional Lagrange multiplier calculated from

$$\int_{t_i}^{t_f} \sum_{\ell=1}^{m} \frac{1}{4\,\mu^2\, w_{\ell}(t)} \left(\frac{\partial H}{\partial u_{\ell}}\right)^2 dt = \epsilon^2 \tag{213}$$

The weighting functions, $w_j(t)$, are introduced to permit a greater flexibility in the numerical convergence procedure, especially in regions of high sensitivity. For initial trials, one may take $w_{\ell}(t) = 1$ for all ℓ.

Now by defining

$$\eta^2 = \int_{t_i}^{t_f} \sum_{\ell=1}^{m} \frac{1}{w_{\ell}(t)} \left(\frac{\partial H}{\partial u_{\ell}}\right)^2 dt \tag{214}$$

we may write Eq. (213) as

$$\frac{\eta^2}{4\,\mu^2} = \epsilon^2 \tag{215}$$

Using this to eliminate μ from (212), the latter becomes

$$\delta u_{\ell} = \frac{\epsilon}{\eta w_{\ell}} \frac{\partial H}{\partial u_{\ell}} \tag{216}$$

It can be shown that the change in P is then given by(158)

$$dP = \epsilon \eta \tag{217}$$

which means that ϵ must be chosen negative, since a minimum of P is sought.

The (n-s) initial conditions, which are unspecified, are now modified according to

$$d\,x_j\,(t_i) = \frac{d\,P}{\lambda_j(t_i)} \tag{218}$$

$$j = s+1,\ s+2,\ \cdots\cdot,\ n$$

It is noted that the $\lambda_j(t_i)$ are the Green's functions for the initial conditions.

The general computation procedure is now summarized as follows.

1. Select a nominal control vector, u(t), and integrate the equations of motion, (199), using the initial conditions, (200). If only s of the n initial conditions are given, choose arbitrary values for the (n-s) unspecified initial conditions. These will then be optimized in the course of the computational procedure. The integration is terminated when condition (206) is satisfied. This yields the nominal trajectory and the nominal terminal time, t_f. In general, not all the terminal constraints, (201) and (202), will be satisfied.

2. Calculate P from Eq. (205).

3. Integrate the adjoint equations, (208), "backwards," using the terminal conditions given by Eq. (209). Note that the matrix, A(t), is defined by Eq. (163).

4. Calculate the hamiltonian, Eq. (207), and the Green's functions, $\partial H/\partial u_\ell$.

5. Choosing some appropriate values for the weighting functions, $w_\ell(t)$, evaluate η^2 by Eq. (214).

6. Using an appropriate step size, ϵ (which must be a negative number), calculate δu_ℓ, with $\ell = 1,2,\cdots\cdot,m$. The resulting $u^{new} = u^{old} + \delta u$ must be "trimmed" to satisfy the inequality constraints, (203).

7. Using $u^{new}(t)$, obtain a new trajectory in the manner of step 1. The unspecified initial conditions are modified in accordance with Eq. (218).

8. Calculate P from Eq. (205), and check whether this is less than the value of P from the previous iteration. If not, reduce the size of ϵ and return to step 6.

9. Stop the iteration when some suitable accuracy criterion is satisfied. This may take the form

$$\eta^2 \lesssim \sigma$$

$$|\psi_\alpha| \leq \rho_\alpha \qquad \alpha = 1, 2, \cdots, p$$

$$\psi_\beta \leq 0 \qquad \beta = 1, 2, \cdots, q$$

where σ and ρ_α are prescribed constants.

Remark: Various arbitrary constants must be chosen in the computational procedure described above. The values for these constants will generally depend on the particular problem under investigation, and no one particular set of values is best in all cases. Nevertheless, certain guidelines are generally valid. First, as far as the K_γ of Eq. (205) are concerned, a large value will yield a new trajectory that will tend to satisfy the terminal constraints to a greater extent than it optimizes the criterion function, J. In short, the early iterations will seek to yield an admissible trajectory that satisfies all the terminal constraints before optimizing J. The value of K_γ may be periodically reduced in the course of the computation. This process indicates a willingness to accept mild excursions from the terminal constraints in favor of optimizing J, which is the primary goal.

The proper step size ϵ is difficult to determine initially, since it reflects the permissible linearity range. Too low a value means long computing time, while too large a value violates linearity and yields inaccurate results. A few preliminary trials should yield an acceptable range of values.

The weighting functions, $w_\ell(t)$, are introduced mainly to afford a greater degree of flexibility in the convergence procedure. Normally one may take $w_\ell(t) = 1$ for all ℓ and t. It may happen that near the optimum, large changes in P result from even small step sizes, ϵ. In this region, the $w_\ell(t)$ may be manipulated so that the δu_ℓ may be varied with time. This sometimes produces a smoother convergence to the optimum.

Example 4: We will now investigate the problem of maximizing the payload for a single-stage boost vehicle that is required to attain a prescribed altitude, velocity, and flight path angle. It is assumed that: the vehicle trajectory is contained in a plane passing through the center of the earth; the earth is spherical; and the gravity forces obey the inverse square law. In this case, the motion is described by (see Fig. 8)

$$m\dot{V} = t\cos\alpha - D - mg\sin\gamma \tag{219}$$

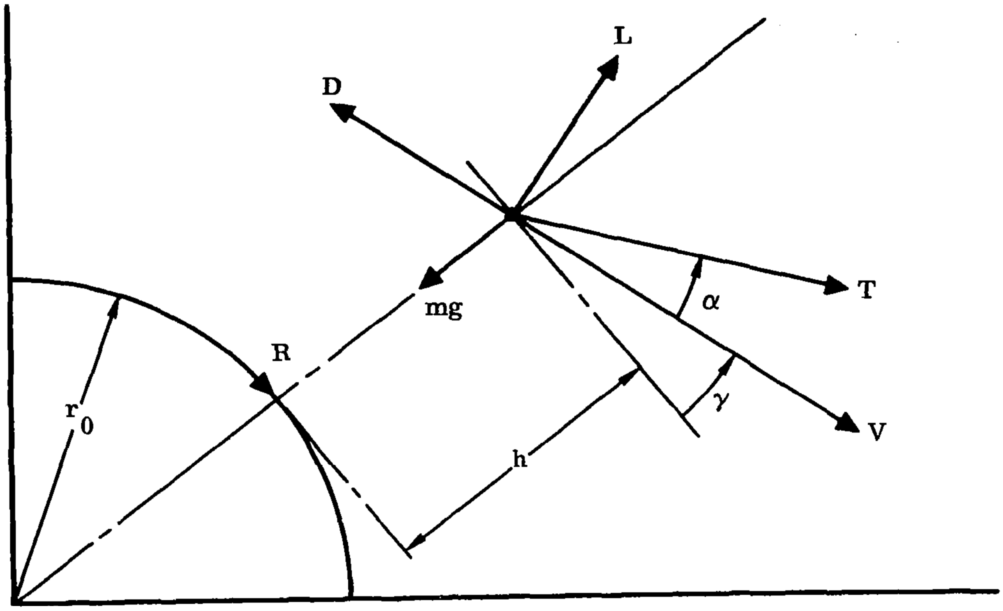


Figure 8. Coordinate System for Example 4

$$m V \dot{\gamma} = T \sin \alpha + L - m g \cos \gamma$$

$$+ \frac{m V^2 \cos \gamma}{r_0 + h} - 2\, m V\, \omega_E \cos \varphi_E \tag{220}$$

$$\dot{h} = V \sin \gamma \tag{221}$$

$$\dot{R} = \left(\frac{r_0}{r_0 + h} \right) V \cos \gamma \tag{222}$$

$$\dot{m} = - \beta \tag{223}$$

with

$$T = V_e \beta \tag{224}$$

$$g = g_0 \left(\frac{r_0}{r_0 + h} \right)^2 \tag{225}$$

The symbols have the following meaning.

C ≡ speed of sound; a known function of altitude

D ≡ drag; a known function of Mach number and angle of attack

g ≡ gravity acceleration; a known function of altitude

g_0 ≡ gravity acceleration at earth's surface

h ≡ altitude

L ≡ lift; a known function of Mach number and angle of attack

m ≡ instantaneous mass of vehicle

M ≡ Mach number = V/c

R ≡ range measured along earth's surface

r_0 ≡ radius of earth

T ≡ thrust

V ≡ velocity

V_e ≡ velocity of exhaust gases in rocket engine; a known constant

α ≡ angle of attack

β ≡ thrust variation parameter

γ ≡ flight path angle

φ_E ≡ angle of earth's polar axis with perpendicular to plane of motion

ω_E ≡ angular velocity of earth about polar axis

The control variables are β and α; i.e., the thrust magnitude and orientation.[†]

These are required to satisfy the inequality constraints

$$\alpha_1 \leqq \alpha \leqq \alpha_2 \tag{226}$$

$$0 \leqq \beta \leqq \beta_M \tag{227}$$

It is also required to limit the axial and normal loads as follows.

$$(T + L \sin\alpha - D \cos\alpha)/mg \leqq L_A \tag{228}$$

[†] Short-period dynamics are neglected; therefore the thrust angle and angle of attack are equivalent.

$$(L \cos \alpha + D \sin \alpha)/mg \le L_N \tag{229}$$

where L_A and L_N are prescribed constants.

The following boundary conditions are specified.

$$\left.\begin{array}{llll} V(t_i) = 0 & & & \\ \gamma(t_i) = \pi/2 & & V(t_f) = V_T & \\ R(t_i) = 0 & & \gamma(t_f) = \gamma_T & \\ h(t_i) = 0 & & h(t_f) = h_T & \\ m(t_i) = m_I & & & \\ t_i = 0 & & t_f \equiv \text{free} & \end{array}\right\} \tag{230}$$

We may therefore take, as a stopping condition,

$$\Omega \equiv h(t_f) - h_T = 0 \tag{231}$$

and the remaining terminal constraints become

$$\psi_1 \equiv V(t_f) - V_T = 0 \tag{232}$$

$$\psi_2 \equiv \gamma(t_f) - \gamma_T = 0 \tag{233}$$

The inequality constraints, (228) and (229), involve state variables. We therefore define

$$x_6 = \int_{t_i}^{t} G_1 \, dt \tag{234}$$

$$x_7 = \int_{t_i}^{t} G_2 \, dt \tag{235}$$

where

$$\begin{aligned} G_1 &= 0 \quad \text{if } g_1 \leq 0 \\ &= g_1^2 \quad \text{if } g_1 > 0 \end{aligned} \tag{236}$$

$$\begin{aligned} G_2 &= 0 \quad \text{if } g_2 \leq 0 \\ &= g_2^2 \quad \text{if } g_2 > 0 \end{aligned} \tag{237}$$

and

$$g_1 = \frac{T + L \sin\alpha - D \cos\alpha}{m g} - L_A \tag{238}$$

$$g_2 = \frac{L \cos\alpha + D \sin\alpha}{m g} - L_N \tag{239}$$

We add to the state equations

$$\dot{x}_6 = G_1 \tag{240}$$

$$\dot{x}_7 = G_2 \tag{241}$$

with initial conditions

$$x_6 (t_i) = 0 \tag{242}$$

$$x_7 (t_i) = 0 \tag{243}$$

In this fashion, the state variable inequality constraints, (228) and (229), are transformed to the terminal constraints.

$$\psi_3 \equiv x_6 (t_f) = 0 \tag{244}$$

$$\psi_4 \equiv x_7 (t_f) = 0 \tag{245}$$

If we let

$$x_1 = V \qquad\qquad u_1 = \alpha$$

$x_2 = \gamma \qquad u_2 = \beta$

$x_3 = h$

$x_4 = R$

$x_5 = m$

then the problem may be expressed in the standard format as follows.

Given the system

$$\dot{x}_1 = \frac{V_e u_2 \cos u_1 - D\left(x_1, x_3, u_1\right)}{x_5} - g_0 \left(\frac{r_0}{r_0 + x_3}\right)^2 \sin x_2 \equiv f_1 \qquad (246)$$

$$\dot{x}_2 = \frac{V_e u_2 \sin u_1 + L\left(x_1, x_3, u_1\right)}{x_1 x_5} - \frac{g_0}{x_1}\left(\frac{r_0}{r_0 + x_3}\right)^2 \cos x_2$$

$$+ \frac{x_1 \cos x_2}{r_0 + x_3} - 2\, \omega_E \cos \varphi_E \equiv f_2 \qquad (247)$$

$$\dot{x}_3 = x_1 \sin x_2 \equiv f_3 \qquad (248)$$

$$\dot{x}_4 = \left(\frac{r_0}{r_0 + x_3}\right) x_1 \cos x_2 \equiv f_4 \qquad (249)$$

$$\dot{x}_5 = - u_2 \equiv f_5 \qquad (250)$$

$$\dot{x}_6 = G_1 \equiv f_6 \qquad (251)$$

$$\dot{x}_7 = G_2 \equiv f_7 \qquad (252)$$

where

$$G_1 = 0 \text{ if } g_1 \leqq 0$$

$$= g_1^2 \text{ if } g_1 > 0$$

$$\begin{aligned} G_2 &= 0 \text{ if } g_2 \leq 0 \\ &= g_2^{\,2} \text{ if } g_2 > 0 \end{aligned}$$

$$g_1 = \left[\frac{V_e\, u_2 + L\left(x_1, x_3, u_1\right) \sin u_1 - D\left(x_1, x_3, u_1\right)}{x_5\, g_0\, r_0^{\,2}} \right] \left(r_0 + x_3\right)^2 - L_A$$

$$g_2 = \left[\frac{L\left(x_1, x_3, u_1\right) \cos u_1 + D\left(x_1, x_3, u_1\right) \sin u_1}{x_5\, g_0\, r_0^{\,2}} \right] \left(r_0 + x_3\right)^2 - L_N$$

with the terminal constraints

$$\psi_1 \equiv x_1\,(t_f) - V_T = 0 \tag{253}$$

$$\psi_2 \equiv x_2\,(t_f) - \gamma_T = 0 \tag{254}$$

$$\psi_3 \equiv x_6\,(t_f) = 0 \tag{255}$$

$$\psi_4 \equiv x_7\,(t_f) = 0 \tag{256}$$

stopping condition

$$\Omega \equiv x_3\,(t_f) - h_T = 0 \tag{257}$$

and initial conditions

$$\left.\begin{aligned} x_1\,(t_i) &= 0 & x_5\,(t_i) &= m_I \\ x_2\,(t_i) &= \pi/2 & x_6\,(t_i) &= 0 \\ x_3\,(t_i) &= 0 & x_7\,(t_i) &= 0 \\ x_4\,(t_i) &= 0 & & \end{aligned}\right\} \tag{258}$$

It is required to determine $u_1(t)$ and $u_2(t)$ such that

$$J = x_3\,(t_f) \tag{259}$$

is a maximum, subject to the control variable constraints

$$\alpha_1 \leq u_1 \leq \alpha_2 \tag{260}$$

$$0 \leq u_2 \leq \beta_M \tag{261}$$

The computational procedure now continues in the manner outlined previously.

3.1.4 Dynamic Programming

The theory of dynamic programming$^{(16)}$ is one of those rarities -- a genuinely original contribution to a long outstanding problem. As is often the case with the appearance of a new idea, a flood of papers dealing with extensions, ramifications, and applications has resulted. The theory has helped influence the development of such diverse areas as management science, information theory, sequential analysis, optimum filtering, adaptive control, learning theory, optimal trajectories, and variational calculus. The passage of time, however, has had a sobering effect on some of the overly optimistic claims made in the early stages of the theory's development. Space limitations preclude discussion of all facets of the theory save those immediately relevant to this monograph. We shall content ourselves with the barest outline of the main ideas and discuss a variety of applications that hopefully will clarify the method and exhibit its utility. In the next section we will show how the basic premise of dynamic programming includes, as a special case, each of the optimization techniques thus far considered.

Of fundamental concern in what follows will be the concept of a multistage decision process. At any one stage in this process, one may make a decision (select a control function), following which the next stage is reached. Successive stages are related by a known transformation. Each stage is characterized by a state vector, and each decision results in some cost. These ideas may be made explicit as follows.

Let the state of a dynamic process be characterized by a state vector, x. Suppose that successive states are related by

$$x^{(j+1)} = T\left(x^{(j)}, u^{(j)}\right) \tag{262}$$

where $u^{(j)}$ is a decision (control) vector, and T is a known transformation. Eq. (262) is a statement of the fact that when in stage (state) j, a control $u^{(j)}$ is chosen, and then the new stage $x^{(j+1)}$ is reached, depending on the explicit form of $T\left(x^{(j)}, u^{(j)}\right)$. Assume there are N stages in the process, and a performance index (cost) is given by

$$J = \sum_{j=1}^{n} g\left(x^{(j)}, u^{(j)}\right) \tag{263}$$

We pose the problem of choosing the control sequence, $u^{(1)}, u^{(2)}, \cdots, u^{(N)}$, such that the function J is a minimum (maximum). For ease of discussion, let us call any permissible control sequence, $u^{(1)}, u^{(2)}, \cdots, u^{(N)}$, a policy. If the initial state of the process is characterized by the state vector, c, the value of J is obviously a function of c and the policy (for a given value of N). Let us now define

$W_N(c) \equiv$ the value of J when the process starts in state c, having N stages to go, and using an optimal policy.

An optimal policy, of course, is one that optimizes the cost function, J. The fundamental premise of dynamic programming is embodied in the following.

> The Principle of Optimality:[16] An optimal policy has the property that whatever the initial state and the initial decision, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

The principle of optimality is used to derive, for the $W_j(c)$, recurrence relations that yield both the optimal policy and the value of the optimal cost function. The reasoning proceeds as follows. We are in the initial state, c, and have N stages to go. We pick an initial value of the control vector, $u^{(1)}$. As a result, we reach a new state,

$$x^{(2)} = T\left(c, u^{(1)}\right) \tag{264}$$

and incur a cost, $g\left(c, u^{(1)}\right)$, with (N-1) stages left in the process. If we proceed in an optimal fashion from this new state, the cost incurred is $W_{N-1}\left[T\left(c, u^{(1)}\right)\right]$, by definition. By way of the principle of optimality, we see that for some specific value of $u^{(1)}$ we incur a cost

$$g\left(c, u^{(1)}\right) + W_{N-1}\left[T\left(c, u^{(1)}\right)\right]$$

But we wish to choose $u^{(1)}$ so that it is maximized.† Therefore

$$W_N(c) = \underset{u^{(1)}}{\text{Max}} \left\{ g\left(c, u^{(1)}\right) + W_{N-1}\left[T\left(c, u^{(1)}\right)\right] \right\} \tag{265}$$

† For definiteness, we consider the problem of maximizing J. The argument, of course, holds equally well for minimization.

Consider now the function, $W_1(c)$; i.e., the optimal value of J with only one stage to go. It is easy to see that

$$W_1(c) = \underset{u^{(1)}}{\mathrm{Max}}\, g\left(c, u^{(1)}\right) \tag{266}$$

For purposes of computation, one proceeds as follows. Assume that $u^{(1)}$ may take any one of a prescribed finite set of values. Then calculate $W_1(c)$, using (266), for a range of values of c. This yields a table of $W_1(c)$, each with its corresponding optimal $u^{(1)}$ which is then utilized to calculate $W_2(c)$ via (265) for a range of values of c, along with its corresponding optimal $u^{(1)}$, etc. Proceeding in this fashion, we find $W_1(c)$, $W_2(c), \cdots, W_N(c)$, together with the optimal policy for each. Note that this is a <u>flooding technique</u>; in effect, we solve the specific problem by imbedding it in a class of related problems. This is a highly efficient technique but has one major limitation. It is what Bellman refers to as the "curse of dimensionality." Suppose that the vector, x (or c), has n components, and assume that it is required to evaluate 100 different values for each component. Then to store $W_j(c)$ in the computer would require 100^n storage cells. Since modern-day computers have on the order of 30,000 storage cells, we see that n=3 is a marginal figure. Various means of overcoming this limitation are discussed in Refs. 9, 10, 168, 169, and 170. We will digress for a moment to consider an elementary application.

<u>Example 5</u>: Given the system

$$\dot{x} = x^2 + u \tag{267}$$

where x and u are both <u>scalars</u>, determine u(t) such that the function

$$J = \int_0^{t_f} |x - u^3|\, dt \tag{268}$$

is a minimum; the final time, t_f, is given. The control function u(t) must satisfy the inequality constraint

$$-2 \leq u \leq 2 \tag{269}$$

This simple, rather contrived problem illustrates the computational procedure using dynamic programming and also exhibits both the power and limitations of the method. The problem is, first of all, nonlinear, having a cost function that is not everywhere continuous, and must also satisfy inequality constraints on the control variable. These features introduce distressing complications in the classical approach. The dynamic programming approach, however, is completely indifferent to analytic aberrations.

To apply dynamic programming, the problem must first be cast in the form of a multistage decision process, which means that we must use the discrete versions of (267) and (268); viz.,

$$dx = \left(x^2 + u\right) dt \tag{270}$$

$$J = \sum_{k=1}^{n} |x_k - u_k^3| dt \tag{271}$$

$$t_f = N\,dt \tag{272}$$

This respresentation is valid as long as the increment, dt, is sufficiently small. Successive states are then related by

$$x^{(j+1)} = x^{(j)} + \left[\left(x^{(j)}\right)^2 + u^{(1)}\right] \Delta \tag{273}$$

$$\Delta \equiv dt$$

which is the transformation law, (262), for this problem.

Now define

$W_N(c)$ = the value of J, Eq. (271), obtained by starting in state c, having N stages to go, and using an optimal policy.

A direct application of the principle of optimality yields the recurrence relation

$$W_N(c) = \underset{u}{\text{Min}} \left\{ |c - u^3| \Delta + W_{N-1} \left[c + \left(c^2 + u\right)\Delta\right]\right\} \tag{274}$$

This simply means that in state c, with N stages to go, if we choose a particular (admissible) u, then we incur a cost $|c - u^3|\Delta$ for this first stage and end in state $[c + (c^2 + u)\Delta]$. Assuming that we proceed optimally thereafter, the remaining cost is, by definition, $W_{N-1}\,[c + (c^2 + u)\Delta]$ for the remaining (N-1) stages. However, we choose not an arbitrary u for the first stage, but the one that minimizes this total cost. Hence the form of Eq. (274).

It is easy to see that when there is only one stage to go, we have

$$W_1(c) = \underset{u}{\text{Min}}\; | c - u^3|\Delta \tag{275}$$

and this represents the "initial condition" for the recurrence relation (274).

We now examine the computational sequence in some detail. To do this, we must decide on the range of interest for the state variables and the number of values of u in the range $-2 \lesssim u \lesssim 2$. For present purposes, we will consider a painfully simplified version wherein

$$c = -5, -4, -3, -2, -1, 0, 1, 2, 3, 4, 5$$

$$u = -2, -1, 0, 1, 2; \quad \Delta = 0.1$$

Then via (275), we calculate the table

c	$W_1(c)$	u
-5	0.3	-2
-4	0.3	-1
-3	0.2	-1
-2	0.1	-1
-1	0	-1
0	0	0
1	0	1
2	0.1	1
3	0.2	1
4	0.3	1
5	0.3	2

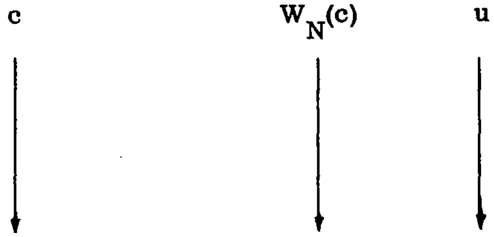
Normally, in realistic problems, this operation is performed on a computer, and the table is printed out. Now we calculate $W_2(c)$ from

$$W_2(c) = \underset{u}{\text{Min}} \left\{ |c - u^3| \, \Delta + W_1 \left[c + \left(c^2 + u \right) \Delta \right] \right\}$$

for a range of values of c. For example, with c = 3, we calculate the bracketed quantity for each of the five permissible values of u; we then select the one that is a minimum as the value of $W_2(3)$. This yields, for example,

$$W_2(3) = 0.50; \quad u = 1$$

Proceeding iteratively, we obtain, ultimately, the tabular form



which represents, along with the tabular forms for $W_j(c)$, $j = 1, 2, \cdots, N-1$, the *solution to the problem*. Note that we have the solution, not only for one specific initial condition, c, but a whole range of possible initial conditions. This is characteristic of the dynamic programming method.

We have thus far considered only "running costs" of the type exemplified by Eq. (263). It may be desired to minimize (or maximize) a function of the final state

$$J = g\left[x(t_f)\right] \tag{276}$$

In this case, we define

$W_N(c)$ = the value of $g\left[x(t_f)\right]$ starting in state c, having N stages to go, and using an optimal policy.

A simple argument shows that the recurrence relations are

$$W_N(c) = \underset{u}{\text{Min}}\; W_{N-1}\left[T(c, u)\right] \tag{277}$$

$$W_0(c) = g(c) \tag{278}$$

In other words, with zero stages to go, the optimal $W_0(c)$ is merely the value $g\left[x(t_f)\right]$, where c replaces $x(t_f)$.

For the functional equations thus far considered, it was possible to write explicit initial conditions; that is, the number of stages in the process was known beforehand. This is not always the case. To examine this situation, consider the problem of driving the solution of

$$\dot{x}_1 = x_2$$

$$\dot{x}_2 = h\left(x_1, x_2, u_1\right)$$

$$x_1(0) = c_1$$

$$x_2(0) = c_2$$

to the equilibrium state, $x_1 = x_2 = 0$, in minimum time, where the control, u_1, is constrained by

$$a \leq u_1 \leq b$$

We consider the discrete version

$$x_1\ (t + \Delta) = x_1(t) + x_2(t)\Delta$$

$$x_2\ (t + \Delta) = x_2(t) + h\left[x_1(t),\ x_2(t),\ u_1(t)\right]\Delta$$

$$\Delta \equiv dt$$

and instead of requiring that x_1 and x_2 be simultaneously zero, it is sufficient to require that $(x_1^2 + x_2^2) \leq \epsilon$, where ϵ is some small positive number.

If we define

$W\ (c_1,\ c_2) \equiv$ the time required to reach the equilibrium region $(x_1^2 + x_2^2) \leq \epsilon$, starting in state (c_1, c_2), and using an optimal policy,

then a simple application of the principle of optimality yields the functional equation

$$W\ (c_1,\ c_2) = \underset{u_1}{\text{Min}} \left\{ \Delta + W\left[c_1 + c_2\Delta,\ c_2 + h\left(c_1,\ c_2,\ u_1\right)\Delta\right]\right\} \qquad (279)$$

The computational procedure for this equation is not immediately apparent, since we are unable to set up "initial" conditions. This equation is of the implicit type; that is, the unknown function appears on both sides of the equation. It is not a recurrence relation of the type previously considered.

We may, nevertheless, obtain a solution via the concept of approximation in policy space, a powerful technique first discovered by Bellman.(16) It is crudely analogous to and predates the gradient technique discussed in Sec. 3.1.3. One proceeds as follows. Select a policy, $u^{(0)}(t)$, that brings the system to the equilibrium state, but that in general, does not result in minimum time. Let this time duration be denoted by $W^{(0)}(c_1,c_2)$, which is a first approximation to the optimal return function, $W(c_1,c_2)$. Do this for a range of values of c_1 and c_2. Calculate an improved return function, $W^{(1)}(c_1,c_2)$, via

$$W^{(1)}\left(c_1, c_2\right) = \underset{u_1}{\text{Min}}\left\{\Delta + W^{(0)}\left[c_1 + c_2\Delta, \; c_2 + h\left(c_1, c_2, u_1\right)\Delta\right]\right\}$$

also for a range of values of c_1 and c_2, storing the improved policy $u_1^{(1)}$. It can be shown(16) that

$$W^{(j)}\left(c_1, c_2\right) \lessgtr W^{(j-1)}\left(c_1, c_2\right)$$

and the iteration stops when

$$\left|W^{(j)}\left(c_1, c_2\right) - W^{(j-1)}\left(c_1, c_2\right)\right| \lessgtr \eta$$

η = small positive constant.

The number and type of optimization problems that can be solved by the simple methods outlined above is truly astonishing. References 16, 17, and 159 contain numerous examples, together with an extensive bibliography. We now consider some simple, though fairly realistic, aerospace applications.

Example 6: We return to the sounding rocket problem already analyzed via the variational calculus (Example 1) and the maximum principle (Example 2). In Example 1 it was found that the optimal thrust control program may contain an ambiguity† for certain forms of the drag function. (See Fig. 3.) This ambiguity is completely resolved in the dynamic programming solution.

We now proceed to express the problem in the format of dynamic programming, using the notation defined in Example 1. The state transition equations are written in discrete form as follows.

$$dV = \left[\frac{\beta\mu - D(V, h)}{m} - g(h)\right]\Delta \tag{280}$$

$$dh = V \tag{281}$$

$$\Delta \equiv dt$$

†This ambiguity also exists via the maximum principle.

We have to take account of the fact that the control function satisfies the relation

$$\beta = -\dot{m} \tag{282}$$

and the constraint

$$0 \lesseqgtr \beta \lesseqgtr \beta_M \tag{283}$$

where β_M is given. Let us suppose that admissible values of β may be taken from the q distinct values

$$\beta \equiv \left\{ \beta_0, \beta_1, \cdots, \beta_q \right\} \tag{284}$$

where

$$\beta_0 = 0$$

$$\beta_q = \beta_M, \text{ the permissible limit}$$

and that

$$\beta_{j+1} - \beta_j = \gamma, \text{ a prescribed positive constant}$$

Now a particular β_k corresponds to a particular mass increment $(d\,m)_k$ determined from

$$\beta_k = -\frac{(d\,m)_k}{\Delta} \tag{285}$$

We now select Δ, γ, and q such that

$$\beta_k = -k \tag{286}$$

where k is a positive integer (or zero) that is interpreted as a number of mass (i.e., fuel) increments. If, at time zero, the total mass is

$$m\,(0) = m_0 + m_F\,(0)$$

where

$m_0 \equiv$ fixed mass (structure and payload)

$m_F \equiv$ fuel mass

then we find the number of "stages" in the process as

$$m_F(0) = N$$

In other words, the size of the mass increments having been established, a particular fuel mass can be expressed as a specified number of these mass increments. In particular, we denote the number of mass increments available at the start of the process by N.

The problem is to determine a thrust control policy that maximizes the final altitude, given a prescribed fuel mass. Define

$W_N(V,h)$ = the altitude attained starting with velocity V, at altitude h, having available N fuel increments, and using an optimal policy.

The functional equation is

$$W_N(V,h) = \underset{\beta_k}{\text{Max}} \left\{ V\Delta + W_{N-k}\left[V + \left(\frac{\beta_k \mu - D}{m_0 + N} - g \right)\Delta,\ h + V\Delta \right] \right\} \qquad (287)$$

The reasoning is as follows. Starting in state (V,h), we acquire an altitude increment, $V\Delta$, and reach a new state

$$V + \left(\frac{\beta_k \mu - D}{m_0 + N} - g \right)\Delta$$

$$h + V\Delta$$

Proceeding optimally thereafter, the altitude acquired is

$$W_{N-k}\left[V + \left(\frac{\beta_k \mu - D}{m_0 + N} - g \right)\Delta,\ h + V\Delta \right]$$

by definition. Note that there now remain (N-k) stages, since the choice β_k means that k fuel increments were expended. But we may choose β_k to maximize this return. Hence the form of Eq. (287). The initial condition, W_0 (V,h), is easily obtained, since with zero stages to go, the altitude acquired is simply the solution of the system

$$\dot{V} = - \frac{D\,(V,\ h)}{m_0} - g\,(h)$$

$$\dot{h} = V$$

where the initial conditions on V and h are merely the arguments of $W_0(V,h)$, and the integration terminates when V = 0. Denoting the altitude acquired during this coast by h_T, we see that

$$W_0 (V, h) = h_T (V, h)$$

the notation h_T (V, h) emphasizing that the altitude acquired during coast is a function of the starting values of V and h. If the altitude is sufficiently high such that the drag is negligible, then

$$W_0 (V, h) = \frac{V^2}{2g}$$

The computational procedure is now straightforward. Some care must be exercised, however, in the choice of increments. One must choose $\Delta \equiv dt$ small enough, for example, such that (280) and (281) represent good approximations to the continuous case. If we consider the numerical values

$$m (0) = 1000 \text{ slugs}$$
$$m_F(0) = 900 \text{ slugs}$$
$$\mu = 2000 \text{ ft/sec}$$
$$\beta_M = 25 \text{ slugs/sec}$$

then we may take

$$\Delta \equiv dt = 0.1 \text{ sec}$$
$$\beta \equiv \left\{0, 0.1, 0.2, \cdots, 25\right\}$$

mass increment = 0.1 slug

$$N = 9000$$

The solution time for this problem would be on the order of a few minutes on a computer of the 7090 type.

Example 7: One of the earliest studies dealing with the application of dynamic programming to trajectory optimization is the following.(160)

Let the motion of an aircraft in the vertical plane be described by (see Fig. 9)

$$m\dot{V} = T - D(V, h) - mg \sin\theta \tag{288}$$

$$\dot{h} = V \sin\theta \tag{289}$$

where

$V \equiv$ velocity

$D \equiv$ drag

$T \equiv$ thrust (constant)

$m \equiv$ mass (constant)

$g \equiv$ gravity acceleration

$h \equiv$ altitude

$\theta \equiv$ angle of thrust vector with respect to horizontal reference

We pose the problem of programming the thrust vector angle, θ, such that the aircraft attains a prescribed velocity and altitude, starting from some given initial velocity and altitude, in minimum time.

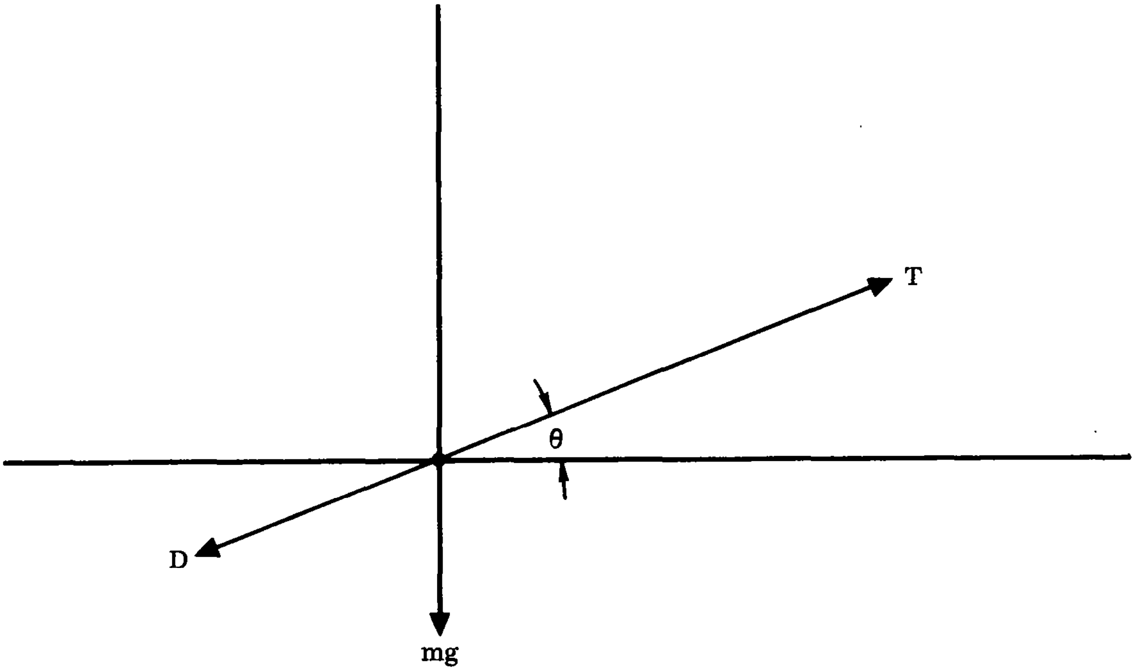


Figure 9. Coordinate System for Example 7

Taking the discrete version of (288) and (289),

$$d\,V = \left[\frac{T - D\,(V,\,h)}{m} - g\sin\theta\right]\Delta \tag{290}$$

$$d\,h = V\Delta\sin\theta \tag{291}$$

$$\Delta \equiv d\,t$$

and defining

$W\,(V,\,h)$ = the time required to reach the prescribed terminal state (V_T, h_T) starting in state (V, h) and using an optimal policy

we obtain, via the principle of optimality,

$$W\,(V,\,h) = \Delta + \underset{\theta}{\text{Min}}\left[V + \left(\frac{T - D}{m} - g\sin\theta\right)\Delta,\ h + V\Delta\sin\theta\right] \tag{292}$$

This is an implicit type of functional equation that must be solved iteratively by the method of approximation in policy space.

Viewing the problem in a different light, Cartaino and Dreyfus$^{(160)}$ obtained a recurrence relation for W (V,h) for which initial conditions could be prescribed, thereby circumventing the need to solve an equation of the explicit type. Their reasoning is based on the observation that the given problem is equivalent to that of determining the minimum time path in the h-V plane. From Eqs. (290) and (291), we see that

$$\Delta = \frac{\left(1 + \frac{V}{g}\frac{d\,V}{d\,h}\right)d\,h}{\left(\frac{V}{w}\right)\left(T - D\right)} \tag{293}$$

$$w = m\,g$$

Imagine now that the h-V plane is subdivided into grids with increments d h and dV. We suppose that the mass point representing the airplane moves in discrete steps from one node to the next, either straight up (V = constant) or horizontally to the right (h = constant). At each node point, a decision must be made whether to proceed vertically up or horizontally to the right. The criterion, of course, is that the grid from the current (initial) V and h to the prescribed V_T and h_T be traversed in minimum time. The "costs" associated with the respective alternatives are

$$\Delta = \frac{w\,dh}{V\,(T - D)} \qquad \text{for constant V}$$

$$= \frac{w\,dV}{g\,(T - D)} \qquad \text{for constant h} \tag{294}$$

via (293).

Let us note at the outset that the insensitive pragmatist might be tempted to enumerate all possible paths and then select the minimum time path by direct comparison. In the case of a p by q grid, the total number of possible paths under the conditions stipulated is

$$\frac{(p + q)!}{(p - 1)!\,(q + 1)!}$$

For a grid with p = q = 100, the number of paths to be enumerated is on the order of 10^{59}. Since this closely approaches the estimated number of atoms in the galaxy, one is compelled to seek another approach.

As a matter of fact, via the principle of optimality, with W (V,h) defined as before, we find

$$W\,(V,\,h) = \text{Min} \begin{cases} \text{A.} \quad \dfrac{w\,dh}{V\,(T - D)} + W\left[V,\, h + dh\right] \\ \\ \text{B.} \quad \dfrac{w\,dV}{g\,(T - D)} + W\left[V + dV,\, h\right] \end{cases} \tag{295}$$

This is arrived at as follows. If we are at the node in the grid whose coordinates are V and h, we have the option of moving vertically upward (constant V), in which case we incur the cost

$$\frac{w\,dh}{V\,(T - D)}$$

(i.e., this is the time consumed in going from h to h + dh) and arrive at the new grid point, (V, h + dh). Proceeding optimally thereafter, the cost is W (V, h + dh). This is alternative A. Alternative B is arrived at by similar reasoning. Finally W (V,h) is taken as the smaller of the two alternatives A and B. Since the terminal values of velocity and altitude, V_T and h_T, are specified, we see that

$$W\left(V_T - dV,\, h_T\right) = \frac{w\,dV}{g\left[T - D\left(V_T,\, h_T\right)\right]} \tag{296}$$

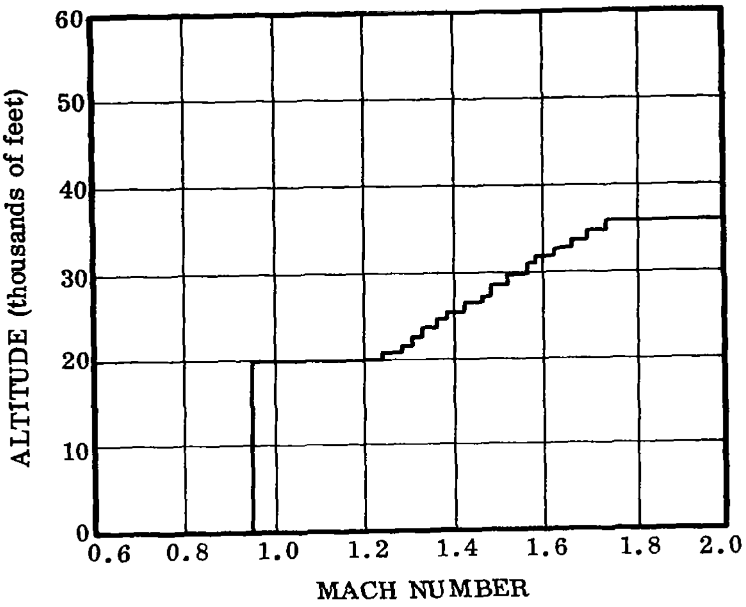


Figure 10. Altitude/Mach Number Trace of Minimum-Time Path

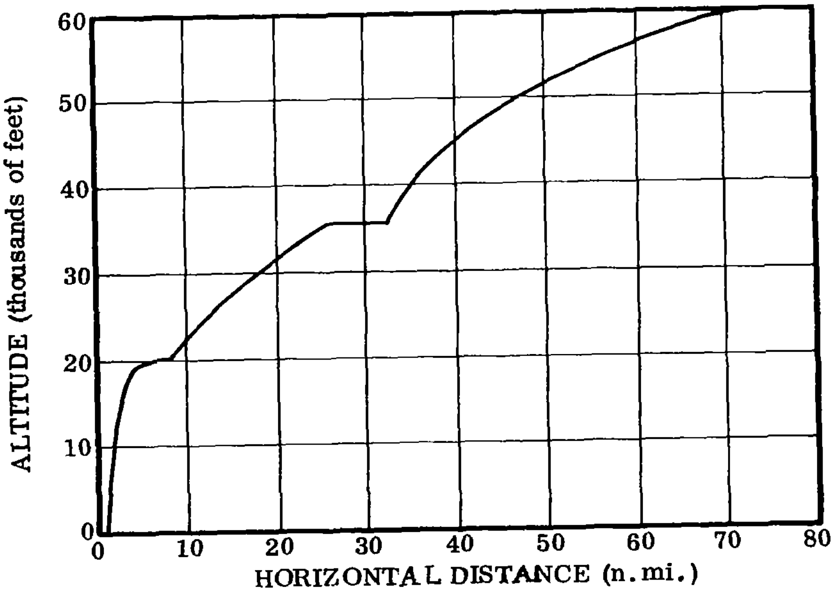


Figure 11. Altitude/Horizontal Distance Profile of Minimum Time Path

$$W\left(V_T, h_T - dh\right) = \frac{w\,dh}{V_T\left[T - D\left(V_T, h_T\right)\right]} \tag{297}$$

These represent the initial conditions for the recurrence relation (295).

It is easy to show that the number of paths to be enumerated in this approach is on the order of pq, compared with the astronomical number of paths to be enumerated by the brute-force approach.

Figs. 10 and 11, abstracted from Ref. 160, show some typical results. From Fig. 10, Eq. (289), and the relation

$$\dot{x} = V \cos \theta$$

$$x \equiv \text{horizontal distance}$$

we may plot the optimal path in the h-x plane, which is depicted in Fig. 11. The constant Mach number and constant altitude subarcs for the optimal path are well delineated in Fig. 10. The unique feature of the present approach is that the complexity of the drag function presents no difficulty whatever, whereas in the classical approach, one may encounter all sorts of analytical obstacles.

3.1.5 Unifying Principles

In the methods presented thus far, only the main results have been stated, together with various examples illustrating their application. The development of detailed derivations in each case would require a document the size of a textbook rather than a monograph. Nevertheless, the omission of derivations constitutes a pedagogic deficiency that precludes complete comprehension. Furthermore, a superficial examination would seem to indicate that the four optimization techniques considered thus far are basically unrelated. It may, in fact, be shown that not only are these techniques intimately related, but that they all fall out as special cases from a more general point of view -- namely, they are all necessary consequences of the principle of optimality. We proceed to demonstrate this in the following way.

Let the system dynamics be given by†

$$\dot{x} = f(x, u) \tag{298}$$

$$x(t_i) = c \tag{299}$$

†An explicit dependence of f on t may be removed by defining an additional state variable as noted earlier.

where, as before, x and u are the state and control vectors respectively. It is required to choose u(t) such that a function of the terminal state

$$J = J\left[x\,(t_f)\right] \tag{300}$$

is minimized.† The control vector must satisfy the inequality constraints

$$\nu_j \lesssim u_j \lesssim \mu_j \qquad j = 1, 2, \cdots, m \tag{301}$$

and a stopping condition is given by

$$\Omega\left[x\,(t_f)\right] = 0 \tag{302}$$

There also exist terminal constraints of the form

$$\psi_\alpha\left[x\,(t_f)\right] = 0 \qquad \alpha = 1, 2, \cdots, q\;(<n) \tag{303}$$

As noted earlier, this is a very general format, since a variety of performance functions may be transformed to type (300), while state variable inequality constraints may be transformed to terminal constraints by defining additional state variables. (See Sec. 3.1.3.)

We now express (298) in the discrete form

$$dx = f\,(x, u)\Delta \tag{304}$$

$$x\,(\Delta) = c + f\,(x, u)\Delta \tag{305}$$

$$\Delta \equiv dt$$

and define

W(x) = the value of J, Eq. (300), when the terminal conditions, (302) and (303), are satisfied, starting in state x and using an optimal policy.

Applying the principle of optimality, we obtain the functional equation

$$W(x) = \operatorname*{Min}_{u} \left\{ W\left[x + f\,(x, u)\Delta\right]\right\} \tag{306}$$

†The arguments that follow are identical if "minimized" is replaced by "maximized."

Via a Taylor expansion,

$$W\left[x + f\,(x,\,u)\,\Delta\right] = W\,(x) + \sum_{j=1}^{n} \frac{\partial\,W(x)}{\partial x_j}\; f_j\,(x,\,u)\,\Delta\; + 0\;(\Delta^2)$$

Substituting in Eq. (306) and taking the limit as Δ→ 0, we obtain

$$0 = \operatorname*{Min}_{u} \left\{ \left[\frac{\partial\,W(x)}{\partial x}\right]^{T} f\,(x,\,u) \right\} \tag{307}$$

This is the partial differential equation that must be satisfied by the optimal return function.

Let us now define

$$\lambda_j = - \frac{\partial\,W}{\partial\,x_j} \tag{308}$$

or, in vector form,

$$\lambda = - \frac{\partial\,W}{\partial\,x} \equiv -\nabla_x\,W \tag{309}$$

and

$$H = \lambda^T f \tag{310}$$

The notation anticipates the interpretation of λ and H as the Lagrange multiplier and hamiltonian respectively. Then from Eq. (307),

$$0 = \operatorname*{Min}_{u} \left\{ \left(\frac{\partial\,W}{\partial\,x}\right)^{T} f \right\} = \operatorname*{Min}_{u} \left\{ -\,\lambda^T f \right\} = \operatorname*{Min}_{u}\;(-H)$$

and finally,

$$0 = \operatorname*{Max}_{u}\,H \tag{311}$$

This states that the control, u, that minimizes J must necessarily maximize H; in other words, Eq. (311) is a concise statement of the maximum principle.

We may express equation (307) by the two equivalent equations

$$\left[\frac{\partial W(x)}{\partial x}\right]^T f\,(x,\,u) \;=\; 0 \tag{312}$$

$$\frac{\partial}{\partial u}\left\{\left[\frac{\partial W(x)}{\partial x}\right]^T f\,(x,\,u)\right\} \;=\; 0 \tag{313}$$

This last equation may be written as

$$\left[\frac{\partial W(x)}{\partial x}\right]^T \left[\frac{\partial f\,(x,\,u)}{\partial u}\right] \;=\; 0 \tag{314}$$

if there are no constraints on u. In the development that follows, it will be assumed that such a constraint has been removed by defining an additional state variable, together with an added differential constraint in the manner discussed in Sec. 3.1.1.

$$\frac{d}{dt}\left[\frac{\partial W}{\partial x_j}\right] = \sum_{k=1}^{n} \frac{\partial^2 W}{\partial x_j\,\partial x_k} \cdot \frac{dx_k}{dt} = \sum_{k=1}^{n} f_k\,(x,u)\,\frac{\partial}{\partial x_j}\left(\frac{\partial W}{\partial x_k}\right) \tag{315}$$

A partial differentiation of Eq. (312) with respect to x_j yields

$$\frac{\partial}{\partial x_j}\left[\sum_{k=1}^{n} \frac{\partial W}{\partial x_k}\,f_k\right] = \sum_{k=1}^{n}\left[f_k\,\frac{\partial}{\partial x_j}\left(\frac{\partial W}{\partial x_j}\right) + \left(\frac{\partial W}{\partial x_k}\right)\left(\frac{\partial f_k}{\partial x_j}\right)\right]$$

$$= \sum_{k=1}^{n}\left\{f_k\,\frac{\partial}{\partial x_j}\left(\frac{\partial W}{\partial x_k}\right) + \frac{\partial W}{\partial x_k}\left[\frac{\partial f_k}{\partial x_j} + \sum_{\ell=1}^{m} \frac{\partial f_k}{\partial u_\ell}\,\frac{\partial u_\ell}{\partial x_j}\right]\right\} = 0 \tag{316}$$

Combining this with (315),

$$\frac{d}{dt}\left(\frac{\partial W}{\partial x_j}\right) + \sum_{k=1}^{n}\left\{\frac{\partial W}{\partial x_k}\left[\frac{\partial f_k}{\partial x_j} + \sum_{\ell=1}^{m} \frac{\partial f_k}{\partial u_\ell}\,\frac{\partial u_\ell}{\partial x_j}\right]\right\} = 0 \tag{317}$$

But Eq. (313) can be written as

$$\frac{\partial}{\partial u_\ell}\left[\sum_{k=1}^{n} \frac{\partial W}{\partial x_k}\,f_k\right] = 0$$

$$\ell = 1,\,2,\,\cdots,\,m$$

or

$$\sum_{k=1}^{m} \frac{\partial W}{\partial x_k} \frac{\partial f_k}{\partial u_\ell} = 0 \qquad \ell = 1, 2, \cdots, m \tag{318}$$

Substituting this in (316), we find

$$\frac{d}{dt}\left(\frac{\partial W}{\partial x_j}\right) + \sum_{k=1}^{n} \frac{\partial W}{\partial x_k} \frac{\partial f_k}{\partial x_j} = 0 \qquad j = 1, 2, \cdots, n \tag{319}$$

which reduces to

$$\dot{\lambda}_j + \sum_{k=1}^{n} \lambda_k \frac{\partial f_k}{\partial x_j} = 0 \qquad j = 1, 2, \cdots, n \tag{320}$$

by combining with (308). It is easy to show that this is an alternate form of the Euler Lagrange equations.

For this purpose, consider the augmented function defined by Eq. (13); viz.,

$$F = \sum_{k=1}^{n} \lambda_k \left(\dot{x}_k - f_k\right)$$

We obtain directly

$$\frac{\partial F}{\partial \dot{x}_j} = \lambda_j$$

$$\frac{\partial F}{\partial x_j} = - \sum_{k=1}^{n} \lambda_k \frac{\partial f_k}{\partial x_j} \qquad j = 1, 2, \cdots, n$$

Consequently, the Euler Lagrange equations

$$\frac{d}{dt}\left(\frac{\partial F}{\partial \dot{x}_j}\right) - \frac{\partial F}{\partial x_j} = 0 \qquad j = 1, 2, \cdots, n$$

may be expressed as

$$\dot{\lambda}_j + \sum_{k=1}^{n} \lambda_k \frac{\partial f_k}{\partial x_j} = 0 \qquad j = 1, 2, \cdots, n$$

which is completely identical with (320).

This leads us to interpret λ_j as the rate of change of the optimal return function with respect to the state variable, x_j.

We have already shown (in Sec. 3.1.3) that the Euler Lagrange operator

$$\frac{d}{dt}\left(\nabla_{\dot{x}} - \nabla_x\right)$$

may be interpreted as a generalized gradient. Furthermore, condition (311) is an alternate expression for Eq. (183), wherein $(\partial H/\partial u)$ is another type of gradient (the Green's functions).

We have shown, therefore, that the basic results embodied in the variational calculus, the gradient technique, and the maximum principle are all necessary consequences of the principle of optimality in dynamic programming.

3.2 STANDARD SOLUTIONS

In certain well defined optimization problems, it is possible to obtain closed-form solutions. If the format of these problems is sufficiently general to include as special cases a wide variety of situations of practical interest, it is tempting to refer to these as standard solutions. These results are useful not only because they solve specific problems, but also because they may be taken as first approximations to more complex problems. This theme will be further developed in Sec. 3.3.

3.2.1 Linear Problems

The most tractable optimization problem, from an analytic point of view, is one whose dynamics are governed by a set of linear differential equations with a performance index of the quadratic type. This is expressed mathematically by

$$\dot{x} = Ax + Bu \qquad (321)$$

$$y = Gx \qquad (322)$$

$$x(0) = c \qquad (323)$$

$$J = \int_0^{t_f} \left(x^T G^T Q G x + u^T R u\right) dt + x^T(t_f) M x(t_f) \qquad (324)$$

where

$x \equiv$ n-dimensional state vector

$u \equiv$ m-dimensional control vector

$y \equiv$ q-dimensional measurement vector

$c \equiv$ initial condition vector

$J \equiv$ performance index (cost); a scalar

$A \equiv n \times n$ constant matrix

$B \equiv n \times m$ constant matrix

$G \equiv q \times n$ constant matrix

$Q \equiv q \times q$ constant matrix (symmetric)

$R \equiv m \times m$ constant matrix (symmetric)

$M \equiv n \times n$ constant matrix (symmetric)

The problem is usually stated in the following form. Given the system, (321) and (322), with the initial condition (323), determine the control vector, u(t), such that the performance index, (324), is a minimum.†

We define the optimal return function as follows.

$$W(x, t) = \underset{u}{\text{Min}} \left\{ \int_t^{t_f} \left(x^T G^T Q G x + u^T R u\right) dt + x^T(t_f) M x(t_f) \right\} \qquad (325)$$

Note that W(x,t) is a function of the current state and the current time. Applying the principle of optimality, we find that W(x,t) satisfies the recurrence relation

$$W(x, t) = \underset{u}{\text{Min}} \left[\left(x^T G^T Q G x + u^T R u\right) \Delta + W\left(x + dx, t + \Delta\right) \right] \qquad (326)$$

$\Delta \equiv dt$

†The arguments are completely analogous if "maximum" replaces "minimum."

Expressing the second term in the brackets by its Taylor expansion, we have

$$W\left(x + dx,\ t + \Delta\right) = W\ (x,\ t) + \left[\frac{\partial W\ (x,\ t)}{\partial x}\right]^T dx + \frac{\partial W}{\partial t}\Delta$$

$$= W\ (x,\ t) + \left[\frac{\partial W\ (x,\ t)}{\partial x}\right]^T (Ax + Bu)\Delta + \frac{\partial W}{\partial t}\ \Delta$$

retaining only first-order terms in Δ. Substituting this in (326), dividing through by Δ, and letting Δ→0, we obtain

$$-\frac{\partial W}{\partial t} = \underset{u}{\text{Min}}\left[\left(x^T G^T QGx + u^T Ru\right) + \left(\frac{\partial W}{\partial x}\right)^T (Ax + Bu)\right] \qquad (327)$$

From (325), we see that a boundary value is given by

$$W\left(x,\ t_f\right) = x^T(t_f)\ Mx\ (t_f) \qquad (328)$$

Eq. (327) is the Hamilton-Jacobi equation for the system which could also be derived from the maximum principle. Despite its formidable appearance, Eq. (327) has a closed-form solution that may be readily obtained as follows. From (327), we write the two equations,

$$x^T G^T QGx + u^T Ru + \left(\frac{\partial W}{\partial x}\right)^T (Ax + Bu) = -\frac{\partial W}{\partial t} \qquad (329)$$

$$\frac{\partial}{\partial u}\left[\left(x^T G^T QGx + u^T Ru\right) + \left(\frac{\partial W}{\partial x}\right)^T (Ax + Bu)\right] = 0 \qquad (330)$$

Now (329) is satisfied only for that value of u obtained by solving (330). This u, when substituted in (329), yields an equation that is completely equivalent to (327). Performing the indicated operations in (330), we find

$$\frac{\partial}{\partial u}\left(u^T Ru\right) + \frac{\partial}{\partial u}\left[\left(\frac{\partial W}{\partial x}\right)^T Bu\right] = 0$$

$$2Ru + B^T \frac{\partial W}{\partial x} = 0$$

or finally,

$$u^* = -\frac{1}{2} R^{-1} B^T \left(\frac{\partial W}{\partial x}\right) \qquad (331)$$

We use the superscript ()* to indicate that this is the optimal control function (policy), which when substituted in the system (321) yields the optimal trajectory x*(t), and is the one which minimizes (or maximizes) the performance function (324).

Substituting (331) in (329) yields

$$-\frac{\partial W}{\partial t} = x^T G^T QGx + \left(\frac{\partial W}{\partial x}\right)^T Ax - \frac{1}{4}\left(\frac{\partial W}{\partial x}\right)^T BR^{-1} B^T\left(\frac{\partial W}{\partial x}\right) \tag{332}$$

This equation is equivalent to (327). We assume a solution of the form

$$W(x,t) = x^T P(t)\, x \tag{333}$$

where P(t) is a symmetric matrix, which is for the moment unknown. Then

$$\frac{\partial W}{\partial x} = 2\, P(t)\, x$$

$$\frac{\partial W}{\partial t} = x^T \dot{P}(t)\, x$$

and Eq. (332) becomes

$$-x^T \dot{P}x = x^T\left(G^T QG + 2PA - PBR^{-1}B^T P\right)x \tag{334}$$

Since the matrix $\dot{P}$ is symmetric, it is convenient to have all the terms inside the parentheses symmetric; in fact, the only one not symmetric is 2 P A, since A in general is not symmetric. It is known, however, that any square matrix may be expressed as the sum of a symmetric and a skew-symmetric matrix. In the present case,

$$2PA = \left(PA + A^T P\right) + \left(PA - A^T P\right)$$

where

$$\left(PA + A^T P\right) \equiv \text{symmetric}$$

$$\left(PA - A^T P\right) \equiv \text{skew-symmetric}$$

Now let

$$a = Ab$$

where a and b are arbitrary vectors. Then

$$a^T = b^T A^T$$

and

$$b^T\left(PA - A^T P\right)b = b^T PAb - b^T A^T Pb = b^T Pa - a^T Pb = 0$$

In other words, $(PA-A^TP)$ is identically zero. Consequently, Eq. (334) becomes

$$x^T\left[\dot{P} + PA + A^T P - PBR^{-1} B^T P + G^T QG\right]x = 0$$

This yields the nonhomogeneous <u>matrix Riccati equation</u>

$$\dot{P} + PA + A^T P - PBR^{-1} B^T P = -G^T QG \tag{335}$$

From (328) and (333), we see that

$$P(t_f) = M \tag{336}$$

The solution of Eq. (335) may be obtained in the following way.(161) Define an associated system of linear matrix equations as follows.

$$\dot{Y} = AY - BR^{-1} B^T Z \tag{337}$$

$$\dot{Z} = -G^T QGY - A^T Z \tag{338}$$

Then the solution of Eq. (335) is given by

$$P = ZY^{-1} \tag{339}$$

This is easily verified by differentiating the above expression and substituting for $\dot{Y}$ and $\dot{Z}$ from (337) and (338). Use is made of the relation

$$\dot{Y}^{-1} = -Y^{-1}\dot{Y}Y^{-1}$$

which follows by differentiating the identity

$$Y^{-1} = Y^{-1} Y Y^{-1}$$

with respect to time.

Writing Eqs. (337) and (338) in the form

$$\dot{V} = FV \tag{340}$$

where

$$V \equiv \begin{bmatrix} Z \\ Y \end{bmatrix}$$

$$F = \left[\begin{array}{c|c} -A^T & -G^T Q G \\ \hline -BR^{-1}B^T & A \end{array}\right]$$

we find

$$V(t) = e^{Ft} V(0) \tag{341}$$

If final rather than initial conditions are known, we write instead

$$V(t) = e^{F(t-t_f)} V(t_f)$$

which is equivalent to (341) if we note that

$$e^{-Ft_f} V(t_f) = V(0)$$

Consequently there is no loss of generality in dealing with V(0) as an "initial condition" matrix.

Let Eq. (341) be partitioned as follows.

$$\begin{bmatrix} Z(t) \\ Y(t) \end{bmatrix} = \left[\begin{array}{c|c} \varphi_{11}(t) & \varphi_{12}(t) \\ \hline \varphi_{21}(t) & \varphi_{22}(t) \end{array}\right] \begin{bmatrix} V_1(0) \\ V_2(0) \end{bmatrix}$$

Then, using (339), we find

$$P(t) = \left[\varphi_{11}(t)\, V_1(0) + \varphi_{12}(t)\, V_2(0)\right]\left[\varphi_{21}(t)\, V_1(0) + \varphi_{22}(t)\, V_2(0)\right]^{-1}$$

$$= \left[\varphi_{11}(t) + \varphi_{12}(t)\, V_0(0)\right]\left[\varphi_{21}(t) + \varphi_{22}(t)\, V_0(0)\right]^{-1} \tag{342}$$

where $V_0(0)$ replaces $V_2(0)V_1^{-1}(0)$. The matrix $V_0(0)$ is evaluated using the boundary condition (336).

Equation (342) is the solution of the matrix Riccati equation (335).

Having P(t), we evaluate the optimal control vector, u*(t), from (331), making use of (333); viz.,

$$u^*(t) = -R^{-1} B^T P(t)\, x(t) \tag{343}$$

An important special case arises when $M \equiv 0$ and $t_f \rightarrow \infty$ in Eq. (324). In this case, W and P do not depend on t. The matrix Riccati equation (335) then reduces to

$$PA + A^T P - PBR^{-1} B^T P = -G^T QG \tag{344}$$

which is merely an algebraic equation for the determination of P. For large n, the computational solution of (344) is not trivial. The recommended procedure is to integrate Eq. (342) with an assumed value for $V_0(0)$, and stop when P(t) reaches a steady-state value, which is then taken as the desired value of P. The speed of convergence depends, of course, on the selected value of $V_0(0)$. This, however, is not too critical, since convergence is generally very rapid.

The optimal control vector is now given by

$$u^*(t) = -K\, x(t) \tag{345}$$

$$K = R^{-1} B^T P, \text{ a constant matrix}$$

In other words, the optimal control is a linear function of the current state. Thus the optimal control is expressed as a closed-loop system rather than an open-loop type, which has characterized the results obtained in previous sections.

3.2.2 Norm-Invariant Systems

Consider the nonlinear system

$$\dot{x} = g\,(x,t) + u \tag{346}$$

where, as before, x is an n-dimensional state vector and u is an n-dimensional control vector. The latter satisfies a constraint of the form[†]

$$\|u\| \lesssim \gamma \tag{347}$$

We pose the problem of determining the control vector, u(t), subject to the above constraint, which drives the system from an arbitrary initial state to the "origin," x=0, in minimum time.

[†] See Appendix C for a discussion of norms.

We suppose that the system is norm-invariant; that is, the homogeneous form of (346)

$$\dot{x} = g\,(x,\,t) \tag{348}$$

has the property

$$\|x(t)\| = \|x(0)\| \quad \text{for all } t \geqslant 0 \tag{349}$$

In this case, Eq. (C15) shows that

$$\|\dot{x}\| = \frac{x^T\dot{x}}{\|x\|} = \frac{x^T g}{\|x\|} = 0$$

which leads to

$$x^T g\,(x,t) = 0 \tag{350}$$

The results that follow were first obtained by Athans.$^{(90)}$ His derivations, however, relied heavily on function space methods and were not in the "mainstream" of optimal control theory.

A simpler and more direct development is possible by application of the basic concepts of dynamic programming as follows.

Define

W(x) = the time required to transfer the system from state x to the origin, x=0, using an optimal policy.

Applying the principle of optimality, we obtain

$$W(x) = \underset{u}{\text{Min}} \left\{ \Delta + W\left[x + (g+u)\Delta\right]\right\} \tag{351}$$

In words, if we are in state x, then by applying a control, we incur a "cost," $\Delta \equiv dt$, and arrive in the new state, $[x + (g+u)\Delta]$. Proceeding optimally thereafter, the cost is $W\,[x + (g+u)\Delta]$ by definition. However, we select u to minimize this total cost. Hence the form of Eq. (351).

Via a Taylor expansion,

$$W\left[x + (g+u)\Delta\right] = W\,(x) + \left[\frac{\partial W(x)}{\partial x}\right]^T (g+u)\Delta + 0\,(\Delta^2)$$

Substituting this in (351), dividing through by Δ, and letting $\Delta \rightarrow 0$, we find

$$0 = \underset{u}{\text{Min}} \left[1 + \left(\frac{\partial W}{\partial x} \right)^T (g + u) \right] \tag{352}$$

Minimizing the expression in the brackets is equivalent to minimizing

$$\left(\frac{\partial W}{\partial x} \right)^T u \tag{353}$$

where u is constrained by (347). Since expression (353) is merely the inner product of two vectors, one of which is arbitrary and the other of which is constrained in magnitude, it is easy to see that (353) is minimized when

$$u = - \frac{\left(\frac{\partial W}{\partial x} \right)}{\left\| \frac{\partial W}{\partial x} \right\|} \gamma \tag{354}$$

Equation (352) is therefore equivalent to

$$0 = 1 + \left(\frac{\partial W}{\partial x} \right)^T g - \left\| \frac{\partial W}{\partial x} \right\| \gamma \tag{355}$$

The solution of this equation is

$$W(x) = \frac{1}{\gamma} \left(x^T x \right) = \frac{1}{\gamma} \|x\| \tag{356}$$

as may be verified by direct substitution; viz.,

$$\frac{\partial W}{\partial x} = \frac{1}{\gamma} \left(x^T x \right)^{-\frac{1}{2}} x$$

which, when substituted in (355), yields

$$0 = 1 + \frac{1}{\gamma} \left(x^T x \right)^{-\frac{1}{2}} x^T g - \frac{1}{\gamma} \left(x^T x \right)^{-\frac{1}{2}} \left(x^T x \right)^{\frac{1}{2}} \gamma$$

The middle term vanishes by virtue of (350). Q.E.D.

Using (356), the optimal control vector, u(t), may be written as

$$u^*(t) = -\frac{x(t)}{\|x(t)\|}\,\gamma \tag{357}$$

Thus $u^*(t)$ is obtained as a function of the current state (feedback principle), and the minimum time is obtained from Eq. (356), with $x^*(t)$ representing the optimal trajectory; i.e., the integration of Eq. (346) using the optimal control, (357).

A similar analysis shows that the control that minimizes

$$\int_0^{t_f} k\,\|u\|\,dt \qquad k = \text{positive constant} \tag{358}$$

while driving the system from an arbitrary state, x, to x = 0, is also given by (357). The terminal time, t_f, is not fixed in advance. This may be interpreted as a minimum-fuel problem.

It may also be shown, via the same techniques, that the control that minimizes

$$\int_0^{t_f} (u^T u)\,dt \tag{359}$$

while transferring the system from an arbitrary initial state, c, to x = 0, is given by

$$u^*(t) = \frac{\|c\|}{t_f} \cdot \frac{x(t)}{\|x(t)\|} \tag{360}$$

$$x(0) = c$$

$$t_f \equiv \text{prescribed}$$

The optimal trajectory, $x^*(t)$, is the solution of Eq. (346), with $u^*(t)$ given by (360). Note that now the optimal control depends explicitly on the initial state and the prescribed time, t_f, whereas previously the optimal control was a function of the current state of the system only. This last case, the minimization of the integral, (359), is a type of minimum-energy problem.

The above results are quite significant, since closed-form solutions for nonlinear optimization problems are hard to come by. More important, several cases of practical interest in aerospace control are indeed characterized by norm-invariant properties.

The latter is often a consequence of conservation of momentum. In particular, two classes of norm-invariant systems are the following.

1. All <u>linear</u> systems of the form

$$\dot{x}(t) = A(t)\, x(t) + u(t) \tag{361}$$

$$\|u(t)\| \overline{<} \gamma$$

$$A(t) = -A^T(t) \tag{362}$$

2. All <u>nonlinear</u> systems of the form

$$\dot{x}(t) = B\Big[x(t),\ t\Big] x(t) + u(t) \tag{363}$$

$$\|u(t)\| \overline{<} \gamma$$

$$B\Big[x(t),\ t\Big] = -B^T\Big[x(t),\ t\Big] \tag{364}$$

<u>Example 8</u>: For purposes of analyzing the short-period motion, a satellite may be assumed suspended in a force-free field. Let I_1, I_2, and I_3 denote the three moments of inertia of the body about the principal axes that pass through the center of mass. Also, denote by x_1, x_2, and x_3 the three angular velocities about the axes, 1, 2, and 3 respectively. The motion is then described by the Euler equations

$$I_1 \dot{x}_1 = \left(I_2 - I_3\right) x_2\, x_3 + T_1 \tag{365}$$

$$I_2 \dot{x}_2 = \left(I_3 - I_1\right) x_3\, x_1 + T_2 \tag{366}$$

$$I_3 \dot{x}_3 = \left(I_1 - I_2\right) x_1\, x_2 + T_3 \tag{367}$$

where T_1, T_2, and T_3 are the components of the torque vector, T.

If we define new variables by

$$\left.\begin{aligned} y_1 &= I_1\, x_1 \\ y_2 &= I_2\, x_2 \\ y_3 &= I_3\, x_3 \end{aligned}\right\} \tag{368}$$

then the Euler equations take the form

$$\dot{y}_1 = \left(\frac{1}{I_3} - \frac{1}{I_2}\right) y_2 \, y_3 + T_1 \tag{369}$$

$$\dot{y}_2 = \left(\frac{1}{I_1} - \frac{1}{I_3}\right) y_3 \, y_1 + T_2 \tag{370}$$

$$\dot{y}_3 = \left(\frac{1}{I_2} - \frac{1}{I_1}\right) y_1 \, y_2 + T_3 \tag{371}$$

Quantities y_1, y_2, and y_3 are the components of the angular momentum vector, y. It is easy to show that when $T_1 = T_2 = T_3 = 0$,

$$\frac{d}{dt} \| y(t) \| = 0 \tag{372}$$

In fact, in the present case, this is merely a statement of the principle of conservation of angular momentum. Now if the constraint on the torque vector is of the form

$$\| T(t) \| \overline{<} \gamma \tag{373}$$

the theory of Sec. 3.2.2 is directly applicable. Therefore, the control law

$$T^* = - \frac{y(t)}{\| y(t) \|} \gamma \tag{374}$$

will bring the system from an arbitrary angular momentum to zero angular momentum in minimum time. This corresponds to stopping the tumbling motions of the body in the shortest possible time.

Eq. (374) shows that the torque vector takes on its maximum absolute value and is directed opposite to the angular momentum vector.

3.2.3 Optimal Trajectories

A simple form of optimal launch trajectory problem may be formulated in the following way. It is assumed that the vehicle is a point mass in a uniform gravitational field; aerodynamic effects are neglected, and the flat earth approximation is used. The equations of motion then take the form (see Fig. 12)

$$\dot{x}_1 = \frac{A \, u_1 \cos u_2}{x_3} \equiv f_1 \tag{375}$$

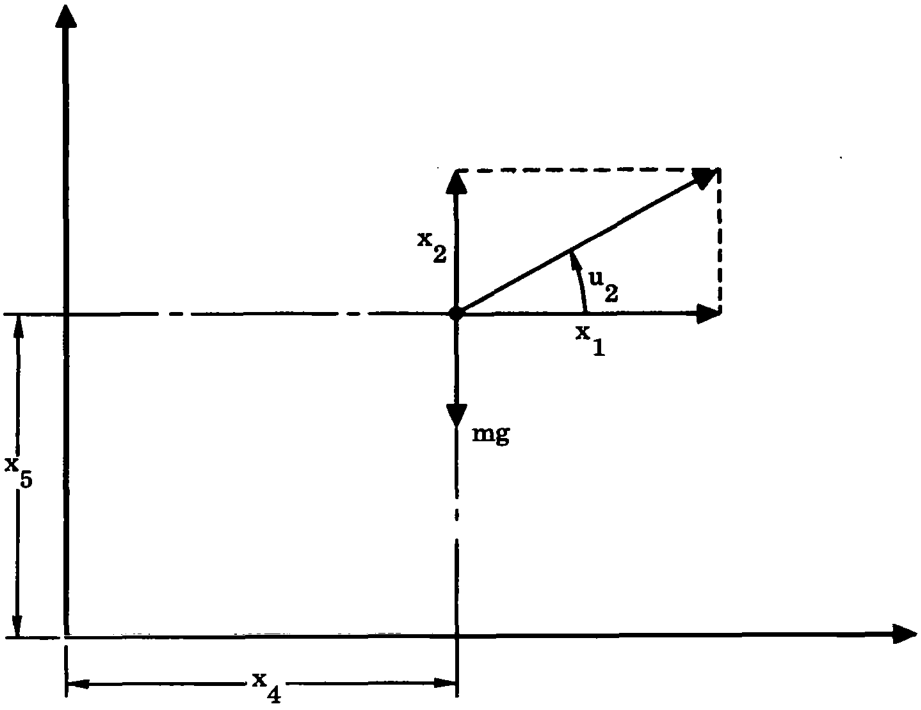


Figure 12. Coordinate System for Optimal Trajectory Problem

$$\dot{x}_2 = \frac{A\,u_1 \sin u_2}{x_3} - g \equiv f_2 \tag{376}$$

$$\dot{x}_3 = -\alpha u_1 \equiv f_3 \tag{377}$$

$$\dot{x}_4 = x_1 \equiv f_4 \tag{378}$$

$$\dot{x}_5 = x_2 \equiv f_5 \tag{379}$$

Here

$x_1(t)$ ≡ horizontal component of velocity

$x_2(t)$ ≡ vertical component of velocity

$x_4(t)$ ≡ range

$x_5(t)$ ≡ height

$x_3(t) \equiv$ nondimensional mass $= \dfrac{m(t)}{m(0)}$

$u_1(t) \equiv$ nondimensional thrust $= \dfrac{T}{T_{max}}$

$u_2(t) \equiv$ inclination of the thrust vector with respect to the horizontal

The thrust is given by

$$T = -\dot{m}\mu \qquad \mu = \text{constant} \tag{380}$$

and satisfies the constraint

$$0 \overline{<} T \overline{<} T_{max} \tag{381}$$

which means that

$$0 \overline{<} u_1(t) \overline{<} 1$$

Constants A and α are defined by

$$A = \frac{T_{max}}{m(0)}$$

$$\alpha = \frac{T_{max}}{m(0)\mu}$$

It is required to program the thrust magnitude and direction (u_1 and u_2 respectively) such that a suitable index of performance is optimized. In various guises this problem has been studied by many investigators.(18,67,70,71) Even so simplified a problem as this does not yield a closed-form solution. However, certain general features of the optimal control functions can be obtained, and the complete solution requires that one solve a set of differential equations with prescribed initial and terminal boundary conditions.

The development that follows is based on the work of Isaev(164) and includes the results of many other investigators as special cases.

We formulate the general problem as follows. Given the initial conditions

$$x_j(0) = a_j \qquad j = 1, 2, \cdots, 5 \tag{382}$$

find the control (u_1, u_2) that transfers the system described by Eqs. (375) through (379) from the given initial state to a prescribed final state such that the function

$$J = \sum_{j=1}^{5} c_j x_j (t_f) \tag{383}$$

is a minimum (maximum), where the final time, t_f, is given.

As shown below, the problems of maximum range, altitude, final velocity, and minimum fuel are special cases of this general formulation. The prescribed final state and performance function will, of course, differ in each case. Nevertheless, certain general features of the control functions are common to all.

We will seek to obtain a solution via the maximum principle. The components of the Lagrange multiplier (costate) vector are defined by Eq. (78); viz.,

$$\dot{\lambda}_j = - \sum_{k=1}^{5} \lambda_k \frac{\partial f_k}{\partial x_j} \tag{384}$$

or

$$\dot{\lambda}_1 = - \lambda_4 \tag{385}$$

$$\dot{\lambda}_2 = - \lambda_5 \tag{386}$$

$$\dot{\lambda}_3 = \frac{A u_1}{x_3^2} \left(\lambda_1 \cos u_2 + \lambda_2 \sin u_2 \right) \tag{387}$$

$$\dot{\lambda}_4 = 0 \tag{388}$$

$$\dot{\lambda}_5 = 0 \tag{389}$$

The hamiltonian is then given by

$$H = \sum_{j=1}^{5} \lambda_j f_j$$

or, in expanded form,

$$H = u_1\left[\frac{A}{x_3}\left(\lambda_1 \cos u_2 + \lambda_2 \sin u_2\right) - \alpha \lambda_3\right] - \lambda_2 g$$

$$+ \lambda_4 x_1 + \lambda_5 x_2 \tag{390}$$

Minimizing (or maximizing) H with respect to u_1 and u_2 is equivalent to minimizing (or maximizing)

$$\bar{H} = u_1\left[\frac{A}{x_3}\left(\lambda_1 \cos u_2 + \lambda_2 \sin u_2\right) - \alpha \lambda_3\right]\frac{1}{A} \tag{391}$$

An elementary calculation shows that $\bar{H}$ is maximized by

$$\begin{aligned} u_1 &= 1 \qquad \Phi_1 > 0 \\ &= 0 \qquad \Phi_1 < 0 \end{aligned} \tag{392}$$

$$u_2 = \tan^{-1}\frac{\lambda_2}{\lambda_1} \tag{393}$$

and minimized by

$$\begin{aligned} u_1 &= 1 \qquad \Phi_2 > 0 \\ &= 0 \qquad \Phi_2 < 0 \end{aligned} \tag{394}$$

$$\begin{aligned} u_2 &= \tan^{-1}\frac{\lambda_2}{\lambda_1} - \pi \\ &= - \tan^{-1}\frac{\lambda_1}{\lambda_2} - \frac{\pi}{2} \end{aligned} \tag{395}$$

where

$$\Phi_1 = \frac{1}{x_3}\sqrt{\lambda_1^{\,2} + \lambda_2^{\,2}} - \frac{\lambda_3}{\mu} \tag{396}$$

$$\Phi_2 = \frac{1}{x_3}\sqrt{\lambda_1^{\ 2} + \lambda_2^{\ 2}} + \frac{\lambda_3}{\mu} \qquad (397)$$

For definiteness, we will consider only the case of maximizing† the performance index, J, of Eq. (383). In this case, the maximum principle requires that we minimize H. Accordingly, the optimal control functions are given by Eqs. (394) and (395).

Now Eqs. (388) and (389) yield immediately

$$\lambda_4 = \lambda_{40} \qquad (398)$$

$$\lambda_5 = \lambda_{50} \qquad (399)$$

and, in turn,

$$\lambda_1 = -\lambda_{40}\,t + \lambda_{10} \qquad (400)$$

$$\lambda_2 = -\lambda_{50}\,t + \lambda_{20} \qquad (401)$$

where λ_{10}, λ_{20}, λ_{40}, and λ_{50} are constants.

It follows that the optimal $u_2(t)$ is given by

$$u_2^*(t) = \frac{\lambda_{20} - \lambda_{50}\,t}{\lambda_{10} - \lambda_{40}\,t} \qquad (402)$$

We therefore have the results that the optimal thrust inclination is a bilinear function of time and the thrust magnitude is of the bang-bang type. This is a very general feature of the optimal control, since no stipulations have as yet been imposed on the performance index, J, or the boundary conditions, $x_j(t_f)$.

If we define

$$\begin{aligned} \text{Sg}\,\Phi_2 &= 1, \qquad \Phi_2 > 0 \\ &= 0\,, \qquad \Phi_2 < 0 \end{aligned} \qquad (403)$$

† Minimizing J is equivalent to maximizing (-J).

then after substituting (394) and (395) into the set of Eqs. (375) through (379), (385) through (389), we obtain

$$\dot{x}_1^* = \frac{A\lambda_1 \, \mathrm{Sg}\,\Phi_2}{x_3\left(\lambda_1^2 + \lambda_2^2\right)^{1/2}} \tag{404}$$

$$\dot{x}_2^* = -\frac{A\lambda_2 \, \mathrm{Sg}\,\Phi_2}{x_3\left(\lambda_1^2 + \lambda_2^2\right)^{1/2}} - g \tag{405}$$

$$\dot{x}_3^* = -\alpha \, \mathrm{Sg}\,\Phi_2 \tag{406}$$

$$\dot{x}_4^* = x_1 \tag{407}$$

$$\dot{x}_5^* = x_2 \tag{408}$$

$$\dot{\lambda}_1 = -\lambda_4 \tag{409}$$

$$\dot{\lambda}_2 = -\lambda_5 \tag{410}$$

$$\dot{\lambda}_3 = -\frac{A\left(\lambda_1^2 + \lambda_2^2\right)^{1/2} \mathrm{Sg}\,\Phi_2}{x_3^2} \tag{411}$$

$$\dot{\lambda}_4 = 0 \tag{412}$$

$$\dot{\lambda}_5 = 0 \tag{413}$$

The set of Eqs. (404) through (408) describes the optimal trajectory. It is easy to show that the switching function, Φ_2, can change sign no more than twice. From Eq. (397)

$$\dot{\Phi}_2 = -\frac{\left(\lambda_4 + \lambda_5\right)}{x_3\left(\lambda_1^2 + \lambda_2^2\right)^{1/2}}$$

which is obtained by elementary differentiation and making use of (406) and (411). By virtue of (409), (410), (412), and (413), this may be written as

$$\dot{\Phi}_2 = \frac{\beta_0}{x_3\left(t^2 + \beta_1 t + \beta_2\right)^{1/2}} \tag{414}$$

where the β_i are constants. Since $x_3(t)$ is a bounded monotonic function, it follows that $\dot{\Phi}_2$ can vanish for only one value of t, say t_0. Therefore, by Rolle's theorem, Φ_2 can vanish at no more than two points.

It follows that the optimal trajectory can have no more than two powered phases.

We now consider several special cases.

1. Maximum Horizontal Velocity

Given:

$$x_j(0) = a_j \qquad j = 1, 2, \cdots, 5 \tag{415}$$

$$\begin{aligned} x_2(t_f) &= 0 \\ x_3(t_f) &= b_3 \qquad t_f \text{ given} \\ x_5(t_f) &= b_5 \end{aligned} \tag{416}$$

Maximize $J = x_1(t_f)$

Therefore $c_1 = 1$

$$c_2 = c_3 = c_4 = c_5 = 0$$

Using the methods of Sec. 3.1.2 to determine the boundary conditions for the λ_i, we find

$$\begin{aligned} \lambda_1(t_f) &= -1 \equiv \lambda_{10} \\ \lambda_4(t_f) &= 0 \equiv \lambda_{40} \end{aligned} \tag{417}$$

The solution of the problem is completely determined by integrating the 10 differential equations (404) through (413) whose boundary conditions are given by (415) through (417), with Φ_2 defined by Eq. (397).

Note that by virtue of (417), the control law, Eq. (402), reduces to

$$u_2(t) = \lambda_{50}\, t - \lambda_{20} \qquad (418)$$

2. Maximum Altitude

Given:

$$x_j(0) = a_j \qquad j = 1, 2, \cdots, 5 \qquad (419)$$

$$\begin{aligned} x_3(t_f) &= b_3 \\ x_4(t_f) &= b_4 \end{aligned} \qquad t_f \text{ given} \qquad (420)$$

Maximize $J = x_5\,(t_f)$

Therefore $c_5 = 1$

$$c_1 = c_2 = c_3 = c_4 = 0$$

and

$$\begin{aligned} \lambda_1\,(t_f) &= 0 \\ \lambda_2\,(t_f) &= 0 \\ \lambda_5\,(t_f) &= -1 \end{aligned} \qquad (421)$$

In this case, the solution to the problem is obtained by integrating the set of equations (404) through (413), with the boundary conditions given by (419) through (421).

We note also that via (421), the control law (402) becomes

$$u_2^*(t) = -\frac{1}{\lambda_{40}} \equiv \text{constant} \qquad (422)$$

3. Maximum Range

Given:

$$x_j(0) = a_j \qquad j = 1, 2, \cdots, 5 \tag{423}$$

$$\left.\begin{aligned} x_3(t_f) &= b_3 \\ x_5(t_f) &= 0 \end{aligned}\right\} \quad t_f \text{ given} \tag{424}$$

Maximize $J = x_4(t_f)$

$$c_4 = 1$$

$$c_1 = c_2 = c_3 = c_5 = 0$$

$$\begin{aligned} \lambda_1(t_f) &= 0 \equiv \lambda_{10} \\ \lambda_2(t_f) &= 0 \equiv \lambda_{20} \\ \lambda_4(t_f) &= -1 \equiv \lambda_{40} \end{aligned} \tag{425}$$

Again the solution is obtained by integrating the set of equations (404) through (413), but with the boundary conditions (423) through (425).

The conditions (425) now yield the control law

$$u_2^*(t) = -\lambda_{50} \equiv \text{constant} \tag{426}$$

4. Minimum Fuel

Given:

$$x_j(0) = a_j \qquad j = 1, 2, \cdots, 5 \tag{427}$$

$$\left.\begin{aligned} x_4(t_f) &= b_4 \\ x_5(t_f) &= 0 \end{aligned}\right\} \quad t_f \text{ given} \tag{428}$$

Maximize $J = -x_3(t_f)$

$$
\begin{aligned}
c_3 &= -1 \\
c_1 = c_2 = c_4 = c_5 &= 0 \\
\lambda_1 (t_f) &= 0 \equiv \lambda_{10} \\
\lambda_2 (t_f) &= 0 \equiv \lambda_{20} \\
\lambda_3 (t_f) &= 1 \equiv \lambda_{30}
\end{aligned}
\tag{429}
$$

As before, the problem is solved by integrating (404) through (413), but for this case using the boundary conditions (427) through (429).

Using conditions (429), the optimal thrust inclination is again a constant; viz.,

$$
u_2^* (t) = \frac{\lambda_{50}}{\lambda_{40}} \equiv \text{constant} \tag{430}
$$

Remark: It has been shown that the optimal control for this problem is characterized by the following properties.

1. The thrust magnitude is either zero or maximum (bang-bang).
2. The thrust inclination is, in the general case, a bilinear function of time. For specialized cases, it may be a constant or a linear function of time.
3. The optimal trajectory contains no more than two powered phases.

In order to obtain the complete solution (exact form of the optimal trajectory), it is necessary to solve a set of nonlinear differential equations with two-point boundary conditions. This is not a trivial computational task. Some promising techniques for doing this efficiently have been developed in recent years; a description of these is contained in Appendix A.

It should also be noted that in the above analysis, there was the implicit assumption that a solution exists. While this may be true in most cases of practical interest, it must be recognized that not all combinations of conditions give a meaningful or well-defined problem. Generally, however, if there exists a control that satisfies the given boundary conditions, one may proceed to calculate an optimal control with the assurance that this is physically and mathematically meaningful.

There is one further mathematical feature of the solution: the control laws derived represent only necessary conditions for the optimum. They may not be sufficient. Sufficiency proofs are at present available only for linear systems. Again, in most cases of practical interest, this is a mathematical sophistry that does not seriously compromise the solution.

3.3 AEROSPACE APPLICATIONS

The literature on application of optimal control theory to guidance and control of aerospace vehicles is very extensive. Even the large list of references in this monograph is far from complete. One of the best survey papers on this subject is the one by Paiewonsky,(19) which reviews the highlights and historical development of the theory and gives a critical evaluation of basic contributions. Other basic sources are the books by Lawden,(2) Chang,(1) Leitmann,(3), and Merriam(146) and the papers by Miehle,(18) Lawden,(70) and Ho.(135)

The choice of examples to illustrate "typical" applications is largely subjective. Among the factors that may guide such a selection are theoretical novelty, practical importance, or mathematical elegance. Since the primary aim for the moment is pedagogic exposition, the examples that follow were selected for their didactic value and practical interest. The extensive detailed treatment that characterizes practical design problems of high order was avoided, since the details tend to obscure the essence.

3.3.1 Lunar Soft Landing(52)

A problem of some practical interest is the following. Assume that a lunar vehicle has arrived in the vicinity of the moon's surface. The motion of the vehicle is vertically downward, and the only forces acting are the thrust and lunar gravity (assumed constant). The thrust magnitude may be throttled between zero and some prescribed maximum. It is desired to calculate the thrust program that will result in a soft touchdown (altitude and altitude rate simultaneously zero) while minimizing the fuel expenditure.

With these assumptions, the motion is governed by

$$\ddot{h} = -\frac{c\,\dot{m}}{m} - g \qquad (431)$$

where

$h \equiv$ altitude of vehicle above lunar surface

$m \equiv$ instantaneous mass

$g \equiv$ lunar gravity acceleration

The thrust is given by $-c\dot{m}$, where c is a positive constant and $\dot{m} \leq 0$. We also have the boundary values

$$\begin{aligned} h\,(0) &= x_{10} \\ \dot{h}\,(0) &= x_{20} \end{aligned} \tag{432}$$

$$\begin{aligned} h\,(t_f) &= 0 \\ \dot{h}\,(t_f) &= 0 \end{aligned} \tag{433}$$

and the thrust magnitude constraint

$$-\alpha \leq \dot{m}\,(t) \leq 0 \tag{434}$$

where α is a positive constant.

Since the vehicle is assumed to be in the terminal descent phase, physical considerations indicate that $x_{10} > 0$ and $x_{20} < 0$.

Minimizing the fuel expenditure is equivalent to minimizing the performance index

$$J = m\,(0) - m\,(t_f) \tag{435}$$

This completes the mathematical statement of the problem. Before proceeding with a formal solution, certain simplifications may be effected in the following way. We have

$$\frac{\dot{m}}{m} = \frac{d}{dt}\,(\ell n\, m)$$

where $\ell n()$ denotes natural log of.

Therefore Eq. (431) may be written as

$$\ddot{h} = -c\,\frac{d}{dt}\,(\ell n\, m) - g \tag{436}$$

Integrating between the limits of 0 and t_f,

$$\dot{h}\,(t) = -c\,\ell n\,\frac{m\,(t)}{m\,(0)} - g\,t + \dot{h}\,(0)$$

Now $\dot{h}(t_f) = 0$ if, and only if,

$$c \, \ell n \frac{m\,(t_f)}{m\,(0)} = \dot{h}\,(0) - g t_f$$

Solving for $m\,(t_f)$,

$$m\,(t_f) = m\,(0) \exp\left[\frac{\dot{h}\,(0) - g\,t_f}{c}\right] \tag{437}$$

Substituting this in (435),

$$J = m\,(0)\left\{1 - \exp\left[\frac{\dot{h}\,(0) - g\,t_f}{c}\right]\right\}$$

It is apparent, therefore, that minimizing the fuel expenditure is completely equivalent to minimizing the total time, t_f. In other words, we reduce to a minimum-time problem.

We introduce the definitions

$$h = x_1$$

$$\dot{x}_1 = x_2$$

$$x_3 = m$$

$$u = \dot{m}$$

Eq. (431) is then represented by the system

$$\dot{x}_1 = x_2 \equiv f_1 \tag{438}$$

$$\dot{x}_2 = -\frac{c}{x_3} u - g \equiv f_2 \tag{439}$$

$$\dot{x}_3 = u \equiv f_3 \tag{440}$$

with the control constraint given by

$$-\alpha \lesssim u(t) \lesssim 0 \tag{441}$$

and boundary conditions

$$\begin{aligned} x_1(0) &= x_{10} & x_1(t_f) &= 0 \\ x_2(0) &= x_{20} & x_2(t_f) &= 0 \\ x_3(0) &= x_{30} & x_3(t_f) &\equiv \text{free} \end{aligned} \tag{442}$$

In accordance with the discussion of Sec. 3.2.1, the minimum time problem

$$\min \int_0^{t_f} dt$$

is converted to a problem of minimizing $x_4(t_f)$ where

$$x_4 = \int_0^t dt$$

$$\dot{x}_4 = 1 \tag{443}$$

$$x_4(0) = 0 \tag{444}$$

Proceeding via the maximum principle, we find, for the hamiltonian,

$$H = \lambda_1 x_2 - \lambda_2 \left(\frac{cu}{x_3} - g \right) + \lambda_3 u + \lambda_4 \tag{445}$$

where the Lagrange multipliers satisfy

$$\dot{\lambda}_1 = 0 \tag{446}$$

$$\dot{\lambda}_2 = -\lambda_1 \tag{447}$$

$$\dot{\lambda}_3 = - \frac{\lambda_2 cu}{x_3^2} \tag{448}$$

$$\dot{\lambda}_4 = 0 \tag{449}$$

In order to minimize $x_4\ (t_f)$, we must maximize H. The optimal control is therefore given by†

$$u^*(t) = -\alpha \qquad \text{when } \Phi < 0$$
$$= 0 \qquad \text{when } \Phi > 0 \tag{450}$$

$$\Phi = \lambda_3 - \frac{c\lambda_2}{x_3} \tag{451}$$

In Ref. 52, a rather elaborate analysis is used to show that Φ cannot change sign more than once in the interval $0 \lesssim t \lesssim t_f$. It is possible, in fact, to derive this result very easily as follows.

Differentiating (451) and using (440), (447), and (448), we find

$$\dot{\Phi} = \frac{c\lambda_1}{x_3} \tag{452}$$

But Eq. (446) shows that λ_1 = constant, while $x_3 = m(t)$ is a positive monotonic function of t. Therefore $\dot{\Phi}$ cannot change sign in the interval, $0 \lesssim t \lesssim t_f$, which, in turn, means that Φ can change sign no more than once in this interval. Consequently, once the full thrust is switched on, it remains on until touchdown. We now seek to determine the relation between $x_1(t)$ and $x_2(t)$, at which switching occurs. Let the interval over the thrusting phase be denoted by $(0,t_T)$. Denote by x'_{10}, x'_{20}, and x'_{30} the altitude, altitude rate, and mass at the instant of switching to full thrust. Then from Eq. (439),

$$x_2\ (t) = -c\ \ln\left(1 - \frac{\alpha}{x'_{30}}\, t\right) - g\, t + x'_{20} \tag{453}$$

$$0 \lesssim t \lesssim t_T$$

Using Eq. (438), we obtain

$$x_1\ (t) = \int_0^t x_2\ (\tau)\, d\tau + x'_{10} = \frac{c x'_{30}}{\alpha}\left(1 - \frac{\alpha}{x'_{30}}\, t\right) \ln\left(1 - \frac{\alpha}{x'_{30}}\, t\right)$$

$$+ ct - \frac{1}{2}\, g t^2 + x'_{20}\, t + x'_{10} \tag{454}$$

$$0 \lesssim t \lesssim t_T$$

†A somewhat lengthy analysis shows that the singularity condition $\Phi = 0$ is incompatible with the physical constraints of the problem. See Ref. 52.

For a soft landing, we must have

$$x_1\,(t_T) = x_2\,(t_T) = 0$$

Therefore Eqs. (453) and (454) become

$$0 = -\,c\,\ln\left(1 - \frac{\alpha}{x'_{30}}\,t_T\right) - g\,t_T + x'_{20} \tag{455}$$

$$0 = \frac{c\,x'_{30}}{\alpha}\left(1 - \frac{\alpha}{x'_{30}}\,t_T\right)\ln\left(1 - \frac{\alpha}{x'_{30}}\,t_T\right) + c\,t_T - \frac{1}{2}\,g\,t_T^2 + x'_{20}\,t_T + x'_{10} \tag{456}$$

These may be solved for x'_{10} and x'_{20} as follows.

$$x'_{20} = c\,\ln\left(1 - \frac{\alpha}{x'_{30}}\,t_T\right) + g\,t_T \tag{457}$$

$$x'_{10} = -\,\frac{c\,x'_{30}}{\alpha}\,\ln\left(1 - \frac{\alpha}{x'_{30}}\,t_T\right) - c\,t_T - \frac{1}{2}\,g\,t_T^2 \tag{458}$$

If we eliminate t_T between these two equations we obtain a relation of the form $F(x'_{10}, x'_{20}) = 0$, which determines the altitude and altitude rate at which one switches to full thrust. A fairly simple form for F is obtained by using the approximation

$$\ln\left(1 - \frac{\alpha}{x'_{30}}\,t_F\right) \approx -\,\frac{\alpha\,t_T}{x'_{30}} - \frac{1}{2}\left(\frac{\alpha\,t_T}{x'_{30}}\right)^2$$

which for

$$\frac{m\,(t_T)}{x'_{30}}\ \overline{>}\ 0.75 \tag{459}$$

is accurate to within 2 percent.

In this case, Eqs. (457) and (458) become

$$x'_{10} = a\,t_T^2 \tag{460}$$

$$x'_{20} = -\,2\,a\,t_T - b\,t_T^2 \tag{461}$$

where

$$a = \frac{1}{2}\left[\frac{c\alpha - gx'_{30}}{x'_{30}}\right]$$

$$b = \frac{c}{2}\left(\frac{\alpha}{x'_{30}}\right)^2$$

Since the ratio of maximum thrust to initial mass must be greater than g, we see that $a > 0$. Also, since x'_{10} is positive, it is apparent that the only value of t_T that is physically meaningful is

$$t_T = \sqrt{\frac{x'_{10}}{a}} \qquad (462)$$

Substituting this into (460) yields

$$F\left(x'_{10}, x'_{20}\right) = \frac{b}{a}x'_{10} + x'_{20} + 2\,a\sqrt{\frac{x'_{10}}{a}} \qquad (463)$$

Physical considerations dictate that $x'_{10} > 0$ and $x'_{20} < 0$, since altitude is positive and the vehicle velocity is in the negative x_1 direction. Furthermore, the inequality (459), together with the relation

$$x_3\,(t) = m\,(t) = x'_{30} - \alpha\, t \qquad 0 \le t \le t_T$$

leads to

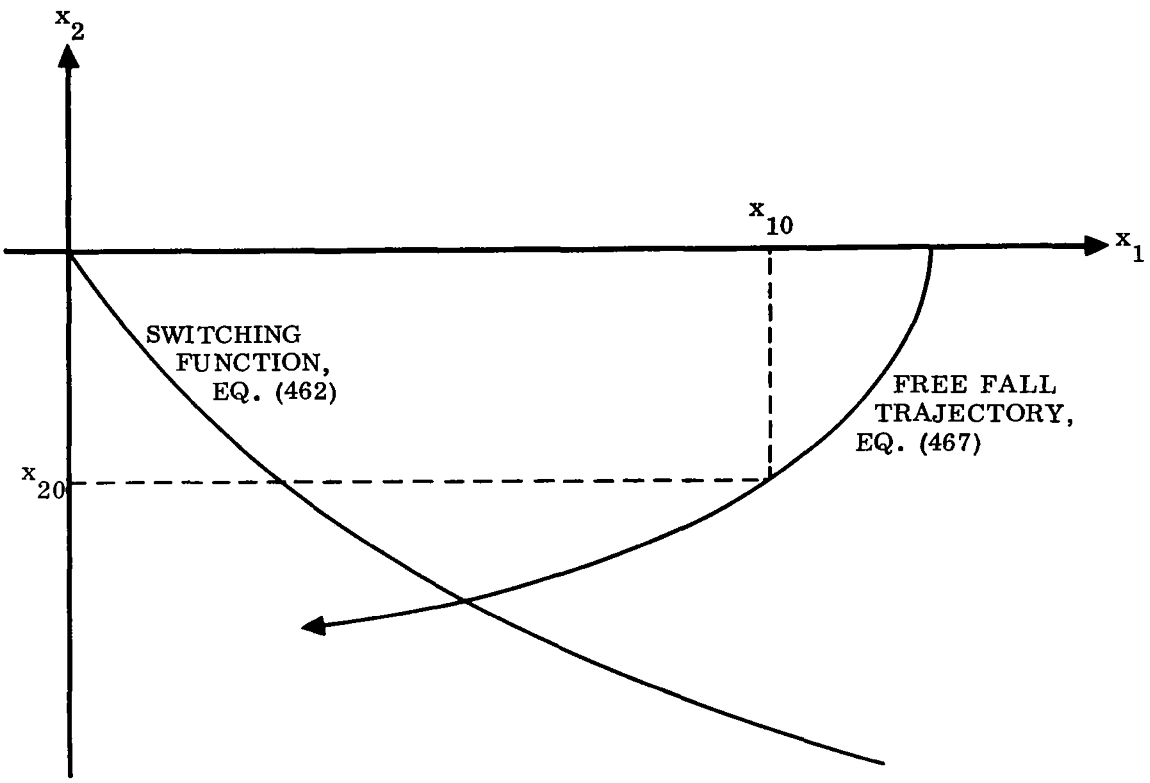
$$0 \le t_T \le \frac{x'_{30}}{4\,\alpha} \qquad (464)$$

Using $t_T\Big]_{max} = \frac{x'_{30}}{4\alpha}$, we find that the range of values of interest in the $x_1 - x_2$ plane is given by

$$0 \le x_1 \le \frac{a}{16}\left(\frac{x'_{30}}{\alpha}\right)^2 \qquad (465)$$

$$-\frac{a}{2}\left(\frac{x'_{30}}{\alpha}\right)^2 - \frac{b}{16}\left(\frac{x'_{30}}{\alpha}\right)^2 \le x_2 \le 0 \qquad (466)$$

A plot of the switching function, (463), for the range of values given by (465) and (466) is shown in Fig. 13.



x_2
x_{10}
x_1
SWITCHING FUNCTION, EQ. (462)
FREE FALL TRAJECTORY, EQ. (467)
x_{20}

Figure 13. Plot of Switching Function and Free-Fall Trajectory

Assume now that the vehicle is at some initial altitude, x_{10}, and with an initial altitude rate, x_{20}. The free-fall trajectory is readily obtained from

$$x_1 = x_{10} - \frac{1}{2g}\left[x_2^2 - x_{20}^2\right] \tag{467}$$

This is also plotted in Fig. 13.

The form of the optimal trajectory is now apparent. If the initial conditions are such that $F(x_{10},x_{20}) > 0$ — i.e., the point (x_{10},x_{20}) lies above the switching curve as shown in Fig. 13 — the vehicle is allowed to continue in free fall until $F(x_1,x_2) = 0$. At this point, full thrust is switched on and remains on until touchdown.

3.3.2 Optimal Control of Booster Vehicles

The fact that a complete solution is available for the problem of optimal control of a linear system with a quadratic performance index (Sec. 3.2.1) has motivated the effort to relate pragmatic design requirements to this format. As a matter of fact, a far from trivial problem is that of formulating a meaningful optimality criterion for a launch vehicle autopilot. One usually requires that the control system stabilize the unstable airframe. In addition, it is desired that the deviations from desired attitude be small, and that the angle of attack be small in order to reduce bending loads, engine angle deflections, etc. With the exception of stability, none of these automatically results from the quadratic performance criterion.

Several recent studies have attempted to relate meaningful performance requirements in booster autopilot systems to the quadratic performance criterion. We shall consider in detail the papers by Tyler and Tuteur,(13) Fisher,(14) Bailey, (122) and Tyler.(136) Rather than discuss these individually, we will take a unified point of view. The results of particular studies will be shown to be special cases of a general approach. Among the problems investigated are the following.

1. Stabilizing a flexible booster using multiple sensors.(14)
2. A model-following system.(136)
3. Minimizing deviations from desired states.(122)
4. Relating the characteristic equation of the optimal system to the weighting matrix elements in the performance index.(13)

These will be referred to as problems 1 through 4. As a preliminary, the results of Sec. 3.2.1 will be expressed in a variety of alternative forms to facilitate treatment of individual problems.

In the notation of Sec. 3.2.1, the system considered takes the form

$$\dot{x} = Ax + Bu \tag{468}$$

$$y = Gx \tag{469}$$

which are merely Eqs. (321) and (322) repeated here for convenience. For the performance index, we take the following version of Eq. (324)

$$J = \lim_{t_f \to \infty} \int_0^{t_f} \left(x^T G^T Q G x + u^T R u \right) dt \tag{470}$$

In this case, the optimal control is given by Eq. (345).

$$u^*(t) = -K x(t) \tag{471}$$

$$K = R^{-1} B^T P$$

where R, B, and P are constant matrices, the latter of which is obtained as the solution of the steady-state Riccati equation

$$PA + A^T P - PBR^{-1} B^T P = -G^T Q G \tag{472}$$

This is the simplest case; the optimal control (471) is merely a constant matrix times the current state of the system (feedback principle). If we take for the performance criterion

$$J = \int_0^{t_f} \left(x^T G^T QGx + u^T Ru\right) dt \tag{473}$$

then the form of the optimal control is still given by (471) except that P, and therefore K, are now time-varying, with P obtained as the solution of

$$-\dot{P} = PA + A^T P - PBR^{-1}B^T P + G^T QG \tag{474}$$

The above is merely a summary of the results obtained in Sec. 3.2.1.

Alternative forms of this solution have been obtained by Kalman(93) and Merriam(146) and have been used in the studies by Tyler,(136) Bailey,(14) and Tyler and Tuteur,(13) which will be discussed here. It will clarify the discussion to relate these results to the ones derived in Sec. 3.2.1.

To obtain Kalman's equations, we first postmultiply Eq. (474) by x, obtaining

$$-\dot{P}x = PAx + A^T Px - PBR^{-1}B^T Px + G^T QGx \tag{475}$$

The costate (Lagrange multiplier) vector is obtained from the optimal return function, W(x,t), via Eq. (309); viz.,

$$\lambda = -\frac{\partial W}{\partial x}$$

which in the present case may be expressed as

$$\lambda = -2\,Px$$

using Eq. (333). However Kalman's performance index contains the factor 1/2 in front of the integral of Eq. (470), which means that his optimal return function is half of that defined by Eq. (325). Furthermore, to obtain the minimum of the performance function, he minimizes the hamiltonian defined by Eq. (310), whereas the conventional (maximum principle) approach requires the maximization of the hamiltonian. This means that instead of λ, as defined above, we must take

$$\lambda = \frac{1}{2}\frac{\partial W}{\partial x} = Px \tag{476}$$

to conform with Kalman's assumptions. Noting that

$$\dot{\lambda} = \dot{P}x + P\dot{x} \tag{477}$$

we obtain, after substituting (476) and (477) in (475),

$$\dot{\lambda} = -A^T\lambda - G^TQGx \tag{478}$$

where we have used (468) and (471). Furthermore, Eqs. (468) and (471) show that the optimal trajectory is described by

$$\dot{x} = Ax - BR^{-1}B^T\lambda \tag{479}$$

in terms of the costate vector, (476). The last two relations may be written as

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -G^TQG & -A^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} \tag{480}$$

It is apparent that the costate vector, λ, is the adjoint of x. Consequently, the eigenvalues (closed-loop roots) in the characteristic equation

$$\begin{vmatrix} (Is - A) & BR^{-1}B^T \\ G^TQG & (Is + A) \end{vmatrix} = 0 \tag{481}$$

consist of the eigenvalues of the optimal system and their mirror images about the imaginary axis in the s plane, which belong to the adjoint system.

We note further that the optimal trajectory is also described by

$$\dot{x}^* = (A - BK)x \tag{482}$$

using (468) and (471). The characteristic equation of (482), together with its adjoint, is therefore given by

$$|Is - (A - BK)| \cdot |Is - (A + BK)| = 0 \tag{483}$$

Thus the roots of (481) and (483) must be identical. This fact will be used later to establish a relationship between the eigenvalues of the optimal system and its feedback gains.

We turn now to a discussion of Merriam's method. The performance criterion is taken as

$$J = \int_0^{t_f} \left[\left(y_D - y\right)^T Q \left(y_D - y\right) + \left(u_D - u\right)^T R \left(u_D - u\right)\right] dt \qquad (484)$$

Here y is the output vector given by Eq. (469), and y_D is a reference or desired output. Similarly, u_D is the desired or reference control, and u is the actual system control vector. The matrices Q and R are assumed to be diagonal matrices that may be time-varying.

The optimal return function is defined in a manner similar to Eq. (325) as follows.

$$W(x, t) = \underset{u}{\text{Min}} \int_t^{t_f} \left[\left(y_D - y\right)^T Q \left(y_D - y\right) + \left(u_D - u\right)^T R \left(u_D - u\right)\right] dt \qquad (485)$$

Proceeding as in Sec. 3.2.1, we find

$$-\frac{\partial W}{\partial t} = \underset{u}{\text{Min}} \left\{ \left(y_D - y\right)^T Q \left(y_D - y\right) + \left(u_D - u\right)^T R \left(u_D - u\right) + \left(\frac{\partial W}{\partial x}\right)^T (Ax + Bu) \right\} \qquad (486)$$

The optimal control is found to be

$$u^* = u_D - \frac{1}{2} R^{-1} B^T \left(\frac{\partial W}{\partial x}\right) \qquad (487)$$

Assuming a solution of the form

$$W(x, t) = p(t) - 2 \sum_{\alpha=1}^{n} p_\alpha (t)\, x_\alpha (t) + \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} p_{\beta\gamma} (t)\, x_\beta(t)\, x_\gamma(t) \qquad (488)$$

we find

$$\frac{\partial W}{\partial t} = \dot{p} - 2 \sum_{\alpha=1}^{n} \dot{p}_\alpha x_\alpha + \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} \dot{p}_{\beta\gamma}\, x_\beta\, x_\gamma \qquad (489)$$

$$\frac{\partial W}{\partial x_\alpha} = -2\, p_\alpha + 2 \sum_{\beta=1}^{n} p_{\beta\alpha}\, x_\beta \tag{490}$$

Substituting (487) through (490) in (486) and assuming that $p_{\beta\gamma} = p_{\gamma\beta}$ leads to

$$(\dot{p} + F) - 2 \sum_{\alpha=1}^{n} \left[\dot{p}_\alpha + F_\alpha\right] x_\alpha + \sum_{\alpha=1}^{n} \left[\dot{p}_{\alpha\alpha} + F_{\alpha\alpha}\right] x_\alpha^2$$

$$+ 2 \sum_{\beta=1}^{n} \sum_{\gamma=1}^{n} \left[\dot{p}_{\beta\gamma} + \frac{1}{2} F_{\beta\gamma} + \frac{1}{2} F_{\gamma\beta}\right] x_\beta\, x_\gamma = 0 \tag{491}$$

The quantities F, F_α, and $F_{\beta\gamma}$ depend only on the components of the matrices A, B, Q, R, and G. In order for Eq. (491) to be satisfied, each of the coefficients on the left-hand side of this equation must vanish independently. This leads to

$$-\dot{p} = F \tag{492}$$

$$-\dot{p}_\alpha = F_\alpha \tag{493}$$

$$-\dot{p}_{\beta\gamma} = \frac{1}{2}\left(F_{\beta\gamma} + F_{\gamma\beta}\right) \tag{494}$$

$$\alpha, \beta, \gamma, = 1, 2, \cdots, n$$

$$-\dot{p}_{\alpha\alpha} = F_{\alpha\alpha}$$

Eqs. (492) through (494) are a set of [1/2 n (n-1) + 2n + 1] differential equations for determining the various p's. Noting that

$$W(x, t_f) = 0$$

via (485), we find, for the initial conditions,

$$p(t_f) = p_\alpha(t_f) = p_{\beta\gamma}(t_f) = 0 \tag{495}$$

Having calculated the p's, the optimal control is obtained directly from Eq. (487), making use of (490).

We now consider in detail the four problems stated at the beginning of this section.

Problem 1

In order to investigate the possibility of stabilizing a flexible launch vehicle using the theory developed here, we will consider the system whose dynamics is expressed by

$$m U_0 (\dot{\alpha} - \dot{\theta}) = T_c \delta - L_\alpha \alpha - (m g \cos \theta_0) \theta \tag{496}$$

$$\ddot{\theta} = \mu_c \delta + \mu_\alpha \alpha \tag{497}$$

$$\ddot{q}_1 + 2 \xi_1 \omega_1 \dot{q}_1 + \omega_1^2 q_1 = - \frac{T_c}{M_1} \delta \tag{498}$$

The symbols have the following meaning.[†]

g = gravity acceleration

I_y = moment of inertia of vehicle about pitch axis

ℓ_c = distance from mass center of vehicle to engine swivel point

ℓ_α = distance from mass center of vehicle to engine swivel point

L_α = aerodynamic load per unit angle of attack

m = mass of vehicle

M_1 = generalized mass of first bending mode

q_1 = generalized displacement of first bending mode

T_c = control thrust

U_0 = forward velocity of vehicle

α = angle of attack

δ = thrust angle deflection

θ = attitude angle

θ_0 = steady-state attitude angle

$\mu_c = \dfrac{T_c \ell_c}{I_y}$

[†] See Ref. 167 for a more complete account of this problem.

$$\mu_\alpha = \frac{L_\alpha \ell_\alpha}{I_y}$$

ξ_1 = relative damping factor for first bending mode

ω_1 = undamped natural frequency of first bending mode

Conventional instrumentation whose output is a linear combination of the state variables is described as

Rate Gyro:

$$\theta_{RG} = K_R\left(\dot{\theta} + \sigma_G^{(1)} \dot{q}_1\right) \tag{499}$$

Position Gyro:

$$\theta_{PG} = \left(\theta + \sigma_G^{(1)} q_1\right) K_P \tag{500}$$

Accelerometer:

$$\theta_A = K_a\left[\frac{T_c \delta - L_\alpha \alpha}{m} + \alpha_T\left(\theta - \sigma_A^{(1)} q_1\right)\right] \tag{501}$$

Angle of Attack Sensor:

$$\theta_\alpha = K_\alpha\left(\alpha - \sigma_\alpha^{(1)} q_1\right) \tag{502}$$

Here, K_R, K_a, K_α, and K_p are the gains associated with the respective sensors; α_T is the thrust acceleration, and the $\sigma^{(1)}$ quantities are the normalized bending mode slopes whose subscripts indicate the location of the sensor along the vehicle.

If we define

$$\begin{aligned} x_1 &= \theta & x_4 &= q_1 \\ x_2 &= \dot{\theta} & x_5 &= \dot{q}_1 \\ x_3 &= \alpha & u &= \delta \end{aligned} \tag{503}$$

then the system described by Eqs. (496) through (498) may be expressed as

$$\dot{x} = Ax + Bu$$

with

$$A = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & \mu_\alpha & 0 & 0 \\ g \cos \theta_0/U_0 & 1 & -L_\alpha/mU_0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & -\omega_1^{\,2} & -\xi_1\,\omega_1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 \\ \mu_c \\ T_c/m\,U_0 \\ 0 \\ -T_c/M_1 \end{bmatrix}$$

This is equivalent to the system (468) and (469), with G = I, the unit matrix.

Taking a performance index of the form (470), the optimal control is given by Eq. (471); viz.,

$$u^*(t) = -K\,x\,(t) \qquad (504)$$

Note that in the present case, R is a scalar, since u is a scalar and K in the above expression is a 1xn matrix (i.e., a row vector). The control is optimal in the sense that the performance index (470) is minimized; this critèrion has as yet not been directly related to control system requirements, such as response speed, overshoot, and damping. These may indeed be determined once K is calculated, since the closed-loop poles of the system are then directly obtainable.

Obviously, K is a function of Q and R, and therefore the properties of the "optimal system" are directly (though not simply) related to the choice of Q and R. In Fisher's paper,(14) R is taken as unity, and an iterative process that will lead to a system of acceptable dynamic qualties is performed on Q (starting with some arbitrary choice). Basically, one guesses Q, determines K, and then calculates the closed-loop poles of the resulting system. If this does not yield an acceptable system from the point of view of dynamic response properties, the process is repeated with a new choice of Q, etc. The method is completely cut-and-try. Fisher gives no technique whereby successive iteration converges to some desired form. Previous experience no doubt is a valuable guide in obtaining meaningful results.

A basic property of the optimal control (504) is that there is one feedback loop for each control variable. It is known, however, that acceptable systems may be designed with only one or two feedback loops. Of course, the more feedback loops employed, the greater the capability to achieve specified performance. It is instructive, therefore, to investigate the means whereby an n^{th}-order system may approximate the optimal control (504) with fewer than n feedback loops. Let us assume that m sensors are used and that the output of each sensor is given by

$$z_i = \sum_{j=1}^{n} b_{ij} x_j \qquad i = 1, 2, \cdots, m \tag{505}$$

We note, for example, that two rate gyros may be viewed as two different sensors if their K_R and $\alpha_G^{(1)}$ are different. Thus for the system (496) through (498), the use of one position gyro and two rate gyros and two accelerometers (having different gains) will provide five independent measurements that are linear combinations of all the state variables.

If now we use

$$u' = \sum_{i=1}^{m} c_i z_i \tag{506}$$

to denote the control function employing the sensors (the c_i are as yet undetermined constants), then (504) and (506) will be equivalent if

$$-\sum_{i=1}^{n} k_i x_i = \sum_{i=1}^{m} c_i z_i = \sum_{i=1}^{m} c_i \sum_{j=1}^{n} b_{ij} x_j \tag{507}$$

This leads to

$$b^T c = - K^T \tag{508}$$

$b \equiv m \times n$ matrix

$c \equiv m \times 1$ matrix (i.e., m vector)

$K \equiv 1 \times n$ matrix (i.e., n-dimensional row vector)

If m = n, then the components of c are uniquely determined from Eq. (508). If m < n, then there is no unique solution for c. However, for an approximate solution, we may seek the value of c that minimizes the error criterion

$$E = \|b^T c + K^T\|^2 = \left(b^T c + K^T\right)^T \left(b^T c + K^T\right) \tag{509}$$

This is easily found to be

$$\frac{\partial E}{\partial c} = 2\left(bb^T c + bK^T\right) = 0$$

or

$$bb^T c = -bK^T \tag{510}$$

The matrix equation (510) is a system of m linear equations for the determination of the m components of c.

It should be noted that the approximate solution will not be optimal and that stability is not guaranteed. Thus each particular solution must be investigated individually to determine if performance is acceptable.

Problem 2

The given system is again described by

$$\dot{x} = Ax + Bu \tag{511}$$

$$y = Gx \tag{512}$$

It is desired that the system behave as the "model"

$$\dot{\eta} = L\eta \tag{513}$$

In order to achieve this, we formulate the performance criterion as

$$J = \frac{1}{2} \lim_{t_f \to \infty} \int_0^{t_f} \left[\left(\dot{y} - Ly\right)^T Q \left(\dot{y} - Ly\right) + u^T R u\right] dt \tag{514}$$

Presumably, an appropriate selection of the weighting matrices, Q and R, will yield the desired result. An analysis completely identical with that of Sec. 3.2.1 shows that in the present case, the optimal control has the form

$$u(t) = -\hat{R}^{-1} B^T \left[P + G^T Q (GA - LG)\right] x \tag{515}$$

where P is the solution of the steady-state form of the matrix Riccati equation

$$P\hat{A} + \hat{A}^T P + \hat{G}^T \hat{Q}\hat{G} - PB\hat{R}^{-1} B^T P = 0 \tag{516}$$

and

$$\hat{R} = B^T G^T QGB + R \tag{517}$$

$$\hat{A} = A - B\hat{R}^{-1} B^T G^T Q\,(GA - LG) \tag{518}$$

$$\hat{Q} = Q - QGB\hat{R}^{-1} B^T G^T Q \tag{519}$$

$$\hat{G} = GA - LG \tag{520}$$

It is convenient to write Eq. (515) in the form

$$u\ (t) = -\left(K_r + K_m\right)x = -Kx \tag{521}$$

where

$$K_r = \hat{R}^{-1} B^T P \tag{522}$$

$$K_m = \hat{R}^{-1} B^T G^T Q\,(GA - LG) \tag{523}$$

The control is thus expressed as a sum of two terms, one of which depends on the solution of the Riccati equation, and the other of which depends on the properties of the model.

We now seek to obtain some insight into this result. More specifically, what must be the stipulations on Q and R in order to have the given system exhibit the dynamic properties of the model? For this purpose, we consider the second-order system described by

$$A = \begin{bmatrix} 0 & 1 \\ -A_{21} & -A_{22} \end{bmatrix} \qquad B = \begin{bmatrix} 0 \\ B_{21} \end{bmatrix}$$

$$G = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

with weighting matrices

$$Q = \begin{bmatrix} Q_{11} & 0 \\ 0 & Q_{22} \end{bmatrix} \qquad R = R_{11}, \text{ a scalar}$$

and the matrix describing the model

$$L = \begin{bmatrix} 0 & 1 \\ -L_{21} & -L_{22} \end{bmatrix}$$

We then find

$$\hat{R}_{11} = Q_{22} B_{21}^{2} + R_{11}$$

$$\hat{A}_{11} = 0 \qquad \hat{A}_{12} = 1$$

$$\hat{A}_{21} = - A_{21} - \frac{B_{21}^{2} Q_{22} (L_{21} - A_{21})}{B_{21}^{2} Q_{22} + R_{11}}$$

$$\hat{A}_{22} = - A_{21} - \frac{B_{21}^{2} Q_{22} (L_{22} - A_{22})}{B_{21}^{2} Q_{22} + R_{11}}$$

$$\hat{Q}_{22} = Q_{22} - \frac{B_{21}^{2} Q_{22}^{2}}{B_{21}^{2} Q_{22} + R_{11}}$$

$$\hat{Q}_{11} = Q_{11}$$

$$K_{m} = \frac{B_{21} Q_{22}}{B_{21}^{2} Q_{22} + R_{11}} \left[\left(L_{21} - A_{21}\right) \left(L_{22} - A_{22}\right) \right]$$

The determination of K_r depends on the components of the P matrix, which are found from Eq. (516) as follows.

$$P_{21} = P_{12} = - \frac{\hat{A}_{21}\,\hat{R}_{11}}{B_{21}^{2}} + \left[\left(\frac{\hat{A}_{21}\,\hat{R}_{21}}{B_{21}^{2}} \right)^{2} + \frac{\hat{Q}_{22}\,\hat{R}_{11}}{B_{21}^{2}} \left(L_{21} - A_{21} \right)^{2} \right]^{1/2}$$

$$P_{22} = - \frac{\hat{A}_{22}\,\hat{R}_{11}}{B_{21}^{2}} + \left[\left(\frac{\hat{A}_{22}\,\hat{R}_{11}}{B_{21}^{2}} \right)^{2} + \frac{2\,P_{21}\,\hat{R}_{11}}{B_{21}^{2}} + \frac{\hat{R}_{11}\,\hat{Q}_{22}}{B_{21}^{2}} \left(L_{22} - A_{22} \right)^{2} \right]^{1/2}$$

It is not necessary to calculate P_{11}, since u(t) depends on the product of B^T and P in Eq. (522), and the B_{11} element that would multiply P_{11} is zero.

Now since we are interested primarily in minimizing the "error" term ($\dot{y}$-Ly) in (514), it is logical to take $Q \gg R$. In the present case, it is sufficient to let $Q_{22} \gg R_{11}$. This leads to

$$K_m \approx \frac{1}{B_{21}} \left[\left(L_{21} - A_{21} \right) \left(L_{22} - A_{22} \right) \right]$$

while

$$\hat{Q}_{22} \rightarrow 0$$

which means that

$$P_{22} = P_{21} = P_{12} \approx 0$$

and, in turn,

$$K_r \approx 0$$

Consequently, the optimal control reduces to

$$\begin{aligned} u\,(t) &\approx - K_m\, x \\ &= \frac{-1}{B_{21}} \left[\left(L_{21} - A_{21} \right) \left(L_{22} - A_{22} \right) \right] \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \frac{-1}{B_{21}} \left[\left(L_{21} - A_{21} \right) x_1 + \left(L_{22} - A_{22} \right) x_2 \right] \end{aligned} \qquad (524)$$

Substituting this in (511),

$$\begin{aligned}\dot{x} &= Ax + Bu \\ &= (A - BK_m)x\end{aligned}$$

But

$$(A - BK_m) = \begin{bmatrix} 0 & 1 \\ -A_{21} & -A_{22} \end{bmatrix} - \frac{1}{B_{21}} \begin{bmatrix} 0 \\ B_{21} \end{bmatrix} \left[\left(L_{21} - A_{21}\right)\left(L_{22} - A_{22}\right)\right].$$

$$= \begin{bmatrix} 0 & 1 \\ -L_{21} & -L_{22} \end{bmatrix}$$

Thus the optimal system behaves exactly as the model. It is instructive to examine these results in terms of conventional transfer functions and feedback loops.

Fig. 14 depicts the feedback loops introduced by the optimal control function (524). The system with the feedback loops, shown in part (a), reduces to the form shown in part (b) after some elementary simplifications. The distinctive feature of this result is that near-infinite gains are not required to make the given plant behave as some prescribed model. In the conventional model-following scheme, shown in Fig. 15, the overall transfer function is

$$\frac{c\,(s)}{R\,(s)} = \frac{G\,(s)\,[1 + K\,M\,(s)]}{1 + K\,G\,(s)}$$

$$= \frac{\frac{G\,(s)}{K} + G\,(s)\,M\,(s)}{\frac{1}{K} + G\,(s)}$$

from which it is apparent that

$$\frac{c\,(s)}{R\,(s)} \rightarrow M\,(s) \text{ for } K \rightarrow \infty$$

Thus the scheme generated by optimal control theory appears extremely attractive. In practice, however, there are two fundamental limitations. First of all, there is no guarantee that response to external disturbances is acceptable. Second, all the state

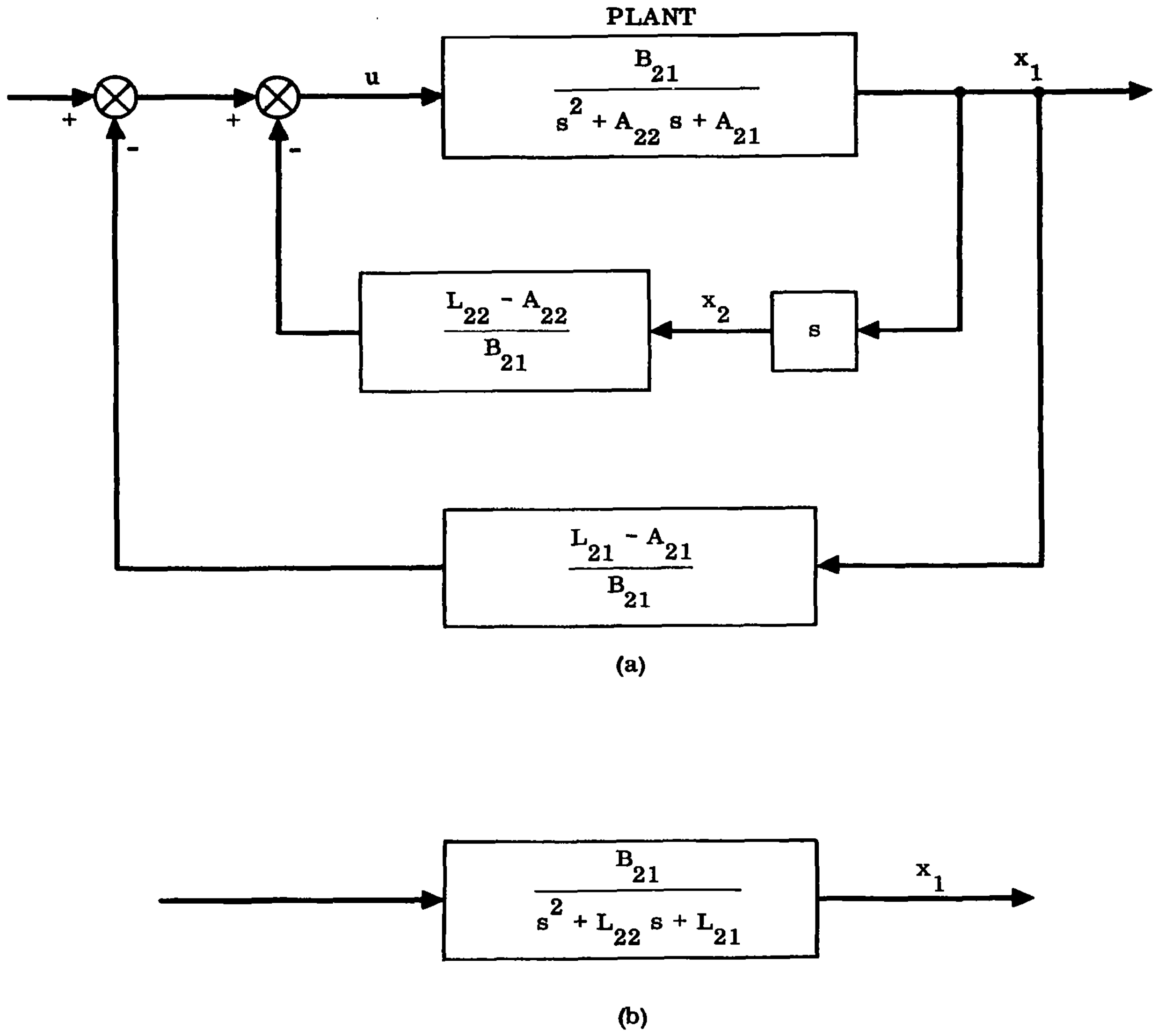

PLANT
$\frac{B_{21}}{s^2 + A_{22} s + A_{21}}$
u
x_1
+
-
+
-
$\frac{L_{22} - A_{22}}{B_{21}}$
x_2
s
$\frac{L_{21} - A_{21}}{B_{21}}$
(a)
$\frac{B_{21}}{s^2 + L_{22} s + L_{21}}$
x_1
(b)

Figure 14. A Model-Following Optimal Control System

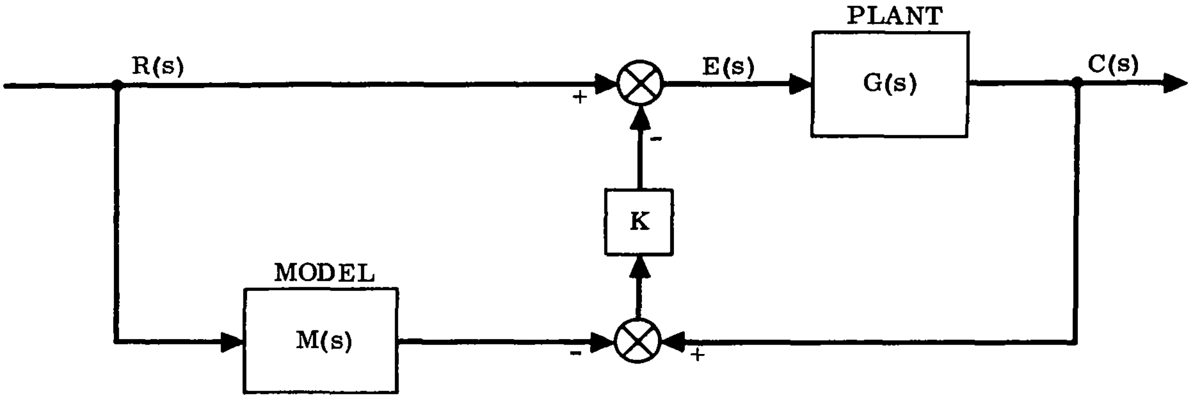


Figure 15. Conventional Model-Following System

variables must be accessible for measurement. For moderately high-order systems, the latter condition may be difficult to realize. Nevertheless, the above methods are a novel approach to an outstanding problem and may prove useful in certain applications.

Problem 3

Instead of including the model directly in the performance index, one may attempt to match the system output to the model output. This leads to a performance index of the form (484), which is repeated below.

$$J = \int_0^{t_f} \left[\left(y_D - y\right)^T Q \left(y_D - y\right) + \left(u_D - u\right)^T R \left(u_D - u\right) \right] dt \tag{525}$$

The distinction between this and the types previously considered is that the upper limit of integration, t_f, is finite. Consequently, it is anticipated that the optimal control function will contain time-varying gains. Bailey(122) applied Merriam's technique to the problem of optimizing the performance index (525) for the yaw dynamics of a launch vehicle whose motion is described by

$$\ddot{Y} = -\frac{L_\beta}{m U_0}\dot{Y} + \frac{1}{m}\left(T_c + L_\beta\right)\psi + \frac{T_c}{m}\delta \tag{526}$$

$$\ddot{\psi} = \frac{L_\beta \ell_\beta}{I_z}\psi - \frac{L_\beta \ell_\beta}{I_z U_0}\dot{Y} - \frac{T_c \ell_c}{I_z}\delta \tag{527}$$

Here

I_z = moment of inertia of vehicle about yaw axis

ℓ_c = distance from mass center of vehicle to engine swivel point

ℓ_β = distance from mass center of vehicle to center of pressure

L_β = aerodynamic load per unit angle of attack

m = mass of vehicle

T_c = control thrust

U_0 = vehicle forward velocity

Y = normal displacement measured parallel to an inertial reference

δ = control engine deflection

ψ = yaw attitude angle

Via the definitions

$x_1 = \dot{Y}$

$x_2 = Y$

$x_3 = \dot{\psi}$

$x_4 = \psi$

$u = \delta$

The pertinent system matrices become

$$A = \begin{bmatrix} -\frac{L_\beta}{m\,U_0} & 0 & 0 & \frac{1}{m}\,(T_c + L_\beta) \\ 1 & 0 & 0 & 0 \\ -\frac{L_\beta \ell_\beta}{I_z\,U_0} & 0 & 0 & \frac{L_\beta \ell_\beta}{I_z} \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

$$B = \begin{bmatrix} \frac{T_c}{m} \\ 0 \\ \frac{T_c \ell_c}{I_z} \\ 0 \end{bmatrix}$$

$G = I$, the unit matrix

The optimal control in this case is given by Eq. (487), which, in combination with (490) and the above value of B, reduces to

$$u = u_D - \frac{T_c}{R_{11}}\left[\frac{1}{m}\left(p_1 + \sum_{\beta=1}^{4} p_{\beta 1}\,x_\beta\right) + \frac{\ell_c}{I_z}\left(p_3 + \sum_{\beta=1}^{4} p_{\beta 3}\,x_\beta\right)\right] \tag{528}$$

where $R = R_{11}$ is a scalar, since u is a scalar. The p's are obtained from Eqs. (492) through (494), which in the present case take the form†

$$-\dot{p}_1 = Q_{11}\,\dot{Y}_D + A_{11}\,p_1 + p_2 + A_{31}\,p_3 - \frac{B_{11}^{\,2}}{R_{11}}\,p_1\,p_{11}$$

$$- \frac{B_{31}\,B_{11}}{R_{11}}\left(p_{11}\,p_3 + p_1\,p_{13}\right) - \frac{B_{31}^{\,2}}{R_{11}}\,p_{13}\,p_3 \tag{529}$$

†We recall that $p_{\alpha\beta} = p_{\beta\alpha}$

$$-\dot{p}_2 = Q_{22} Y_D - \frac{B_{11}^2}{R_{11}} p_1 p_{12} - \frac{B_{31} B_{11}}{R_{11}} \left(p_{12} p_3 + p_{23} p_1 \right) - \frac{B_{31}^2}{R_{11}} p_{23} p_3 \tag{530}$$

$$-\dot{p}_3 = Q_{33} \dot{\psi}_D + p_4 - \frac{B_{11}^2}{R_{11}} p_1 p_{13} - \frac{B_{31} B_{11}}{R_{11}} \left(p_3 p_{13} + p_1 p_{33} \right) - \frac{B_{31}^2}{R_{11}} p_3 p_{33} \tag{531}$$

$$-\dot{p}_4 = Q_{44} \psi_D + p_1 A_{14} + p_3 A_{34} - \frac{B_{11}^2}{R_{11}} p_1 p_{14} - \frac{B_{31} B_{11}}{R_{11}} \left(p_3 p_{14} + p_1 p_{34} \right) - \frac{B_{31}^2}{R_{11}} p_3 p_{34} \tag{532}$$

$$\dot{p}_{11} = Q_{11} + 2 A_{11} p_{11} + 2 p_{12} + 2 A_{31} p_{13} - \frac{B_{11}^2}{R_{11}} p_{11}^2 - \frac{2 B_{31} B_{11}}{R_{11}} p_{11} p_{13} - \frac{B_{31}^2}{R_{11}} p_{13}^2 \tag{533}$$

$$-\dot{p}_{12} = A_{11} p_{12} + p_{22} + A_{31} p_{23} - \frac{B_{11}^2}{R_{11}} p_{12} p_{11} - \frac{B_{31} B_{11}}{R_{11}} \left(p_{12} p_{13} + p_{23} p_1 \right) - \frac{B_{31}^2}{R_{11}} p_{13} p_{23} \tag{534}$$

$$-\dot{p}_{13} = A_{11} p_{13} + p_{23} + A_{31} p_{33} + p_{14} - \frac{B_{11}^2}{R_{11}} p_{11} p_{13}$$

$$- \frac{B_{31} B_{11}}{R_{11}} \left(p_{13}^2 + p_{11} p_{33}\right) - \frac{B_{31}^2}{R_{11}} p_{13} p_{33} \qquad (535)$$

$$-\dot{p}_{14} = A_{11} p_{14} + p_{24} + A_{31} p_{34} + A_{14} p_{11} + A_{34} p_{13} - \frac{B_{11}^2}{R_{11}} p_{11} p_{14}$$

$$- \frac{B_{31} B_{11}}{R_{11}} \left(p_{14} p_{13} + p_{11} p_{34}\right) - \frac{B_{31}^2}{R_{11}} p_{13} p_{34} \qquad (536)$$

$$-\dot{p}_{22} = Q_{22} - \frac{B_{11}^2}{R_{11}} p_{12}^2 - \frac{2 B_{31} B_{11}}{R_{11}} p_{12} p_{23} - \frac{B_{31}^2}{R_{11}} p_{23}^2 \qquad (537)$$

$$-\dot{p}_{23} = p_{24} - \frac{B_{11}^2}{R_{11}} p_{12} p_{13} - \frac{B_{31} B_{11}}{R_{11}} \left(p_{13} p_{23} + p_{12} p_{33}\right)$$

$$- \frac{B_{31}^2}{R_{11}} p_{23} p_{33} \qquad (538)$$

$$-\dot{p}_{24} = A_{14} p_{12} + A_{34} p_{23} - \frac{B_{11}^2}{R_{11}} p_{12} p_{14} - \frac{B_{31} B_{11}}{R_{11}} \left(p_{23} p_{14} + p_{12} p_{34}\right)$$

$$- \frac{B_{31}^2}{R_{11}} p_{23} p_{34} \qquad (539)$$

$$-\dot{p}_{33} = Q_{33} + 2 p_{34} - \frac{B_{11}^2}{R_{11}} p_{13}^2 - \frac{2 B_{31} B_{11}}{R_{11}} p_{13} p_{33} - \frac{B_{31}^2}{R_{11}} p_{33}^2 \qquad (540)$$

$$-\dot{p}_{34} = A_{14} p_{13} + A_{34} A_{33} + p_{44} - \frac{B_{11}^{2}}{R_{11}} p_{13} p_{14}$$

$$- \frac{B_{31} B_{11}}{R_{11}} \left(p_{13} p_{34} + p_{33} p_{14}\right) - \frac{B_{31}^{2}}{R_{11}} p_{33} p_{34} \qquad (541)$$

$$-\dot{p}_{44} = Q_{44} + 2 A_{14} p_{14} + 2 A_{34} p_{34} - \frac{B_{11}^{2}}{R_{11}} p_{14}^{2}$$

$$- \frac{2 B_{31} B_{11}}{R_{11}} p_{14} p_{34} - \frac{B_{31}^{2}}{R_{11}} p_{34}^{2} \qquad (542)$$

The weighting matrix Q has been assumed diagonal.

$$Q = \begin{bmatrix} Q_{11} & & & 0 \\ & Q_{22} & & \\ & & Q_{33} & \\ 0 & & & Q_{44} \end{bmatrix}$$

and the subscripted A and B terms are the components of the respective matrices.

The set of equations (530) through (542), when solved with the initial conditions

$$p_{\alpha}(t_f) = p_{\alpha\beta}(t_f) = 0 \qquad (543)$$

$$\alpha, \beta = 1, 2, 3, 4$$

yield the p's which, when substituted in Eq. (528), give a control function of the form

$$u(t) = u_D(t) + K_{\psi}(t)\psi + K_{\dot{\psi}}(t)\dot{\psi} + K_Y(t)\,Y + K_{\dot{Y}}(t)\,\dot{Y} \qquad (544)$$

Bailey(122) gives some results for a specific selection of R_{11} and the Q_{ii} that represents a compromise between the desirability for "tight" control and minimum drift. The superior results obtained by using time-varying gains in the feedback loops must be weighed against the added complexity of the control system. Furthermore, the influence of the neglected higher-order dynamic effects, nonlinearities, and parameter uncertainties, in addition to extraneous disturbances, remains to be evaluated.

Problem 4

We again consider the system

$$\dot{x} = Ax + Bu \tag{545}$$

$$y = Gx \tag{546}$$

with the performance criterion

$$J = \lim_{t_f \to \infty} \int_0^{t_f} \left(x^T G^T Q G x + u^T R u\right) dt \tag{547}$$

The optimal control has been found to be†

$$u(t) = -Kx(t) \tag{548}$$

where K is a constant matrix. Substituting this in Eq. (545), we obtain the equation of the optimal system

$$\dot{x} = (A - BK)\,x \tag{549}$$

We seek to determine the relationship between the components of K (feedback gains) and the components of the weighting matrices, Q and R.

An alternative derivation has shown that the optimal system and its adjoint satisfy‡

$$\begin{bmatrix} \dot{x} \\ \dot{\lambda} \end{bmatrix} = \begin{bmatrix} A & -BR^{-1}B^T \\ -G^T QG & -A^T \end{bmatrix} \begin{bmatrix} x \\ \lambda \end{bmatrix} \tag{550}$$

The characteristic equation of the optimal system, together with its adjoint, may therefore be obtained from (549) as

$$|Is - (A - BK)| \cdot |Is - (A + BK)| = 0 \tag{551}$$

or, from (550), as

$$\begin{bmatrix} Is - A & BR^{-1}B^T \\ G^T QG & Is + A^T \end{bmatrix} = 0 \tag{552}$$

†See Eq. (471).
‡See Eq. (480).

Since (551) and (552) are the characteristic equation for the same system, they must have the same roots. Therefore, by equating coefficients of like powers of s, we obtain a set of equations that relates the components of K to the components of Q and R. However, this brute-force approach soon runs into an avalanche of complicated algebra that yields little insight and no general results.

Tyler and Tuteur(13) show that if Q is a diagonal matrix and R is taken as the unit matrix, then for large Q_{ii}, the optimal system exhibits the properties of a Butterworth function. Their design approach involves the expansion of the determinant equation (552), which is the characteristic equation of the optimal system, together with its adjoint. A root locus is drawn for a specific Q_{ii} as the "system gain" and with all other Q_{ii} set equal to zero. After a Q matrix is determined in this fashion, the calculation of the K matrix is straightforward.

Rather than explore this technique in detail, we will indicate a simpler and more elegant approach that affords a higher degree of insight. This is based on a crucial simplification of the characteristic equation (552), which will now be developed. To do this, we will make use of the following two relations between determinants.

$$\begin{vmatrix} \alpha & \beta \\ \gamma & \delta \end{vmatrix} = |\alpha| \cdot |\delta - \gamma \alpha^{-1} \beta| \tag{553}$$

$$|\mu I_n - \gamma \beta| = \mu^{n-m} |\mu I_m - \beta \gamma| \tag{554}$$

where

$\alpha \equiv$ m × m matrix (nonsingular)

$\beta \equiv$ m × n matrix

$\gamma \equiv$ n × m matrix

$\delta \equiv$ n × n matrix

$\mu \equiv$ scalar

$I_i \equiv$ i × i unit matrix

The first of these is proven in Bodewig's book(172) (p. 217), and the second is due to Plotkin.(173) Applying (553) to (552) yields

$$|s I_n - A| \cdot |(s I_n + A^T) - G^T QG (s I_n - A)^{-1} BR^{-1} B^T| = 0 \tag{555}$$

But $|s\,I_n - A| = 0$ only for those values of s that correspond to the open-loop poles. Consequently, the characteristic equation for the closed-loop (optimal) system reduces to

$$|\left(s\,I_n + A^T\right) - G^T Q G \left(s\,I_n - A\right)^{-1} B R^{-1} B^T| = 0 \tag{556}$$

Furthermore,

$$\begin{aligned}&\left(s\,I_n + A^T\right) - G^T Q G \left(s\,I_n - A\right)^{-1} B R^{-1} B^T \\ &\quad = \left(s\,I_n + A^T\right)\left[I - \left(s\,I_n + A^T\right)^{-1} G^T Q G \left(s\,I_n - A\right)^{-1} B R^{-1} B^T\right]\end{aligned}$$

which means that

$$|s\,I_n + A^T| \cdot |I_n - \left(s\,I_n + A^T\right)^{-1} G^T Q G \left(s\,I_n - A\right)^{-1} B R^{-1} B^T| = 0$$

Now since $|s\,I_n + A^T| = 0$ only for those values of s that correspond to the open-loop poles of the adjoint system, Eq. (556) is further reduced to

$$|I_n - \left(s\,I_n + A^T\right)^{-1} G^T Q G \left(s\,I_n - A\right)^{-1} B R^{-1} B^T| = 0 \tag{557}$$

An application of (554) yields

$$|I_m - R^{-1} B^T \left(s\,I_n + A^T\right)^{-1} G^T Q G \left(s\,I_n - A\right)^{-1} B| = 0 \tag{558}$$

We now observe that the quantity

$$L(s) = G\left(I\,s - A\right)^{-1} B \tag{559}$$

is the matrix transfer function† for the system (545) and (546). Also,

$$L^T(-s) = -B^T\left(I\,s + A^T\right)^{-1} G^T \tag{560}$$

Therefore, Eq. (558) becomes

$$|I_m + R^{-1} L^T(-s)\,Q\,L(s)| = 0 \tag{561}$$

†See Ref. 174, Sec. 3.4.

This is, in fact, a multidimensional form of Chang's Root Square Locus method.(1) In principle, one could plot the loci for Q_{ii}/R_{jj} as a parameter and determine the Q and R that yield acceptable dynamic response. From this, the gain matrix, K, and hence the optimal system, could be determined.

Consider, for example, the single-input/single-output system for which

$$G \equiv 1 \times n \text{ matrix}$$

$$B \equiv n \times 1 \text{ matrix}$$

$$Q \equiv Q_{11} \equiv \text{scalar}$$

$$R \equiv R_{11} \equiv \text{scalar}$$

Then Eq. (561) takes the form

$$1 + \frac{Q_{11}}{R_{11}} L(-s) L(s) = 0 \tag{562}$$

where

$$\frac{Y(s)}{U(s)} = L(s) = \frac{N(s)}{D(s)} = \frac{k \prod_{j=1}^{r}\left(s - z_j\right)}{\prod_{i=1}^{n}\left(s - p_i\right)} \tag{563}$$

$$r \lessgtr n$$

and z_j and p_i are the zeros and poles respectively of the open-loop system, with k a known factor.

Defining

$$\Omega = - s^2$$

$$\mu_j = - z_j^2$$

$$\eta_i = - p_i^2$$

we express Eq. (562) as

$$0 = 1 + \left[\frac{k^2 Q_{11}}{R_{11}}\right] \frac{\prod_{j=1}^{r}\left(\Omega - \mu_j\right)}{\prod_{i=1}^{n}\left(\Omega - \eta_i\right)} = T(s)\, T(-s) \tag{564}$$

Thus standard root locus techniques can be used with $k^2\, Q_{11}/R_{11}$ as the variable gain. A given value of this variable gain yields pairs of roots Ω_i, each pair representing a point in the s plane, together with its mirror image about the imaginary axis; i.e., a closed-loop pole for the optimal and adjoint system. If these closed-loop roots are acceptable from a dynamic response point of view, then the corresponding Q_{11} and R_{11} are used to determine the optimal gain matrix, K.

Note that the optimal system contains only those roots in the left-hand s plane. It can be shown that the optimal system obtained by this procedure is always stable. Therefore, the mirror image poles belong to the adjoint of the optimal system.

In simple situations, the components of K may be directly related to Q_{11}/R_{11}. For the second-order case described by

$$\begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -b & -a \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \begin{bmatrix} 0 \\ c \end{bmatrix} u$$

$$y = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

we obtain, after equating Eq. (551) to (552),

$$\left[s^2 + s\left(a + c\,K_2\right) + \left(b + c\,K_1\right)\right]\left[s^2 - s\left(a + c\,K_2\right) + \left(b + c\,K_1\right)\right]$$

$$= \left(s^2 + a\,s + b\right)\left(s^2 - a\,s + b\right) + \frac{Q_{11}}{R_{11}}\,c^2$$

Equating coefficients of like powers of s and solving for K_1 and K_2 yields

$$K_1 = -\frac{b}{c} + \frac{b}{c}\sqrt{1 + \frac{Q_{11}}{R_{11}}\,\frac{c^2}{b^2}}$$

$$K_2 = -\frac{a}{c} + \sqrt{\frac{a^2}{c^2} + \frac{2\,K_1}{c}}$$

$$K = \begin{bmatrix} K_1 & K_2 \end{bmatrix}$$

Thus, because the relationships between the dynamic characteristics (such as closed-loop frequency and damping ratio) and the feedback gains are known, these can be related directly to the components of the Q and R matrices.

3.3.3 Optimal Re-entry From Orbit

The guidance and control problem for a manned spacecraft during the atmospheric re-entry phase must take account of two primary figures of merit: the surface heating rate and the accelerations experienced by the crew. To formulate the problem mathematically, the vehicle is assumed to be a point mass moving about a spherical, non-rotating earth with an inertial coordinate frame as shown in Fig. 16. The motion is described by

$$\dot{h} = - V \sin \gamma \tag{565}$$

$$\dot{V} = g \sin \gamma - \frac{D}{m} \tag{566}$$

$$\dot{\gamma} = \frac{g \cos \gamma}{V} - \frac{V \cos \gamma}{r_E + h} - \frac{L}{m V} \tag{567}$$

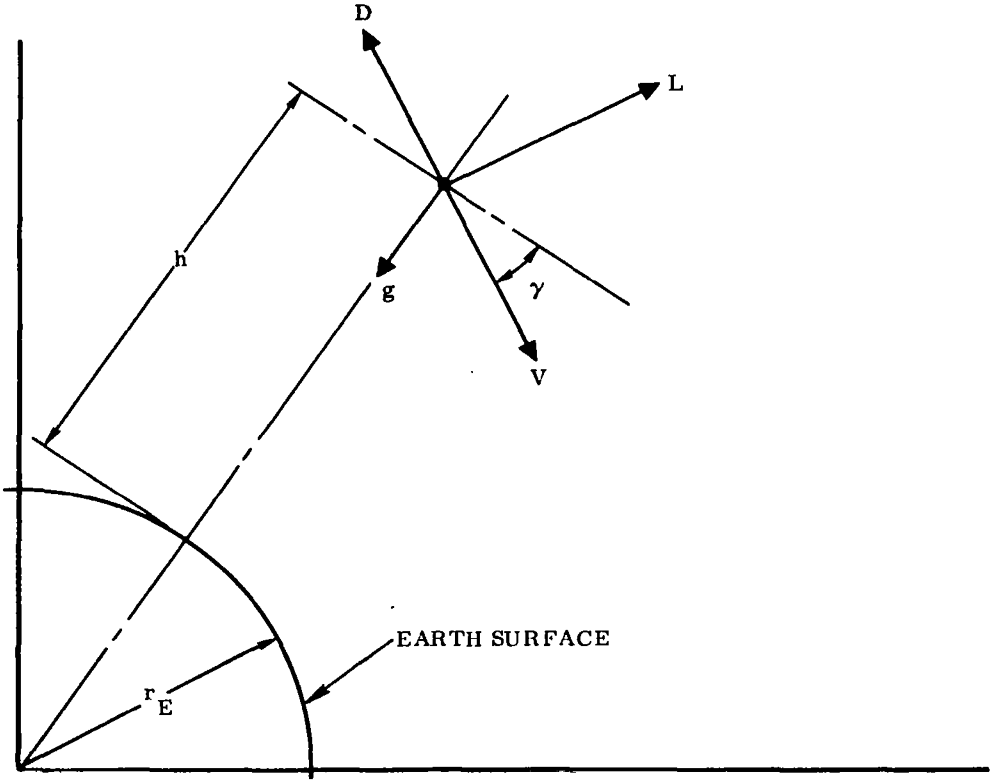


Figure 16. Coordinate System for Re-entry Problem

where

D = drag force

g = gravity acceleration

h = altitude above earth's surface

L = lift force

m = mass of vehicle

r_E = radius of earth

V = velocity of vehicle

γ = flight path angle

The rate of heating due to atmospheric friction is given by

$$K_4 \rho^{1/2} V^3 \tag{568}$$

where K_4 is a heating constant and ρ is the atmospheric density. The expression (568) represents the heating rate per unit surface area, so that for a particular vehicle, K_4 is a function of the surface area of the vehicle nose region.

The acceleration sensed by the crew is due only to the aerodynamic forces and is given by

$$\frac{\left(L^2 + D^2\right)^{1/2}}{m} \tag{569}$$

The limit of human endurance is a function of the acceleration magnitude and the length of time applied. Within the range of 5 to 10 g's, this endurance limit is roughly a linear function of the acceleration squared.

In view of these observations, a reasonable measure of performance may be expressed as

$$J = \int_{t_i}^{t_f} \left[K_4 \rho^{1/2} V^3 + \frac{K_7 \left(L^2 + D^2\right)}{m^2} \right] dt \tag{570}$$

where K_7 is a relative weighting constant between the heating and acceleration effects. It is convenient to define the additional state variables

$$x_4 = \int_{t_i}^{t} K_4 \rho^{1/2} V^3 \, dt \tag{571}$$

$$x_5 = \int_{t_i}^{t} \frac{\left(L^2 + D^2\right)}{m^2} \, dt \tag{572}$$

with

$$x_4 (t_i) = 0 \tag{573}$$

$$x_5 (t_i) = 0 \tag{574}$$

and add the equations

$$\dot{x}_4 = K_4 \rho^{1/2} V^3 \tag{575}$$

$$\dot{x}_5 = \frac{L^2 + D^2}{m^2} \tag{576}$$

to the system (565) through (567).

The performance criterion becomes, therefore,

$$J = x_4 (t_f) + K_7 x_5 (t_f) \tag{577}$$

For notational convenience, we define also

$$x_1 = h \qquad r_E = K_1$$

$$x_2 = V \qquad g = K_2$$

$$x_3 = \gamma \qquad m = K_3$$

We approximate the atmospheric density by the exponential model

$$\rho = K_5^{\,2} e^{K_6 x_1} \tag{578}$$

where K_5 and K_6 are constants.

The aerodynamic forces are expressed as

$$L = \frac{1}{2} \rho x_2^2 K_{10} C_L \tag{579}$$

$$D = \frac{1}{2} \rho x_2^2 K_{10} C_D \tag{580}$$

where K_{10} is a reference area and C_L and C_D are the lift and drag coefficients. The latter are approximated by

$$C_L = K_{11} \sin u \cos u \tag{581}$$

$$C_D = K_{12} + K_{13} \sin^2 u \tag{582}$$

where K_{11}, K_{12}, and K_{13} are appropriate constants for the lift drag polar and u is the angle of attack, which is taken as the control variable.

Combining all the above relations, the state equations for the system become

$$\dot{x}_1 = - x_2 \sin x_3 \equiv f_1 \tag{583}$$

$$\dot{x}_2 = K_2 \sin x_3 - \frac{K_5^2 K_{10}}{K_3} e^{K_6 x_1} x_2^2 \left(K_{12} + K_{13} \sin^2 u\right) \equiv f_2 \tag{584}$$

$$\dot{x}_3 = K_2 \frac{\cos x_3}{x_2} - \frac{x_2 \cos x_3}{K_1 + x_1}$$

$$- \frac{K_5^2 K_{10} K_{11}}{K_3} e^{K_6 x_1} x_2 \sin u \cos u \equiv f_3 \tag{585}$$

$$\dot{x}_4 = K_4 K_5 e^{\frac{1}{2} K_6 x_1} x_2^3 \equiv f_4 \tag{586}$$

$$\dot{x}_5 = \left(\frac{K_7 K_5^2 K_{10}^2}{K_3^2}\right) e^{2K_6 x_1} x_2^4 \left[K_{11}^2 \sin^2 u \cos^2 u + K_{12}^2 \right.$$

$$\left. + 2 K_{12} K_{13} \sin^2 u + K_{13}^2 \sin^4 u \right] \equiv f_5 \qquad (587)$$

The problem of optimal control is now formulated as follows.

"Given the system described by Eqs. (583) through (587), calculate the angle of attack history, u(t), that will minimize the function (577)."

The boundary conditions are

$$\begin{aligned} x_1(t_i) &= h_0 & x_1(t_f) &= h_f \\ x_2(t_i) &= V_0 & x_2(t_f) &= V_f \\ x_3(t_i) &= \gamma_0 & x_3(t_f) &= \gamma_f \\ x_4(t_i) &= 0 \\ x_5(t_i) &= 0 \end{aligned} \qquad (588)$$

This problem was analyzed by Payne,(175) using the maximum principle. We form the hamiltonian

$$H = \sum_{j=1}^{5} \lambda_i f_i \qquad (589)$$

where the λ_i satisfy†

†See Sec. 3.1.2

$$\dot{\lambda}_1 = \lambda_2\left[\left(\frac{K_5^{\,2} K_6 K_{10}}{K_3}\right) e^{K_6 x_1} x_2^{\,2} \left(K_{12} + K_{13} \sin^2 u\right)\right]$$

$$- \lambda_3\left[\frac{x_2 \cos x_3}{\left(K_1 + x_1\right)^2} - \left(\frac{K_5^{\,2} K_6 K_{10} K_{11}}{K_3}\right) e^{K_6 x_1} x_2 \sin u \cos u\right]$$

$$- \lambda_4\left[\left(\frac{K_4 K_5 K_6}{2}\right) e^{\frac{1}{2} K_6 x_1} x_2^{\,3}\right]$$

$$- \lambda_5\left[\left(\frac{2 K_5^{\,4} K_6 K_7 K_{10}^{\,2}}{K_3}\right) e^{2 K_6 x_1} x_2^{\,4} \left(K_{11}^{\,2} \sin^2 u \cos^2 u\right.\right.$$

$$\left.\left. + K_{12}^{\,2} + 2 K_{12} K_{13} \sin^2 u + K_{13}^{\,2} \sin^4 u\right)\right] \tag{590}$$

$$\dot{\lambda}_2 = \lambda_1 \sin x_3 + \lambda_2\left[\left(\frac{2 K_5^{\,2} K_{10}}{K_3}\right) e^{K_6 x_1} x_2 \left(K_{12} + K_{13} \sin^2 u\right)\right]$$

$$+ \lambda_3\left[\frac{K_2 \cos x_3}{x_2^{\,2}} + \frac{\cos x_3}{K_1 + x_1} + \left(\frac{K_5^{\,2} K_{10} K_{11}}{K_3}\right) e^{K_6 x_1} \sin u \cos u\right]$$

$$- \lambda_4\left[3 K_4 K_5 e^{\frac{1}{2} K_6 x_1} x_2^{\,2}\right]$$

$$- \lambda_5\left[\left(\frac{4 K_5^{\,4} K_7 K_{10}^{\,2}}{K_3^{\,2}}\right) e^{2 K_6 x_1} x_2^{\,3} \left(K_{11}^{\,2} \sin^2 u \cos^2 u\right.\right.$$

$$\left.\left. + K_{12}^{\,2} + 2 K_{12} K_{13} \sin^2 u + K_{13}^{\,2} \sin^4 u\right)\right] \tag{591}$$

$$\dot{\lambda}_3 = \lambda_1 x_2 \cos x_3 - \lambda_2 K_2 \cos x_3 + \lambda_3 \left[\frac{K_2 \sin x_3}{x_2} - \frac{x_2 \sin x_3}{K_1 + x_1} \right] \tag{592}$$

$$\dot{\lambda}_4 = 0 \tag{593}$$

$$\dot{\lambda}_5 = 0 \tag{594}$$

In order to minimize J, Eq. (577), we must maximize H, Eq. (598), with respect to u. If u is not bounded, it may be obtained from

$$\frac{\partial H(x, \lambda, u)}{\partial u} = 0 \tag{595}$$

However, a brief examination of the system (583) through (594) indicates that this is not a simple task. Thus it is necessary to use some iterative procedure to determine u such that H is a maximum at all points along the trajectory. This may be done, for example, by computing the optimal u such that the relation

$$\underset{u}{\text{Max}}\ H(x, \lambda, u) \tag{596}$$

is satisfied for all t. It is sufficient to let u assume a range of discrete values and calculate u by direct search. It is necessary simultaneously to solve the two-point boundary value problem represented by the 10 equations (583) through (587) and (590) through (594). Eight boundary conditions are specified by (588), and the remaining two are given by†

$$\lambda_4 (t_f) = -1 \tag{597}$$

$$\lambda_5 (t_f) = -K_7 \tag{598}$$

Payne(175) uses a variant of the Neighboring Optimum method(119) to solve this system. The optimal trajectory thus obtained is shown in Figs. 17 through 21, which depict the time histories of each of the state variables. The time history of the optimal control function is shown in Fig. 22. In each case, the solid line represents the condition $K_7 = 0$, while the dashed line is the case for $K_7 = 200$. The most pronounced effect of including the acceleration in the criterion function is shown in Fig. 21, which

†See Eq. (85) and the discussion in Sec. 3.1.2.

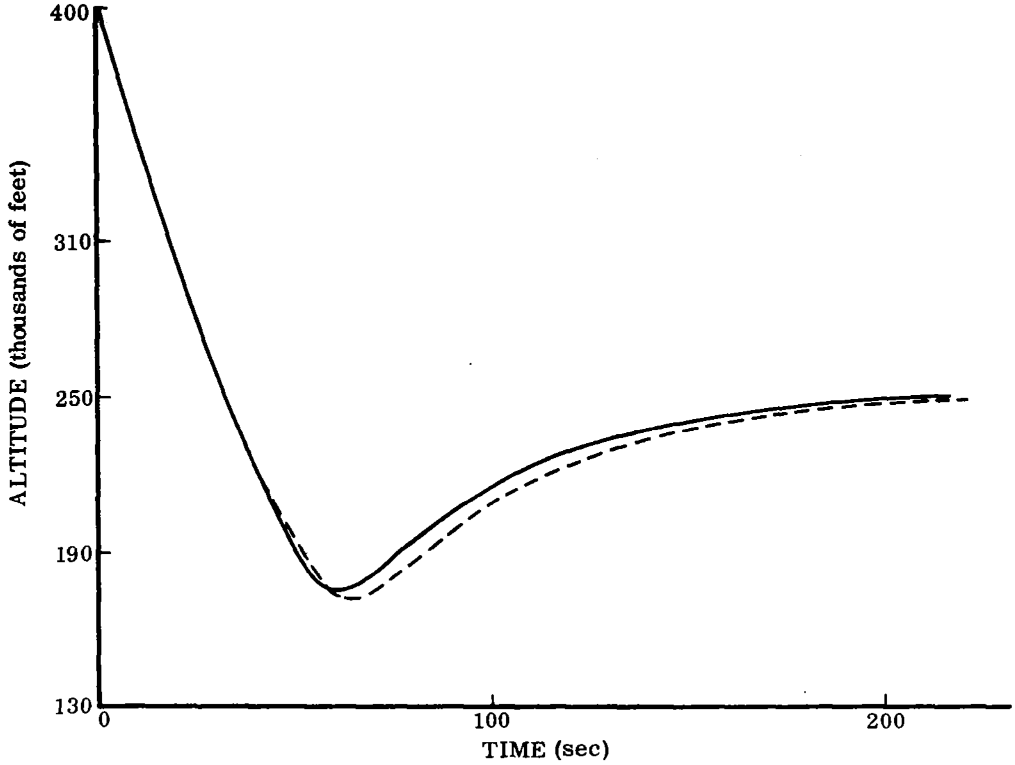


Figure 17. Optimal Trajectories, Time-Altitude

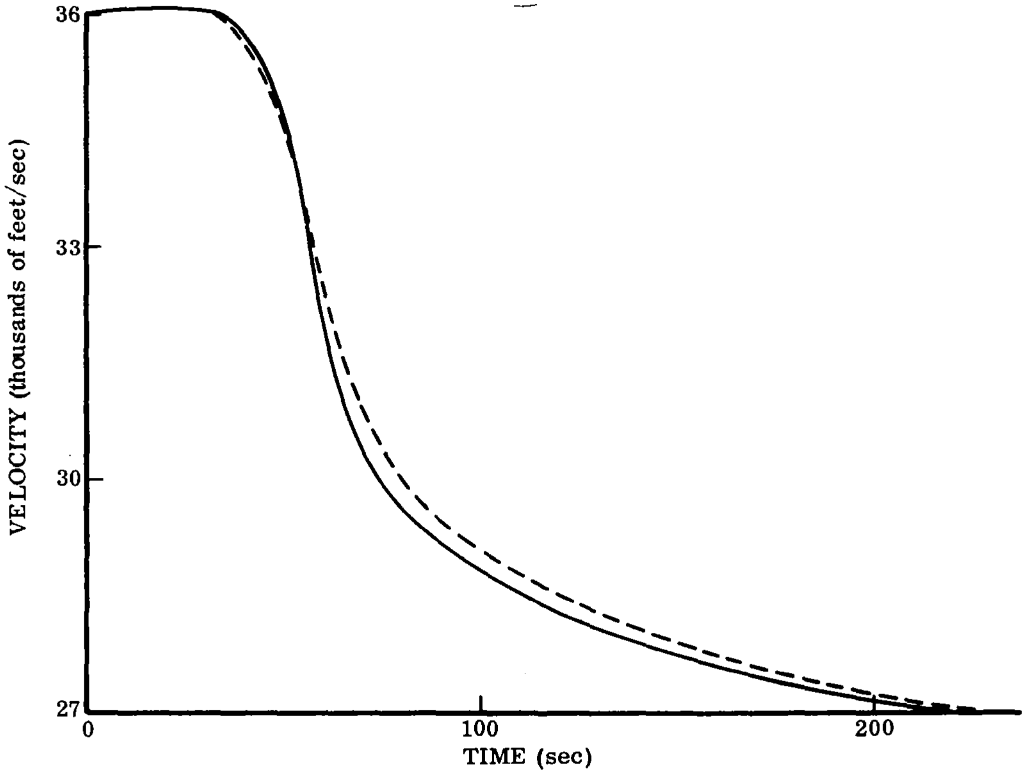


Figure 18. Optimal Trajectories, Time-Velocity

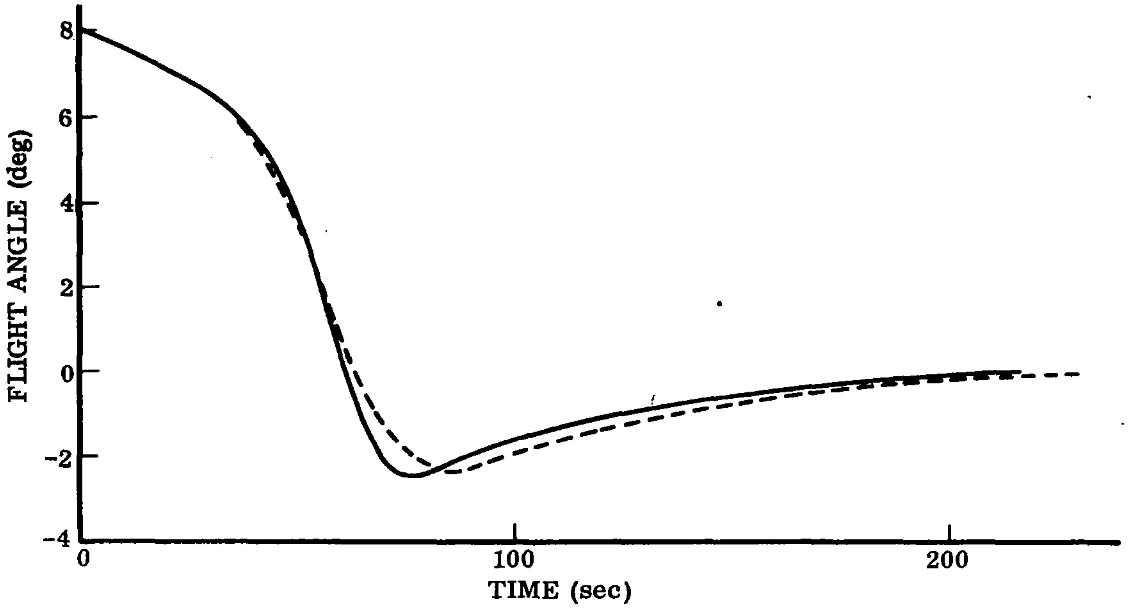


Figure 19. Optimal Trajectories, Time-Flight Angle

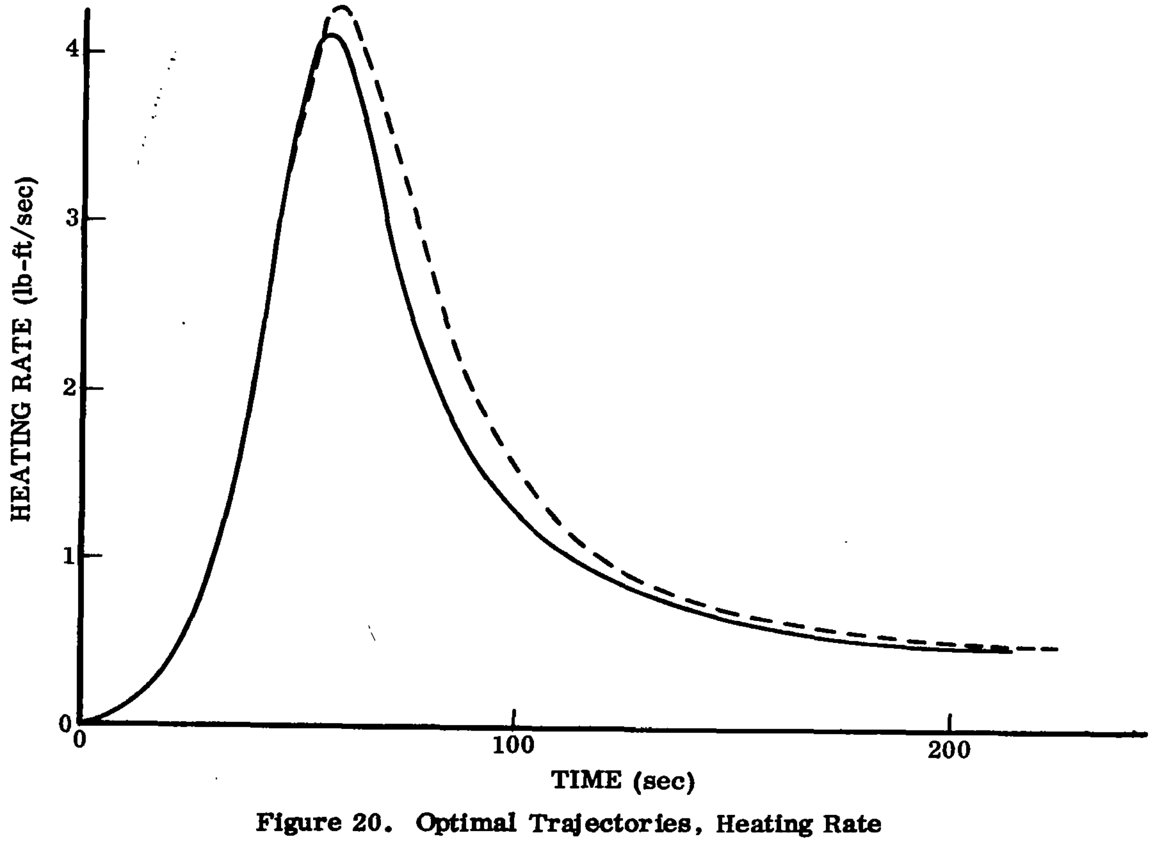


Figure 20. Optimal Trajectories, Heating Rate

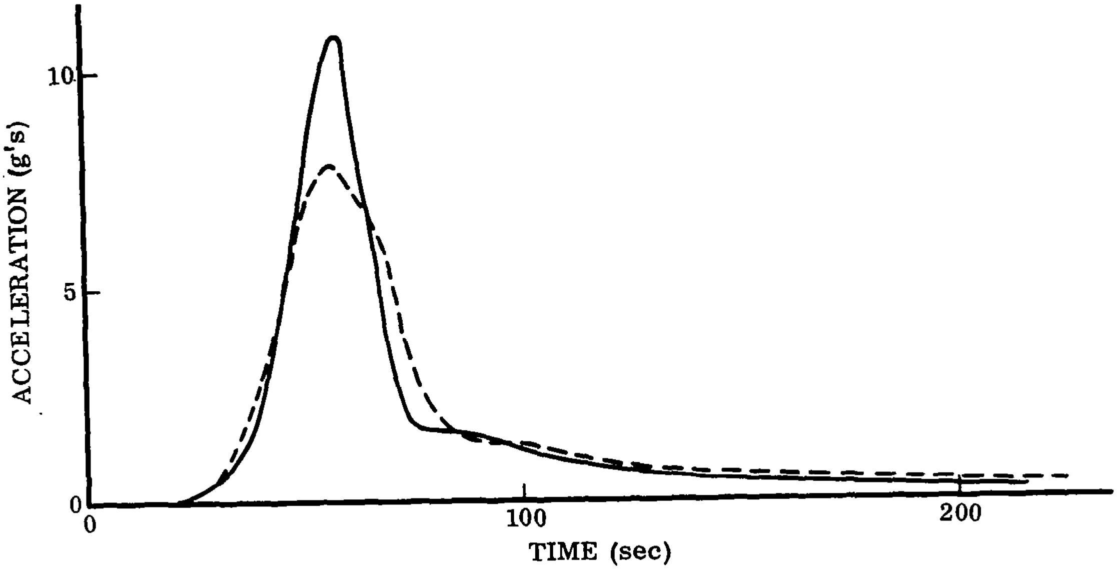


Figure 21. Optimal Trajectories, Acceleration Forces

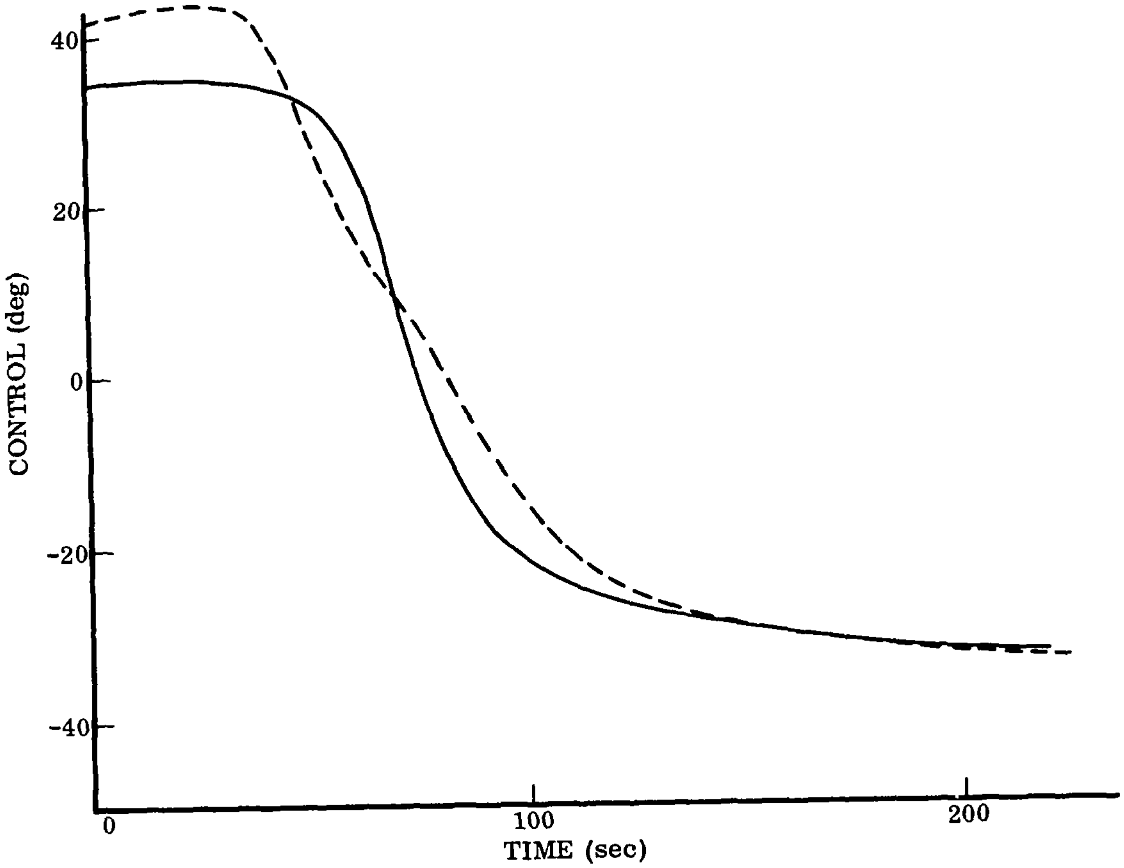


Figure 22. Optimal Trajectories, Control Function

indicates a reduction of about 25 percent in the g forces by using the weighting factor of $K_7 = 200$. The heating rate (Fig. 20) is thereby increased, but not significantly.

Fig. 23 shows how the total heat and acceleration effects along the optimal trajectory vary with parameter K_7. This plot can be used to determine the minimum amount of additional heating that the vehicle must absorb in order to achieve a specified reduction in acceleration effects. For example, to reduce the acceleration as shown in Fig. 21, it is necessary that the vehicle absorb about 10 percent additional heat during the re-entry maneuver.

The basic data for the problem is given below.

$K_1 = 2.09 \times 10^7$ ft

$K_2 = 32.2$ ft/sec^2

$K_3 = 250$ lb sec^2/ft

$K_4 = 1.0 \times 10^{-4}$ (lb)$^{1/2}$ sec

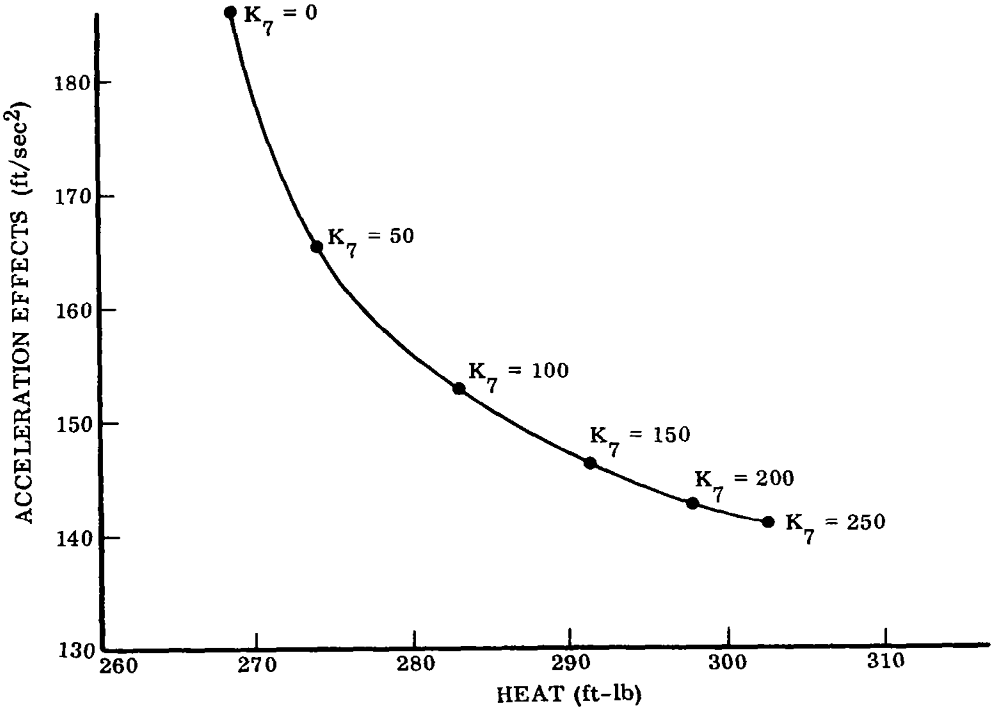


Figure 23. Optimal Tradeoff, Heat-Acceleration Effects

$K_5 = 0.052\ (\mathrm{lb})^{1/2}\ \mathrm{sec/ft}^2$

$K_6 = -4.26 \times 10^{-5}\ \mathrm{ft}^{-1}$

$K_{10} = 66.5\ \mathrm{ft}^2$

$K_{11} = 1.2$

$K_{12} = 0.274$

$K_{13} = 1.8$

Initial conditions:

$x_1(t_i) = 400{,}000$ ft

$x_2(t_i) = 36{,}000$ ft/sec

$x_3(t_i) = 8.09$ deg

Final conditions:

$x_1(t_f) = 250{,}000$ ft

$x_2(t_f) = 27{,}000$ ft/sec

$x_3(t_f) = 0$ deg

SECTION 4

REFERENCES

1. Chang, S. S. L. Synthesis of Optimal Control Systems, McGraw Hill Book Co., Inc., New York, N. Y., 1961.

2. Lawden, D. F. Optimal Trajectories for Space Navigation, Butterworth Mathematical Texts, London, 1963.

3. Leitmann, G. Optimization Techniques with Applications to Aerospace Systems, Academic Press, Inc., New York, N. Y., 1962.

4. Pontryagin, L. S., et al. The Mathematical Theory of Optimal Processes, John Wiley & Sons, Inc., New York, N. Y., 1962.

5. Tsien, H. S., and Evans, R. C. "Optimum Thrust Programming for a Sounding Rocket;" ARS Journ., 1951, p. 99-107.

6. Lee, E. B. "Design of Optimum Multivariable Control Systems," J. Basic Eng., March 1961, p. 85-90.

7. Greensite, A. Optimal Transfer Between Coplanar Orbits -- A General Approach, General Dynamics Convair Report ERR-AN - 229, 19 Nov. 1962.

8. Friedland, B. "Optimum Control of an Unstable Booster with Actuator Position and Rate Limits," AIAA Journal, Vol. 3, No. 7, 1965, p. 1268-1274.

9. Larsen, R. E. "Dynamic Programming with Reduced Computational Requirements," IEEE Trans. on Automatic Control, April 1965, p. 135-143.

10. Kahne, S. J. Feasible Control Computations Using Dynamic Programming, Air Force Cambridge Research Laboratory Report AFCRL - 65-232, April 1965.

11. Greensite, A. The Optimization of Missile Gust Response, General Dynamics Convair Report ERR-AN-135, 4 March 1962.

12. Greensite, A. Optimal Control Subject to State Variable Inequality Constraints with Application to the Booster Loads Problem, General Dynamics Convair Report ERR-AN-490, 13 March 1964.

13. Tyler, J. S., and Tuteur, F. B. "The Use of a Quadratic Performance Index to Design Multivariable Control Systems," IEEE Trans. on Automatic Control, Jan. 1966, p. 84-92.

14. Fisher, E. E. "An Application of the Quadratic Penalty Function Criterion to the Determination of a Linear Control for a Flexible Vehicle," AIAA Journ., Vol. 3, No. 7, 1965, p. 1262-1267.

15. Greensite, A. The Theory of Optimal Control with Application to Missiles and Space Vehicles, General Dynamics Convair Report ERR-AN-131, 21 March 1962.

16. Bellman, R. Dynamic Programming, Princeton University Press, 1957.

17. Bellman, R., and Dreyfus, S. Applied Dynamic Programming, Princeton University Press, 1962.

18. Miele, A. "General Variational Theory of the Flight Paths of Rocket Powered Aircraft, Missiles, and Satellite Carriers," Astronautica Acta, Vol. IV, 1958, p. 264.

19. Paiewonsky, B "Optimal Control: A Review of Theory and Practice," AIAA Journal, 1965, p. 1985-2006.

20. Leitmann, G. "A Calculus of Variations Solution of Goddard's Problem," Astronautica Acta, Vol. II, 1956, p. 55-62.

21. Hohmann, W. Die Erreichbarkeit der Himmelskorper, Oldenbourg, Munich, 1925; also NASA Translation TT-F-44, 1960.

22. Barrar, R. B. "An Analytic Proof That the Hohmann-type Transfer is the True Minimum Two-impulse Transfer," Astronautica Acta, Vol. IX, No. 1, 1963.

23. Lawden, D. F. "Optimal Intermediate-Thrust Arcs in a Gravitational Field," Astronautica Acta , Vol. VIII, No. 2, 1962, p. 106.

24. Lawden, D. F. "Minimal Trajectories," J. Brit. Interplanet. Soc., Vol. 9, July 1950, p. 179-186.

25. Lawden, D. F. "The Determination of Minimal Orbits," J. Brit. Interplanet. Soc., Vol. II, Sept. 1952, p. 216 - 224.

26. Lawden, D. F., "Minimal Rocket Trajectories," ARS Journ., Vol. 23, 1953, p. 360-365.

27. Lawden, D. F. "Stationary Rocket Trajectories," Quart. J. Mech. Appl. Mech., Vol. 7, Dec. 1954, p. 488-504.

28. Lawden, D. F. "Optimal Programming of Rocket Thrust Direction," Astronautica Acta, Vol. I, 1955, p. 41-56

29. Lawden, D. F. "Optimum Launching of a Rocket into an Orbit Around the Earth," Astronautica Acta, Vol. I, 1955, p. 185-190.

30. Edelbaum, T. N., "Some Extensions of the Hohmann Transfer Maneuver," ARS Journ., Vol. 29, 1959, p. 864.

31. Carstens, J. P., and Edelbaum, T.N. "Optimum Maneuvers for Launching Satellites into Circular Orbits of Arbitrary Radius and Inclination," ARS Journ., Vol. 31, 1961, p. 943-949.

32. Munick, H., McGill, R., and Taylor, G. E., "Minimization of Characteristic Velocity for Two-impulse Orbital Transfer," ARS Journ., Vol. 30, 1960, p. 638-639.

33. Altman, S. P., and Pistiner, J. S. "Minimum Velocity Increment Solution for Two-impulse Coplanar Orbital Transfer," AIAA Journ., Vol. I, 1963, p. 435-442.

34. Hoelker, R. F., and Silber, R. The Bi-elliptical Transfer Between Circular Coplanar Orbits, Dept. of the Army Tech. Memo 2-59, Army Ballistic Missile Agency, 1959.

35. Fimple, W. R. "Optimum Midcourse Plane Changes for Ballistic Interplanetary Trajectories," AIAA Journ., Vol. 1, 1963, p. 430-434.

36. Horner, J. M. "Optimum Impulsive Orbital Transfers Between Coplanar Orbits," ARS Journ., Vol. 32, 1962, p. 1082-1089.

37. Ting, L. "Optimum Orbital Transfer by Several Impulses," Astronautica Acta, Vol. VI, 1960, p. 256.

38. Wang, C. J., Anthony, G. W., and Lawrence, H. R. "Thrust Optimization of a Nuclear Rocket of Variable Specific Impulse," ARS Journ., Vol. 29, 1959, p. 341.

39. Melbourne, W. G., "Three-dimensional Optimum Thrust Trajectories for Power-Limited Propulsion Systems," ARS Journ., Vol. 31, 1961, p. 1723.

40. Melbourne, W. G., and Sauer, C. G., Jr. "Optimum Thrust Programs for Power-Limited Propulsion Systems," Astronautica Acta, Vol. VIII, 1962.

41. Dreyfus, S. E., and Elliott, J. R. "An Optimal Linear Feedback Guidance Scheme," J. Math. Anal. Appl., Vol. 8, 1964, p. 364-386.

42. Kelley, H. J. "An Optimal Guidance Approximation Theory," Inst. Elec., Electron. Engrs. Trans. Auto. Control, Vol. AC-9, Oct. 1964, p. 375.

43. Rekasius, Z. V. "A General Performance Index for Analytical Design of Control Systems," Inst. Radio Engrs., Trans. Auto. Control, Vol. AC-6, May 1961.

44. Zaborsky, J. "The Development of Performance Criteria for Automatic Control Systems," Am. Inst. Elec. Engrs. Inst. Tech. Groups Auto. Control, Vol. 1, No. 2, 1962.

45. Gibson, J. E., et al. "A Set of Standard Specifications for Linear Automatic Control Systems," Am. Inst. Elec. Engrs. Trans., Part II, 1961.

46. Walkovitch, J., Magdaleno, R. E., McRuer, D. T., Graham, F. D., and McDonnell, J. D. Performance Criteria for Linear Constant-Coefficient Systems with Deterministic Inputs, Wright Patterson Air Force Base Aeronautical Systems Div. Report TR 61-501, 1961.

47. Reynolds, P. A., and Rynaski, E. G. "Application of Optimal Linear Control Theory of the Design of Aerospace Vehicle Control Systems," Proceedings of the ASD Optimal System Synthesis Conference, 1962.

48. Rynaski, E. G., Reynolds, P. A., and Shed, W. H. Design of Linear Flight Control Systems Using Optimal Control Theory, Wright Patterson Air Force Base Aeronautical Systems Div. Report ASD-TDR-63-376, April 1964.

49. Goldstein, A. A., Greene, A. H., and Johnson, A. J. "Fuel Optimization in Orbital Rendezvous," AIAA Progress in Astronautics and Aeronautics: Guidance and Control II (edited by R. C. Langford and C. J. Mundo), Vol. 13, Academic Press, New York, N.Y., 1964, p. 823-844.

50. Hinz, H. K. "Optimal Low-thrust Near-circular Orbital Transfer," *AIAA Journ.*, Vol. 1, 1963, p. 1367-1371.

51. Neustadt, L. W. "Synthesizing Time Optimal Control Systems," *J. Math. Anal. Appl.*, Vol. 1, Dec. 1960, p. 4.

52. Meditch, J. S. "On the Problem of Optimal Thrust Programming for a Lunar Soft Landing," *Inst. Elec., Electron. Engrs. Trans. Auto. Control*, Vol. AC-9, 1964, p. 233.

53. Hall, B. A., Dietrich, R. G., and Tiernan, K. E. "A Minimum Fuel Vertical Touchdown Lunar Landing Guidance," *AIAA Progress in Astronautics and Aeronautics; Guidance and Control II* (edited by R. C. Langford and C. J. Mundo), Vol. 13, Academic Press, New York, N.Y., 1964, p. 965-994.

54. Neustadt, L. W. *A General Theory of Minimum-fuel Space Trajectories*, University of Michigan Report TR06181-1-T, November 1964.

55. Kelley, H. J. "Successive Approximation Techniques for Trajectory Optimization," *Proceedings of the IAS Symposium on Vehicle Systems Optimization*, 1961, p. 10.

56. Fuller, A. T. "Relay Control Systems Optimized for Various Performance Criteria," *Proceedings of the International Federation on Automatic Control Congress*, Butterworth's Scientific Publications Ltd., London, 1960.

57. Fried, B. D. "On the Powered Flight Trajectory of an Earth Satellite," *Jet Propulsion*, Vol. 27, 1957, p. 641-643.

58. Kelley, H. J. "An Investigation of Optimal Zoom-climb Techniques," *J. Aerospace Sci.*, Vol. 26, 1959, p. 794-802 and 824.

59. Miehle, A. "On the Non-steady Climb of Turbo-jet Aircraft," *J. Aeronaut. Sci.*, Vol. 21, Nov. 1954, p. 781-783.

60. MacKay, John S. *A Variational Method for the Optimization of Interplanetary Round-trip Trajectories*, NASA Report TN D-1660.

61. Hibbs, A. R. "Optimum Burning Program for Horizontal Flight," *ARS Journ.*, Vol. 22, 1952, p. 204-212.

62. Bryson, A. E., and Ross, S. E. "Optimum Rocket Trajectories with Aerodynamic Drag," *Jet Propulsion*, Vol. 28, 1958, p. 465-469.

63. Ross, S. "Composite Trajectories Yielding Maximum Coasting Apogee Velocity," *ARS Journ.*, Vol. 29, 1959, p. 843-848.

64. Kulakowski, L. J., and Stancil, R. L. "Rocket Boost Trajectories for Maximum Burnout Velocity," *ARS Journ.*, Vol. 30, 1960, p. 612-618.

65. Stancil, R. T., and Kulakowski, L. J. "Rocket Boost Vehicle Mission Optimizations," *ARS Journ.*, Vol. 31, 1961, p. 935.

66. Breakwell, J. V. "The Optimization of Trajectories," *J. Soc. Ind. Appl. Math.*, Vol. 7, June 1959, p. 215-247.

67. Leitmann, G. "Optimal Thrust Direction for Maximum Range," *J. Brit. Interplanet. Soc.*, Vol. 16, 1958, p. 503-507.

68. Lawden, D. F. "Optimal Program for Correctional Maneuvers," *Astronautica Acta*, Vol. IV, 1958, p. 264.

69. Striebel, C. T., and Breakwell, J. V. "Minimum Effort Control in Interplanetary Guidance," *IAS Paper*, January 1963, p. 63-80.

70. Lawden, D. F. "Interplanetary Rocket Trajectories," *Advances in Space Sciences*, (edited by F. I. Ordway III), Vol. 1, Academic Press, New York, N.Y., 1959, p. 1053.

71. Leitmann, G. "The Optimization of Rocket Trajectories," *Progress in the Astronautical Sciences*, Vol. 1, North Holland Publishing Co., Amsterdam, Holland, 1962.

72. Garfinkel, B. "Minimal Problems in Airplane Performance," *Quart. Appl. Math*, Vol. 9, No. 2, 1951.

73. Cicala, P., and Miele, A. "Brachistochronic Maneuvers of a Variable Mass Aircraft in a Vertical Plane," *J. Aeronaut. Sci.*, Vol. 22, August 1955, p. 577-578.

74. Hestenes, M. R. A General Problem in the Calculus of Variations with Applications to Paths of Least Time, RAND RM 100, Rand Corp., Santa Monica, Calif., March 1949.

75. Kelley, H. J. "Gradient Theory of Optimal Flight Paths," ARS J., Vol. 30, 1960, p. 947-954.

76. Bryson, A. E., Denham, W. F., Carroll, F.J., and Mikami, M. "Determination of Lift or Drag Programs to Minimize Re-entry Heating," J. Aerospace Sci., Vol. 29, April, 1962.

77. Levinsky, E. S. "Application of Inequality Constraints to Variational Problems of Lifting Re-entry," IAS Paper, January 1961, p. 61-21; also Wright Air Development Div. Rept. TR 60-369, July 1960.

78. Flugge-Lotz, I. Discontinuous Automatic Control, Princeton University Press, Princeton, N. J., 1953.

79. Kazda, L. "Control System Optimization Using Computers as Control System Elements," Proceedings of Computer in Control Systems Conference, American Institute of Electrical Engineers, 1958.

80. Kreindler, E. "Contributions to the Theory of Time-Optimal Control," J. Franklin Inst., April 1963, p. 275.

81. Bellman, R., Glicksberg, I., and Gross, O. "On the Bang-Bang Control Problem," Quart. Appl. Math., Vol. 14, 1956, p. 11-18.

82. LaSalle, J. P. "The Time Optimal Control Problem," Contributions to the Theory of Nonlinear Oscillations, Vol. V, Princeton University Press, Princeton, N. J., 1960.

83. Gamkrelidze, R. V. "The Theory of Time Optimal Processes in Linear Systems," Bull. Acad. Sci. USSR English Transl., Vol. 2, 1958, p. 449-474.

84. Rozonoer, L. I. "L. S. Pontriagin's Maximum Principle in the Theory of Optimum Systems, I, II, III," Avtomat. i Telemeh., Vol. 20, October, November, December 1959, p. 1320-1334, 1441, 1458, 1561-1578; Automation Remote Control Transl., Vol. 20, June, July, August, 1960, p.1288-1302, 1405-1421, 1517-1532.

85. Weiss, H. K. "Analysis of Relay Servomechanisms," J. Aeronaut Sci., Vol. 13, July 1946, p. 364.

86. Wang, P. K. C. "Analytical Design of Electrohydraulic Servomechanisms with Near Time-Optimal Response," Inst. Elec., Electron. Engrs., Trans. Auto Control, Vol. AC-8, January 1963.

87. Flugge-Lotz, I., and Marbach, H. "The Optimal Control of Some Attitude Control Systems for Different Performance Criteria," Proceedings of the Joint Automatic Control Conference, 1962.

88. Meditch, J. S. "On Minimal-Fuel Satellite Attitude Controls," Inst. Elec., Electron. Engrs., Trans. Appl. Ind., No. 71, March 1964, p. 120.

89. Friedland, B. "A Minimum Response-Time Controller for Amplitude and Energy Constraints," Inst. Radio Engrs., Trans. Auto. Control, Vol. AC-7, January 1962.

90. Athans, M., Falb, P. L., and Lacoss, R. T. "Time-, Fuel-, and Energy-Optimal Control of Nonlinear Norm-invariant Systems," Inst. Elec., Electron. Engrs., Trans. Auto. Control, Vol. AC-8, July 1963.

91. Chang, S. S. L. "Minimal Time Control with Multiple Saturation Limits," Inst. Elec., Electron. Engrs., Trans. Auto. Control, Vol. AC-8, January 1963.

92. Knudson, H. K. "An Iterative Procedure for Computing Time Optimal Controls," Inst. Elec., Electron. Engrs., Trans. Auto. Control, Vol. AC-9, January 1964, p. 23.

93. Kalman, R. E. "The Theory of Optimal Control and the Calculus of Variations," Research Institute for Advanced Studies, Baltimore, Md., Report TR 61-3, 1961.

94. Lee, E. B. "Geometric Properties and Optimal Controllers for Linear Systems," Inst. Elec., Electron. Engrs., Trans. Auto. Control, Vol. AC-8, October 1963, p. 379.

95. Kalman, R.E., and Koepche, R. W. "Optimal Synthesis of Linear Sampling Control Systems Using Generalized Performance Indices," Trans. Am. Soc. Mech. Engrs., November 1958, p. 1120.

96. Katz, S. "A Discrete Version of Pontriagin's Maximum Principle," J. Electron. Control, Vol. XIII, August 1962, p. 179.

97. Hestenes, M. R. "Variational Theory and Optimal Control Theory," Computing Methods in Optimization Problems (edited by A. V. Balakrishnan and L. W. Neustadt), Academic Press, Inc., New York, N.W., 1964.

98. Bolza, O. Lectures on the Calculus of Variations, Dover Publications, Inc., New York, N. Y., 1961.

99. Valentine, F. A., "The Problem of Lagrange with Differential Inequalities as Added Side Conditions," Contributions to the Calculus of Variations, 1933-1937, University of Chicago Press, Chicago, Ill., 1937.

100. Berkovitz, L. D. "On Control Problems with Bounded State Variables," J. Math. Anal. Appl., Vol. 5, December 1962.

101. Kipiniak, W. Dynamic Optimization and Control: A Variational Approach, Technical Press, Massachusetts Institute of Technology, Cambridge, Mass., and John Wiley and Sons, Inc., New York, N. Y., 1961.

102. Gamkrelidze, R. V. "Time Optimal Processes with Bounded Phase Coordinates," Dokl. Akad. Nauk. SSSR, Vol. 125, 1959, p. 475-478.

103. Dreyfus, S. Variational Problems with State Variable Inequality Constraints, Rand Corp. Report P-2605, July 1962.

104. Denham, W. F. Steepest-Ascent Solution of Optimal Programming Problems, Raytheon Co., Bedford, Mass., Report BR-2393, April 1963; also Bryson, A. E., and Denham, W. F. , J. Appl. Mech., Vol. 29, June 1962.

105. Bryson, A. E., Denham, W. F., and Dreyfus, S. E. "Optimal Programming Problems with Inequality Constraints I: Necessary Conditions for Extremal Solutions," AIAA J., Vol. 1, 1963, p. 2544-2550.

106. Denham, W.F., and Bryson, A. E. "Optimal Programming Problems with Inequality Constraints II: Solution by Steepest-Ascent," AIAA J., Vol. 2, 1964, p. 25-34.

107. Desoer, C. A. "The Bang-Bang Servo Problem Treated by Variational Technique," Inform. Control, Vol. 2, December 1959, p. 333-348.

108. Neustadt, L. W. "Minimum Effort Control Systems," Soc. Ind. Appl. Math. J., Vol. 1, 1963.

109. Eaton, J. H. "An Iterative Solution to Time-Optimal Control," J Math. Anal. Appl., Vol. 5, 1962, p. 329-344.

110. Ho, Y. C. "A Successive Approximation Technique for Optimal Control Systems Subject to Input Saturation," J. Basic Eng., Vol. 84, March 1962, p. 33-40.

111. Draper, C., and Li, Y. "Principles of Optimizing Control Systems and an Application to the Internal Combustion Engine," Mech. Eng., Vol. 74, February 1952, p. 145.

112. Rose, N. J. "Optimum Switching Criteria for Discontinuous Automatic Controls," Institute of Radio Engineers Convention Record, Part 4, 1956, p. 61.

113. Fadden, E. I. and Gilbert, E. G. "Computational Aspects of the Time-Optimal Control Program," Computing Methods in Optimization Problems (edited by A. V. Balakrishnan and L. W. Neustadt), Academic Press, Inc., New York, N.Y., 1964.

114. Fletcher, R., and Powell, M. J. D. "A Rapidly Convergent Descent Method for Minimization," Computer J., July 1963.

115. Powell, M. J. D. "An Iterative Method for Finding Stationary Values of a Function of Several Variables," Computer J., July 1962.

116. Paiewonsky, B. H., Woodrow, P., Terkelsen, F., and McIntyre, J. A Study of Synthesis Techniques for Optimal Control, Aeronautical Systems Div. Report ASD-TDR-63-239, Wright-Patterson Air Force Base, Ohio, June 1964.

117. Paiewonsky, B., Woodrow, P., Brunner, W., and Halbert, P. "Synthesis of Optimal Controllers Using Hybrid Analog-Digital Computers," Computing Methods in Optimization Problems (edited by A. V. Balakrishnan and L. W. Neustadt), Academic Press, Inc., New York, N.Y., 1964.

118. Elsgolc, L. E. Calculus of Variation, Addison-Wesley Publishing Company, Inc., Reading, Mass., 1962.

119. Breakwell, J. V., Speyer, J. L., and Bryson, A. E. "Optimization and Control of Nonlinear Systems Using the Second Variation," SIAM J., Series A: Control 1, 1963, p. 193.

120. Merriam, C. W., III "An Algorithm for the Iterative Solution of a Class of Two-Point Boundary Value Problems," SIAM J., Series A: Control 2, 1964, p. 1.

121. Ellert, F. J., and Merriam, C.W., III "Synthesis of Feedback Controls Using Optimization Theory—an Example," Inst. Elec., Electron. Engrs. Trans. Auto. Control, Vol. AC-8, April 1963.

122. Bailey, F. B. "The Application of the Parametric Expansion Method of Control Optimization to the Guidance and Control Problem of a Rigid Booster," Inst. Elec., Electron. Engrs., Trans. Auto. Control, Vol. 9, January 1964.

123. Schwartz, L. Optimization Techniques — a Comparison, U.S. Air Force Flight Dynamics Lab. Report TDR-64-21, Wright-Patterson Air Force Base, Ohio, March 1964.

124. Peterson, E. L., and Remond, F. X. Investigations of Dynamic Programming Methods for Flight Control Problems, U. S. Air Force Flight Dynamics Lab. Report TDR-64-11, Wright-Patterson Air Force Base, Ohio, March 1964.

125. Balakrishnan, A. V., and Neustadt, C. W. "Computing Methods in Optimization Problems," Proceedings of the January 1964 UCLA Conference, Academic Press, Inc., New York, N. Y., 1964.

126. Curry, H. B. "The Method of Steepest Descent for Nonlinear Minimization Problems," Quart. Appl. Math., Vol. II, October 1944.

127. Levenberg, K. "A Method for the Solution of Certain Nonlinear Problems in Least Squares," Quart. Appl. Math., Vol. II, July 1944.

128. Edelbaum, T. N. "Theory of Maxima and Minima," Optimization Techniques (edited by G. Leitmann), Academic Press, Inc., New York, N. Y., 1962, Chap. I.

129. Spang, H. A., III "A Review of Minimization Techniques for Nonlinear Functions," SIAM Review, Vol. 4, October 1962.

130. Wilde, D. J. *Optimum Seeking Methods,* Prentice-Hall, Inc., Englewood Cliffs, N. J., 1964.

131. Shah, B. V., Buehler, R. J., and Kempthorne, O. "Some Algorithms for Minimizing a Function of Several Variables," *J. SIAM,* Vol. 12, March 1964, p. 74.

132. Tompkins, C. B. "Methods of Steep Descent," *Modern Methematics for the Engineer* (edited by E. F. Beckenbach), McGraw-Hill Book Co., Inc., New York, N. Y., 1956, Chap. 18.

133. Bellman, R., "On the Application of the Theory of Dynamic Programming to the Study of Control Processes," *Proceedings of the Symposium on Nonlinear Circuit Analysis,* Polytechnic Press, Brooklyn, N. Y., 1956, p. 199-213.

134. Bliss, G. A. *Lectures on the Calculus of Variations,* University of Chicago Press, Chicago, Ill., 1946, p. 296.

135. Ho, Y.C., and Brentani, P. B. *On Computing Optimal Control with Inequality Constraints,* Minneapolis-Honeywell Report 1529 TR-5, Boston, Mass., March 1962; also, *SIAM J. Control,* Vol. 1, 1963, p. 319.

136. Tyler, J. S., Jr. "The Characteristics of Model-following Systems as Synthesized by Optimal Control," *Proceedings of the Joint Automatic Control Conference,* 1964, p. 41.

137. Krasovskii, N.N., and Letov, A.M. "The Theory of Analytic Design of Controllers," *Automation and Remote Control,* Vol. 23, June 1962.

138. Rekasius, Z.V., and Hsia, T.C. "On an Inverse Problem in Optimal Control," *Inst. Elec., Electron. Engrs., Trans. Auto. Control,* Vol. AC-9, October 1964.

139. Knapp, C.H., and Frost, P.A. "Determination of Optimal Control and Trajectories Using the Maximum Principle in Association with a Gradient Technique," *Proceedings of the Joint Automatic Control Conference,* 1964, p. 222ff.

140. Jurovics, S.A., and McIntyre, J. E. "The Adjoint Method and its Application to Trajectory Optimization," *ARS J.,* Vol. 32, 1962, p. 1354.

141. Greenley, R. R. "Comments on 'The Adjoint Method and its Application to Trajectory Optimization,'" *AIAA J.,* Vol. 1, 1963, p 1463.

142. Paiewonsky, B. H. A Study of Time Optimal Control, Aeronautical Research Associates of Princeton, Report 33, June 1961; also Proceedings of International Symposium on Nonlinear Differential Equations and Nonlinear Mechanics (edited by J. P. LaSalle and S. Lefshetz), Academic Press, Inc., New York, N. Y., 1963, p. 333-365.

143. Smith, F. B., Jr., and Lovingood, J. A. "An Application of Time-Optimal Control Theory to Launch Vehicle Regulation," Proceedings of the Optimum System Synthesis Conference, September 11-13, 1962.

144. Kelley, H. "Optimization Techniques," Method of Gradients (edited by G. Leitman), Academic Press, Inc., New York, N. Y., 1962, Chap. 6.

145. McGill, R., and Kenneth, P. "Solution of Variational Problems by Means of a Generalized Newton-Raphson Operator," AIAA J., Vol. 2, 1964, p. 1761.

146. Merriam, C. W. Optimization Theory and the Design of Feedback Control Systems, McGraw Hill Book Co., Inc., New York, N. Y., 1964.

147. Goodman, T. R., and Lance, G. N. "The Numerical Integration of Two Point Boundary Value Problems," Math. Tables and Other Aids to Computation, April 1956, p. 82-86.

148. Long, R. S. "Quasilinearization and Orbit Determination," AIAA Journ., Vol. 3, No. 10, 1965, p. 1937-1940.

149. Kalaba, R. "On Nonlinear Differential Equations; the Maximum Operation and Monotone Convergence," J. Math. and Mech., Vol. 8, 1959, p. 519-574.

150. Miehle, A. General Variational Approach to the Optimum Thrust Programming for the Vertical Flight of a Rocket, Report AFOSR-TN-57, March 1957.

151. Courant, R. "Variational Methods for the Solution of Problems of Equilibrium and Vibrations," Bull. Am. Math. Soc., Vol. 49, No. 1, 1943.

152. Dreyfus, D. "The Numerical Solution of Variational Problems," J. Math. Anal. and Appl., Vol. 5, 1962, p. 30-45.

153. Jazwinski, A. H. "Inequality Constraints in Steepest Descent Trajectory Optimization," *J. Aero. Sci.*, October 1962, p. 1268.

154. Taylor, A. E. *Introduction to Functional Analysis*, John Wiley & Sons, Inc., New York, N.Y. , 1958.

155. Zadeh, L. A., and Desoer, C. A. *Linear System Theory*, McGraw Hill Book Co., Inc., New York, N. Y., 1963.

156. Courant, R. "Variational Methods for the Solution of Problems of Equilibrium and Vibrations," *Bull, Am. Math. Soc.*, 1953, p. 1-23.

157. Butler, T., and Martin, A. V. "On a Method of Courant for Minimizing Functionals," *J. Math. and Physics*, Vol. 41, 1962, p. 291-299.

158. Jazwinski, A. H. *Steepest Descent Trajectory Optimization with Inequality Constraints*, General Dynamics Report ERR-AN-172, 6 June 1962.

159. Greensite, A. *Notes on Dynamic Programming*, General Dynamics Convair Report ERR-AN-217, 25 October 1962.

160. Cartaino, T.F., and Dreyfus, S. E. "Application of Dynamic Programming to the Airplane Minimum Time to Climb Problem," *Aero. Eng. Review*, 1957, p. 74-77.

161. Reid, W. T. "A Matrix Differential Equation of the Riccati Type," *Am. J. of Math.*, Vol. 68, 1946, p. 237-246.

162. Bellman, R. *Introduction to Matrix Analysis*, McGraw Hill Book Co., Inc., New York, N. Y., 1960.

163. Frazer, R. A., Duncan, W. J., and Collar, A. R. *Elementary Matrices*, Cambridge University Press, London, 1938.

164. Isaev, V. K. "Pontryagin's Maximum Principle and Optimal Programming of Rocket Thrust," *Automation and Remote Control*, Vol. 22, No. 8, 1961, p. 986-1001.

165. McGill, R., and Kenneth, P. "A Convergence Theorem on the Iterative Solution of Nonlinear Two Point Boundary Value Systems," *XIVth Int. Astronautical Fed. Congress*, Paris, 1963.

166. Long, R. S. "Newton Raphson Operator; Problems with Undetermined End Points," AIAA Journ., 1965, p. 1351.

167. Greensite, A. Design Criteria for Control of Space Vehicles, Vol. I, Part 1, Short Period Dynamics, Convair Report GDC-DDE65-055, 1 October 1965.

168. Greensite, A. A Functional Approximation Technique in N-Space with Application to Dynamic Programming, General Dynamics Convair Report ERR-AN-107, 15 January 1962.

169. Greensite, A. A Further Note on Functional Approximation in N-Space, General Dynamics Convair Report ERR-AN-262, 9 January 1963.

170. Bellman, R., and Dreyfus, S. "Functional Approximations and Dynamic Programming," Math Tables and Other Aids to Computation, Vol. XIII, 1959, p. 247.

171. Lanczos, C. "Trigonometrical Interpolation of Empirical and Analytical Functions," Journ. of Math. and Physics, 1938, p. 123.

172. Bodewig, E. Matrix Calculus, North Holland Pub. Co., Amsterdam, 1959, 2nd Ed.

173. Plotkin, M. "Matrix Theorem with Applications Related to Multivariable Control Systems," IEEE Trans. on Automatic Control, Vol. AC-9, 1964, p. 120.

174. Greensite, A. Design Criteria for Control of Space Vehicles, Vol. II, part 1, Linear Systems, General Dynamics Convair Report GDC-DDE66-019, 25 April 1966.

175. Payne, J. A. Computational Methods in Optimal Control Problems, Air Force Flight Dynamics Lab. Report TR-65-50, August 1965.

176. Johnson, C. D., and Gibson, J. E. "Singular Solutions in Problems of Optimal Control," IEEE Trans. on Automatic Control, Vol. AC-8, Jan. 1963, p. 4-15.

APPENDIX A

COMPUTATIONAL SOLUTION OF TWO-POINT BOUNDARY VALUE PROBLEMS

As noted repeatedly in this monograph, the solution of an optimization problem by variational methods or the maximum principle leads to the requirement for solving a set of differential equations with two-point boundary conditions. This requirement is indeed one of the main limitations of the classical approach, since the solution of boundary value problems is a formidable computational task. Various solutions have been obtained in special cases for particular problems†, but a reasonably general theory was not available until fairly recently.

The basic idea in the method to be described here is that of "quasilinearization,"(149 in which a nonlinear problem is transformed to a sequence of linear ones by means of a "generalized Newton Raphson operator."(145) As the name implies, the method is closely akin to the Newton Raphson technique for obtaining the roots of transcendental equations by successively replacing the arcs of a curve by its tangents.

Since the solution of the nonlinear problem is reduced to solving its linear equivalent, the latter will be considered first.

A1. THE LINEAR CASE

Given the matrix vector differential equation

$$\dot{y} = A(t)y + h(t) \tag{A1}$$

where y(t) and h(t) are n vectors and A(t) is an nxn matrix. It is required to determine y(t) in the interval (0, t_f) such that the boundary conditions

$$y_j(0) = a_j \qquad j = 1, 2, \ldots, r \tag{A2}$$

$$y_j(t_f) = b_j \qquad j = r+1, \ldots, n \tag{A3}$$

are satisfied.

†Cf. Refs. 64, 77, 110, 113, and 120.

A simple and direct method for obtaining y(t) is the following.(147)

Integrate the homogeneous version of Eq. (A1)

$$\dot{u} = A(t)\,u \tag{A4}$$

(n-r) times. The initial conditions at the m^th time are

$$u_j^{(m)}(0) \begin{array}{ll} = 0 & j \neq r+m \\ = 1 & j = r+m \end{array} \tag{A5}$$

This yields the (n-r) vectors

$$u^{(m)}(t) \;,\quad 0 \lesssim t \lesssim t_f \qquad\qquad m = 1,\, 2,\, \ldots,\, (n-r) \tag{A6}$$

Now integrate the equation

$$\dot{v} = A(t)v + h(t) \tag{A7}$$

using as initial conditions

$$v_j(0) \begin{array}{lll} = a_j \;, & & j = 1,\, 2,\, \ldots,\, r \\ = 0 \;, & & j = (r+1),\, \ldots,\, n \end{array} \tag{A8}$$

Call this solution

$$v(t) \;,\quad 0 \lesssim t \lesssim t_f \tag{A9}$$

The general solution of Eq. (A1) is then given by

$$y_j(t) = \sum_{p=1}^{m} C_p\, u_j^{(p)}(t) + v_j(t) \tag{A10}$$

where the C_p are (scalar) constants determined by solving the (n-r) linear algebraic equations

$$b_j = \sum_{p=1}^{m} C_p\, u_j^{(p)}(t_f) + v_j(t_f) \qquad j = (r+1), \cdots, n \tag{A11}$$

The distinctive feature of this method is that all integrations are performed with <u>initial</u> conditions; it is not necessary to "guess" at final values. The solution is exact (within roundoff error) after a finite number of operations.

A2. THE NONLINEAR CASE

The vector equation to be solved now takes the form

$$\dot{x} = f(x,t) \tag{A12}$$

where x and f are n vectors, and the boundary conditions are given by

$$x_j(0) = a_j \qquad j = 1, 2, \ldots, r \tag{A13}$$

$$x_j(t_f) = b_j \qquad j = (r+1), \ldots, n \tag{A14}$$

Let $x^{(0)}(t)$ be an initial guess for the solution of (A12). Successive improvements are obtained from

$$\dot{x}^{(k+1)} = F^{(k)}\left(x^{(k+1)} - x^{(k)}\right) + f^{(k)} \tag{A15}$$

where $f^{(k)} \equiv f\left(x^{(k)}, t\right)$, and $F^{(k)}$ is the Jacobian matrix whose ijth component is given by

$$F_{ij}^{(k)} = \frac{\partial f_i\left(x^{(k)},t\right)}{\partial x_j} \tag{A16}$$

Eq. (A15) may also be written as

$$\dot{x}^{(k+1)} = F^{(k)}x^{(k+1)} + w^{(k)} \tag{A17}$$

where

$$w^{(k)} = f^{(k)} - F^{(k)}x^{(k)} \qquad \text{(A18}$$

Eq. (A17) is a linear equation, which, together with the boundary conditions, (A13) and (A14), is completely analogous to the system (A1) - (A3). It may therefore be solved by the methods of the previous section.

Note that the conventional Newton Raphson method solves the scalar equation g(x) = 0 via

$$x_{i+1} = x_i - \frac{g(x_i)}{g'(x_i)} \tag{A19}$$

$$g'(x) \equiv \frac{dg}{dx}$$

$$x_0 \equiv \text{initial guess}$$

Writing (A19) in the form

$$0 = g'(x_i)\left[x_{i+1} - x_i\right] + g(x_i) \tag{A20}$$

The analogy with (A15) is evident. The generalized Newton Raphson approach is conceptually identical to the scalar case in that a curve is successively replaced by tangent lines. In other words, the nonlinear problem is replaced by a sequence of linear problems.

A suitable "stopping" criterion is given by

$$\left|x^{(k+1)} - x^{(k)}\right| \lesseqgtr \epsilon \tag{A21}$$

where ϵ is a predetermined error vector.

A relevant question at this point is: "What are the conditions that ensure that this procedure converges to the solution of (A12)?"

It has been shown (149, 165) that sufficient conditions for convergence are

1. $f_j(x,t)$ is strictly convex† for all t in the interval $(0, t_f)$, and
2. $\partial f_i(x,t)/\partial x_j > 0$ when $i \neq j$

†A function $\varphi(x)$ is strictly convex in an interval (a,b) if, for x_1, x_2, λ, with $a < x_1 < x_2 < b$ and $0 < \lambda < 1$, we have

$$\varphi\left[\lambda x_1 - (1-\lambda)\, x_2\right] < \lambda\varphi(x_1) + (1-\lambda)\,\varphi(x_2)$$

If $\varphi''(x)$ exists, then $\varphi(x)$ is strictly convex if $\varphi''(x) > 0$.

We emphasize the word "sufficient," since these conditions may not be necessary. This procedure has been found to converge, for example, when none of the $f_j(x)$ were convex. Each case must therefore be investigated independently and convergence checked by a criterion of the type (A21) unless the aforementioned conditions are satisfied beforehand. It is worthy of note that convergence (when it occurs) is quadratic; that is, the number of correct digits approximately doubles at each iteration. Furthermore, this convergence is monotonic, which means that accurate solutions are obtained with relatively few iterations, even with poor initial guesses.

Thus far, we have assumed that the final time, t_f, was known beforehand. It sometimes happens that t_f is not known initially, and this requires a modified treatment. One approach is described in Ref. 145, which, however, entails numerical differentiation — a procedure to be avoided if possible. A superior method, due to Long,(166) is the following.

Introduce a new independent variable, s, defined by

$$t = as$$

where a is a constant to be determined. The new endpoints are taken as s = 0 and s = 1. Therefore, once a is determined, the value of t_f is given by $t_f = a$.

Suppose that a typical equation of the system (A12) is $\dot{x}_j = f_j(x,t)$. If we write

$$\frac{dx}{ds} = x'$$

then we have

$$x_j' = af_j(x,as)$$

The parameter a is treated as an additional state variable by writing

$$a' = 0$$

Thus the new state vector of the system is of dimension (n+1), and we proceed as before. Note that one makes an initial guess on a $\left(\text{in } x^{(0)}(t)\right)$. This is therefore constant during each iteration, but it varies from one iteration to the next. Thus the determination of t_f becomes an integral part of the quasilinearization procedure.

APPENDIX B

THE VARIATION OPERATOR

Consider the function

$$F = F(\dot{y}, y, t) \tag{B1}$$

where the dot denotes derivative with respect to t, the independent variable. For a fixed value of t, F depends only on y and ẏ.

Suppose now that y is replaced by Y, where

$$Y = y + \epsilon\, \eta(t) \tag{B2}$$

$$\epsilon \equiv \text{a constant}$$

The change $\epsilon\, \eta(t)$ in y (t) is called the variation of y (t) and is denoted by δy;

$$\delta y = \epsilon\, \eta(t) \tag{B3}$$

We may further define the variation of $\dot{y}(t)$ as

$$\delta \dot{y} = \dot{Y} - \dot{y} \tag{B4}$$

which, by virtue of Eq. (B2), becomes

$$\delta \dot{y} = \epsilon\, \dot{\eta} \tag{B5}$$

However, from (B3),

$$\frac{d}{dt}(\delta y) = \epsilon\, \dot{\eta} \tag{B6}$$

Comparing (B5) and (B6), we find

$$\delta\left(\frac{dy}{dt}\right) = \frac{d}{dt}(\delta y) \tag{B7}$$

In other words, the operators δ and d/dt are commutative.

With y replaced by Y of Eq. (B2), the change in F may be expressed as

$$\Delta F = F(\dot{y} + \epsilon \dot{\eta},\ y + \epsilon \eta,\ t) - F(\dot{y},\ y,\ t)$$

Taking a Taylor expansion in which only linear terms are retained, this reduces to

$$\Delta F = \frac{\partial F}{\partial y} \epsilon \eta + \frac{\partial F}{\partial \dot{y}} \epsilon \dot{\eta}$$

After substituting Eqs. (B3) and (B5), the resulting expression is defined as the (first) variation of F:

$$\delta F = \frac{\partial F}{\partial y} \delta y + \frac{\partial F}{\partial \dot{y}} \delta \dot{y} \tag{B8}$$

For a complete analogy with the definition of a differential, one might perhaps anticipate the definition

$$\delta F = \frac{\partial F}{\partial y} \delta y + \frac{\partial F}{\partial \dot{y}} \delta \dot{y} + \frac{\partial F}{\partial t} \delta t$$

However, t is not varied, so that $\delta t \equiv 0$. Hence the analogy is indeed complete.

It is easy to verify from the above definitions that the variation operator obeys the same laws as the differentiation operator. Thus

$$\delta (F_1 F_2) = F_1 \delta F_2 + F_2 \delta F_1 \tag{B9}$$

$$\delta \left(\frac{F_1}{F_2} \right) = \frac{F_2 \delta F_1 - F_1 \delta F_2}{F_2^2} \tag{B10}$$

etc.

The essential distinction between the variation and differential operators is the following. The differential of a function is a first-order approximation to the change in that function along a particular curve. However, the variation of a function is a first-order approximation to the change from curve to curve.

For further details, we refer the reader to standard texts on the calculus of variations.(98, 134)

APPENDIX C

INNER PRODUCT IN FUNCTION SPACE

For any two vectors, a and b, the <u>inner product</u>† is defined by

$$a \cdot b \equiv a^T b = \sum_{j=1}^{n} a_j b_j \tag{C1}$$

while the <u>norm</u> of a vector, a, is defined by

$$\|a\| = (a \cdot a)^{1/2} \tag{C2}$$

Two vectors, a and b, are said to be <u>orthogonal</u> if

$$a \cdot b = 0 \tag{C3}$$

when both a and b are nonzero.

A set of vectors, $\left\{a^{(k)}\right\}$, is termed an <u>orthogonal</u> set if

$$a^{(r)} \cdot a^{(s)} = 0 \quad \text{for } r \neq s \tag{C4}$$

If, in addition, the relation

$$a^{(r)} \cdot a^{(r)} = 1 \quad \text{for all } r \tag{C5}$$

is satisfied, then $\left\{a^{(k)}\right\}$ is called an <u>orthonormal</u> set of vectors.

It may be shown that the inner product of two vectors satisfies the <u>Schwartz inequality</u>

$$\|a \cdot b\| \leq \|a\| \cdot \|b\| \tag{C6}$$

and that for any two vectors, a and b, the following relations hold.

$$\|a\| > 0 \quad \text{for all } a \neq 0 \tag{C7}$$

$$\|a\| = 0 \quad \text{if, and only if,} \quad a = 0 \tag{C8}$$

†Also called scalar or dot product.

$$\|a + b\| \leq \|a\| + \|b\| \tag{C9}$$

$$\|\beta a\| = \beta \|a\|, \quad \beta \equiv \text{scalar} \tag{C10}$$

Furthermore, the gradients of the scalar

$$b^T A a = a^T A^T b$$

where A is an n × n constant matrix, are given by†

$$\frac{\partial}{\partial b}(b^T A a) = A a \tag{C11}$$

$$\frac{\partial}{\partial a}(b^T A a) = A^T b \tag{C12}$$

In particular,

$$\frac{\partial}{\partial a}(a^T A a) = 2 A a \tag{C13}$$

The vector

$$\frac{\partial \gamma}{\partial a} \equiv \begin{bmatrix} \frac{\partial \gamma}{\partial a_1} \\ \vdots \\ \vdots \\ \frac{\partial \gamma}{\partial a_n} \end{bmatrix} \tag{C14}$$

is called the gradient of the scalar, γ.

† The superscript T denotes transpose.

If the vector, a, is a function of a parameter, t, then

$$\frac{d}{dt}\,\|a\| \equiv \|\dot{a}\| = \frac{d}{dt}\,(a^T a)^{1/2}$$

$$= \frac{1}{2}(a^T a)^{-1/2}\sum_{i=1}^{n} 2\, a_i\, \dot{a}_i$$

which reduces to

$$\|\dot{a}\| = \frac{a^T \dot{a}}{\|a\|} \qquad \text{(C15)}$$

Consider now two functions, f (t_i) and g (t_i), which are defined for discrete values of t_i, in an interval, t_1, t_2, ..., t_n. Suppose we define a slightly generalized inner product by

$$\sum_{i=1}^{n} f(t_i)\, g(t_i)\, \Delta t_i$$

where

$$\Delta t_i = t_{i+1} - t_i$$

Then, in the limit

$$\lim_{\substack{n \to \infty \\ \Delta t_i \to 0}} \left[\sum_{i=1}^{n} f(t_i)\, g(t_i)\, \Delta t_i\right] \equiv (f, g) = \int_{t_1}^{t_n} f(t)\, g(t)\, dt \qquad \text{(C16)}$$

The quantity, (f, g), is defined as the inner product of the functions, f (t) and g (t), in function space. Loosely speaking, we say that (C16) gives the inner product of two vectors in an infinite dimensional Euclidean space.

The norm of f(t) is defined by

$$\|f\| = \sqrt{(f, f)} \qquad \text{(C17)}$$

This satisfies a set of relations completely analogous to (C6) - (C10); viz.,

$$\| (f, g) \| \leq \| f \| \cdot \| g \| \qquad \text{(C18)}$$

$$\| f \| > 0 \quad \text{for all } f \neq 0 \qquad \text{(C19)}$$

$$\| f \| = 0 \quad \text{if, and only if,} \quad f = 0 \qquad \text{(C20)}$$

$$\| f + g \| \leq \| f \| + \| g \| \qquad \text{(C21)}$$

$$\| \beta f \| = \beta \| f \| \quad \text{for any scalar } \beta \qquad \text{(C22)}$$

In analogy to (C4), we call a system of functions $\{f_i\}$ orthogonal if for any two functions, f_r and f_s,

$$(f_r, f_s) \equiv \int_{t_1}^{t_n} f_r(t) f_s(t) \, dt = 0 \quad , \quad r \neq s \qquad \text{(C23)}$$

Furthermore, in analogy with (C5), the system of functions $\{f_i\}$ is orthonormal if

$$(f_k, f_k) = 1 \quad \text{for all } k \qquad \text{(C24)}$$

For a more comprehensive account of these ideas, the reader is referred to standard texts. (154, 162, 163)