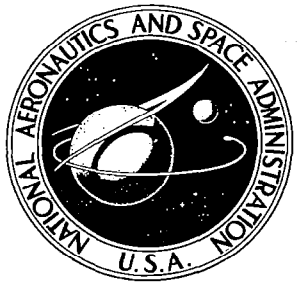
NASA CONTRACTOR REPORT



NASA CR-1012

GUIDANCE, FLIGHT MECHANICS AND TRAJECTORY OPTIMIZATION

Volume XIII - Numerical Optimization Methods

by J. E. McIntyre

Prepared by
NORTH AMERICAN AVIATION, INC.
Downey, Calif.
for George C. Marshall Space Flight Center

NATIONAL AERONAUTICS AND SPACE ADMINISTRATION • WASHINGTON, D. C. • MARCH 1968

GUIDANCE, FLIGHT MECHANICS AND TRAJECTORY OPTIMIZATION

Volume XIII - Numerical Optimization Methods

By J. E. McIntyre

Issued by Originator as Report No. SID 66-1678-5

Prepared under Contract No. NAS 8-11495 by
NORTH AMERICAN AVIATION, INC.
Downey, 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 contract NAS 8-11495 and is one of a series intended to illustrate analytical methods used in the fields of Guidance, Flight Mechanics, and Trajectory Optimization. Derivations, mechanizations and recommended procedures are given. Below is a complete list of the reports in the series.

Volume I	Coordinate Systems and Time Measure
Volume II	Observation Theory and Sensors
Volume III	The Two Body Problem
Volume IV	The Calculus of Variations and Modern Applications
Volume V	State Determination and/or Estimation
Volume VI	The N-Body Problem and Special Perturbation Techniques
Volume VII	The Pontryagin Maximum Principle
Volume VIII	Boost Guidance Equations
Volume IX	General Perturbations Theory
Volume X	Dynamic Programming
Volume XI	Guidance Equations for Orbital Operations
Volume XII	Relative Motion, Guidance Equations for Terminal Rendezvous
Volume XIII	Numerical Optimization Methods
Volume XIV	Entry Guidance Equations
Volume XV	Application of Optimization Techniques
Volume XVI	Mission Constraints and Trajectory Interfaces
Volume XVII	Guidance System Performance Analysis

The work was conducted under the direction of C. D. Baker, J. W. Winch, and D. P. Chandler, Aero-Astro Dynamics Laboratory, George C. Marshall Space Flight Center. The North American program was conducted under the direction of H. A. McCarty and G. E. Townsend.

C O N T E N T S

		Page
1.0	STATEMENT OF THE PROBLEM.	1
2.0	STATE OF THE ART.	2
2.1	Problem Statement and General Considerations	2
2.2	Numerical Methods in Maxima-Minima Theory.	11
2.2.1	One-Dimensional Problem	11
2.2.1.2	Method #2-Newton's Method.	12
2.2.2	-Dimensional Problem	13
2.2.2.1	Method #1-Steepest Descent (Gradient Method)	15
2.2.2.2	Method #2-Newton-Raphson Method.	17
2.2.2.3	Modified Newton-Raphson Method	18
2.2.3	-Dimensional Problem with Constraints.	20
2.2.3.1	Method #1 - Steepest Descent	22
2.2.3.2	Method #2 - Newton-Raphson Technique	24
2.2.4	Discussion.	27
2.3	Numerical Solution of the One-Dimensional Lagrange Problem .	28
2.3.1	Steepest Descent (Gradient) Technique	32
2.3.2	Quasilinearization.	38
2.3.3	Neighboring Extremal.	41
2.3.4	Discussion of Results	45
2.4	Classical Mayer Problem.	47
2.4.1	Quadratic Cost Problem.	49
2.4.2	Steepest Descent.	54
2.4.3	Neighboring Extremal.	64
2.4.4	Quasilinearization.	72
2.4.5	Discussion.	78
2.5	Nonclassical Mayer Problem	80
2.5.1	Problem Statement and Necessary Conditions.	81
2.5.2	Neighboring Extremal Technique.	84
2.5.3	Steepest Descent.	91
2.5.4	Quasilinearization.	96
2.5.5	Final Time Open	100
3.0	RECOMMENDED PROCEDURES.	103
4.0	REFERENCES.	104

1.0 STATEMENT OF THE PROBLEM

A problem of current interest in the area of flight mechanics concerns the selection of a trajectory along which a particular vehicle should fly in order to accomplish its mission in some sort of best fashion. Control over the vehicle's trajectory is achieved through control over some of the forces that are acting on the vehicle; for example, the aerodynamic forces in the case of a conventional aircraft, or the thrusting forces in the case of a space vehicle. What is involved, then, in the selection of a "best" trajectory, is the determination of the optimum time history of the controllable forces.

The first step in the solution of such a problem involves the determination of certain conditions which a trajectory must satisfy in order to be optimal. These conditions are expressible in differential equation form and can be developed by applying to the problem any one of several standard mathematical techniques such as the Maximum Principle (Ref. (1)), Dynamic Programming (Ref. (2)) or the Calculus of Variations (Ref. (3)). The simultaneous solution of these differential equations together with the equations governing the motion of the vehicle constitutes the solution to the problem. In most cases of interest this solution must be developed on a digital computer by means of some iterative numerical technique.

The mathematical procedures to be followed in formulating an optimization problem are reasonably straight forward and have been documented in three previous monographs in this series. (Refs. (4) to (6)). This monograph is concerned with the second step in the optimization process, that of generating numerical solutions. Specifically, the three fundamental numerical procedures, neighboring external, steepest-descent and quasilinearization, are presented and their application to both classical and non-classical problems indicated. Only a passing reference is made to any of the other techniques since they all represent minor variations or extensions of the three fundamental techniques - extensions which may work effectively in some cases and ineffectively in others.

The numerical procedures used in optimization theory are intimately related to those employed in the solution of ordinary maxima-minima problems and in the sections which follow some emphasis is placed on establishing this relationship. The analysis begins with a review of two numerical methods for locating the minimum value of a function. An extrapolation of these methods into a function space leads to the three fundamental techniques previously mentioned for locating the minimum value of a functional. These techniques are applied first to the classical problem of Lagrange and then to the classical and non-classical versions of the problem of Mayer arising in modern trajectory and control applications. Some general remarks are made concerning the rate and range of convergence of each technique along with the relative ease or difficulty involved in a computer mechanization.

2.0 STATE OF THE ART

2.1 PROBLEM STATEMENT AND GENERAL CONSIDERATIONS

Most optimization problems encountered in trajectory analysis and control theory can be cast in the following form: Given the dynamical system

$$\dot{X} = f(X, t, u) \iff \begin{bmatrix} \dot{X}_1 \\ \dot{X}_2 \\ \vdots \\ \dot{X}_n \end{bmatrix} = \begin{bmatrix} f_1(X, t, U) \\ f_2(X, t, U) \\ \vdots \\ f_n(X, t, u) \end{bmatrix} \tag{2.1.1}$$

where x is an n dimensional state vector, u is a r dimensional control vector and f is an n dimensional vector function, determine the control action u from the admissible control set U which drives the system from the specified initial state

$$X = X_o ; \; t = t_o \iff \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix} = \begin{bmatrix} X_{1o} \\ X_{2o} \\ \vdots \\ X_{no} \end{bmatrix} ; \; t = t_o \tag{2.1.2}$$

to the terminal set

$$\psi(X_f) = 0 ; \; t = t_f \iff \begin{bmatrix} \psi_1(X) \\ \psi_2(X) \\ \vdots \\ \psi_M(X) \end{bmatrix} = 0 ; \; t = t_f \tag{2.1.3}$$

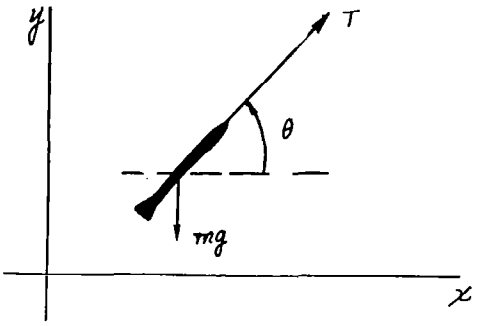
(where ψ is an M dimensional vector function) while at the same time minimizing a scalar function of the terminal state. This scalar function will be denoted

$$J = \phi(X_f) = \text{MINIMUM} \tag{2.1.4}$$

The preceding problem takes the form of the classical Mayer problem in the Calculus of Variations (Ref. (7)). Note that equations (2.1.1) to (2.1.4) have been written both in vector and scalar form. In the paragraphs to follow both forms of these equations will be used.

As an example of this formulation, consider the problem of maneuvering a vehicle over a flat earth having no atmosphere. The governing equations in this case (see sketch to right) take the form

$$\left.\begin{aligned} \ddot{x} &= \frac{T}{m}\cos\theta \\ \ddot{y} &= \frac{T}{m}\sin\theta - g \\ \dot{m} &= \frac{-T}{V} \end{aligned}\right\} \qquad (2.1.5)$$



Sketch (1)

where T is the thrust magnitude, V the exhaust velocity and θ the steering angle. Suppose that the position, velocity and mass of the vehicle are specified initially by

$$\begin{aligned} x &= x_o \\ y &= y_o \\ \dot{x} &= \dot{x}_o \\ \dot{y} &= \dot{y}_o \\ m &= m_o \end{aligned} \qquad AT \quad t = t_o \qquad (2.1.6)$$

and that at the terminal point only the magnitudes position and velocity magnitude are specified

$$\begin{aligned} x &= x_f \\ y &= y_f \\ \dot{x}^2 + \dot{y}^2 &= v_f^2 \end{aligned} \qquad AT \quad t = t_f \qquad (2.1.7)$$

where the final time itself, t_f , is not specified. The problem is to determine the steering angle and thrust time history so that the fuel expended during the flight is a minimum.

To place this problem in the Mayer form as indicated by Eqs. (2.1.1) to (2.1.4), define the new variables x_1, x_2, x_3, x_4, x_5 and u_1 by

$$\begin{aligned} X_1 &= X \\ X_2 &= Y \\ X_3 &= \dot{X} \\ X_4 &= \dot{Y} \\ X_5 &= m \\ U_1 &= \theta \end{aligned} \tag{2.1.8}$$

Also, assume that the thrust magnitude T can vary from zero to some maximum value T_{MAX}, and let

$$T = T_{MAX}\, U_2$$

where u_2 denotes the throttle setting which can vary between zero and unity, as the thrusting engine goes from the "full off" to the "full on" position. Using these definitions, the governing equations become

$$\begin{aligned} \dot{X}_1 &= X_3 \\ \dot{X}_2 &= X_4 \\ \dot{X}_3 &= \frac{T_{MAX}\, U_2}{X_5} \cos U_1 \\ \dot{X}_4 &= \frac{T_{MAX}\, U_2}{X_5} \sin U_1 - g \\ \dot{X}_5 &= \frac{-T_{MAX}\, U_2}{V} \end{aligned} \tag{2.1.9}$$

with the initial conditions

$$\begin{aligned} X_1 &= X_{10} \\ X_2 &= X_{20} \\ X_3 &= X_{30} \\ X_4 &= X_{40} \\ X_5 &= X_{50} \end{aligned} \qquad \text{AT} \quad t = t_0 \tag{2.1.10}$$

and the terminal conditions

$$\begin{aligned} X_1 &= X_{1f} & & & \psi_1 &= X_1 - X_{1f} = 0 \\ X_2 &= X_{2f} & \text{AT} \quad t = t_f \quad &\Longleftrightarrow & \psi_2 &= X_2 - X_{2f} = 0 \qquad t = t_f \\ X_3^2 + X_4^2 &= v_f^2 & & & \psi_3 &= X_3^2 + X_4^2 - v_f^2 = 0 \end{aligned}$$

(2.1.11)

The fuel expended during the flight will be minimized if the final mass is maximized; or alternately, if the negative of the final mass is minimized. This leads to the performance criterion

$$J = -X_5 \quad = \text{MINIMUM}$$

(2.1.12)

where the function ϕ in Eq. (2.1.4) is equal to $-x_5$. The minimization is accomplished through the appropriate selection of the steering angle u_1 and throttle setting u_2 . Since u_1 can take any value at all while u_2 must satisfy the inequality condition

$$0 \leq u_2 \leq 1 ,$$

(2.1.13)

if follows that the admissible control set U in which the control vector must lie is given by

$$U \Longleftrightarrow \begin{cases} 0 \leq u_2 \leq 1 \\ u_1 \quad \text{ARBITRARY} \end{cases}$$

(2.1.14)

The development of the optimizing conditions for the Mayer problem given in Eqs. (2.1.1) to (2.1.4) is most easily accomplished by applying the Pontryagin Maximum Principle (See Refs. (1) or (5)). This requires the introduction of an n dimensional adjoint or multiplier vector p which satisfied the differential equation

$$\dot{P} = -\frac{\partial H}{\partial X} \iff \begin{bmatrix} \dot{P}_1 \\ \dot{P}_2 \\ \vdots \\ \dot{P}_n \end{bmatrix} = \begin{bmatrix} -\frac{\partial H}{\partial X_1} \\ -\frac{\partial H}{\partial X_2} \\ \vdots \\ -\frac{\partial H}{\partial X_n} \end{bmatrix} \tag{2.1.15}$$

where H is the variational Hamiltonian given by

$$H = P^T f = \sum_{i=1}^{n} P_i f_i (X, t, u) \tag{2.1.16}$$

If the final time t_f is not explicitly specified, then the additional terminal condition

$$H = \sum_{i=1}^{n} P_i f_i (x, t, u) = 0 \tag{2.1.17a}$$

must hold.

An n dimensional constant vector P is now introduced and selected to satisfy the terminal conditions

$$P + \left(\frac{\partial \psi}{\partial X}\right)^T \mu + \frac{\partial \phi}{\partial X} = 0 ; \; t = t_f \iff P_i + \sum_{j=1}^{M} \mu_j \frac{\partial \psi_j}{\partial x_i} + \frac{\partial \phi}{\partial x_i} = 0 \qquad i = 1, n \tag{2.1.17b}$$

The optimal control u_{opt} is that control in the admissible set U which maximizes the Hamiltonian H at each instant of time along the solution trajectory; that is, u_{opt} must lie in U and satisfy

$$H(X, t, P, u_{OPT}) \geq H(X, t, P, \bar{u}) ; \qquad t \in (t_o, t_f) \tag{2.1.18}$$

where $\bar{u}$ denotes any other control vector lying in the set U

By way of illustration, the application of the maximum Principle to the rocket problem of Eqs. (2.1.5) to (2.1.14) provides the variational Hamiltonian

$$H = P_1 X_3 + P_2 X_4 + P_3\left(\frac{T_{MAX}}{X_5}\ u_2 \cos u_1\right) + P_4\left(\frac{T_{MAX}}{X_5}\ u_2 \sin u_1 - g\right) + P_5\left(-\frac{T_{MAX}\ u_2}{V}\right) \tag{2.1.19}$$

where the p vector satisfies the differential system

$$\begin{aligned}
\dot{P}_1 &= 0 \\
\dot{P}_2 &= 0 \\
\dot{P}_3 &= -P_1 \\
\dot{P}_4 &= -P_2 \\
\dot{P}_5 &= \frac{P_3\ T_{MAX}\ u_2 \cos u_1}{X_5^2} + \frac{P_4\ T_{MAX}\ u_2 \sin u_1}{X_5^2}
\end{aligned} \tag{2.1.20}$$

Using Eq. (2.1.11) the terminal conditions on the p vector corresponding to Eq. (2.1.17b) are

$$\begin{aligned}
P_1 &= \mu_1 \\
P_2 &= \mu_2 \\
P_3 + 2\mu_3 X_3 &= 0 \qquad AT \quad t = t_f \\
P_4 + 2\mu_3 X_4 &= 0 \\
P_5 &= 1
\end{aligned} \tag{2.1.21}$$

where the μ_i are constants to be selected so the boundary conditions in Eq. (2.1.11) are satisfied. Also, since the final time t_f is not specified. Eq. (2.1.17a) must hold, i.e.,

$$H = \sum_{i=1}^{n} P_i f_i (X, t, u) = 0 \qquad AT \quad t = t_f \tag{2.1.22}$$

Finally, the optimal control is to be selected to maximize H . Letting

$$\frac{P_3}{\sqrt{P_3^2 + P_4^2}} = \cos z \quad ; \quad \frac{P_4}{\sqrt{P_3^2 + P_4^2}} = \sin z$$

the H function of (2.1.19) can be written as

$$H = \sqrt{P_3^2 + P_4^2}\ \frac{T_{MAX}\ u_2}{X_5}\left\{\cos(u_1 - Z) - \frac{P_5 X_5}{V\sqrt{P_3^2+P_4^2}}\right\} + \text{TERMS NOT INVOLVING } u_1 \text{ OR } u_2 \tag{2.1.23}$$

Hence, to maximize H, $\cos(u_1 - Z) = 1$ AND

$$\cos u_1 = \frac{P_3}{\sqrt{P_3^2 + P_4^2}}\ ; \qquad \sin u_1 = \frac{P_4}{\sqrt{P_3^2 + P_4^2}} \tag{2.1.24}$$

Thus Eq. (2.1.23) reduces to

$$H = \frac{T_{MAX}\ u_2}{X_5}\left\{\sqrt{P_3^2 + P_4^2} - \frac{P_5 X_5}{V}\right\} + \text{terms not involving } u_1 \text{ or } u_2$$

but, for u_2 to maximize H and satisfy the inequality in (2.1.13) it follows that

$$\begin{aligned} u_2 &= 0\ ; \quad \Theta < 0 \\ u_2 &= 1\ ; \quad \Theta > 0 \end{aligned} \tag{2.1.25}$$

where

$$\Theta = \sqrt{P_3^2 + P_4^2} - \frac{P_5 X_5}{V} \tag{2.1.26}$$

Collecting results for the rocket problem, it follows that the optimal trajectory is determined from the simultaneous solution of the state and adjoint equations, Eqs. (2.1.9) and (2.1.20), with the control vector, u_1 and u_2, satisfying Eqs. (2.1.24) and (2.1.25). Note that the state and adjoint equations constitute a tenth-order differential system (ten first-order equations). Hence, ten boundary conditions must be specified along with the time at which the conditions are to hold. Five of these boundary conditions are given by the specification of the initial state and time in Eqs. (2.1.10). Three more come from the state terminal constraints of Eqs. (2.1.11). Eq. (2.1.21) provides two more conditions*

$$\begin{aligned} X_3 P_4 - X_4 P_3 &= 0 \qquad \text{AT} \quad t = t_f \\ P_5 &= 1 \end{aligned} \tag{2.1.27}$$

* Since the μ_i are not specified explicitly, the five conditions in (2.1.21) reduce to the two conditions in (2.1.27).

But, since the final time itself is not specified, one additional boundary condition is needed before a solution can be generated; this condition comes from Eq. (2.1.22) which requires the Hamiltonian to vanish at the terminal point.

From the previous discussion, it follows that, in the rocket problem, the specified state boundary conditions together with the boundary conditions on the p vector (the transversality conditions) are just sufficient in number to generate a unique solution to the governing differential equations (Eqs. (2.1.9) and (2.1.20)). Note that the boundary conditions are specified at two distinct instants of time; that is, some conditions are to hold at the initial point $t = t_0$ and others are to hold at the final point $t = t_f$. However, to generate a numerical solution on a computer, all boundary data must be given at one point in time. This requirement is simply a result of the manner in which numerical integration must be performed. That is, a complete set of boundary data at some point t_1, and a finite difference approximation to the differential equations are used to generate a complete set of data at a neighboring point, say $t_1 + \Delta t$ where Δt is small. Hence, to construct a solution to the rocket problem, the value of the p vector at time t_0 (i.e., guess p_{1_0} , p_{2_0} , p_{3_0} , p_{4_0} & p_{5_0}) and the terminal time t_f must be guessed, then the equations must be integrated numerically to the guessed terminal time to determine if the six conditions incorporated in Eqs. (2.1.11), (2.1.22) and (2.1.27) are satisfied. In general, it is not possible to make the correct guess on the first try. However, it is hoped that some iterative process of correcting and recorrecting the initial guesses would eventually converge to the desired optimal solution.

A similar situation holds in most all optimization problems, that is a set of differential equations exists for which boundary data are given at two distinct times. Hence, some iterative procedure must be used to construct a solution. The technique outlined above, in which the unknown quantities are selected at one end of the trajectory to satisfy known conditions at the other end, constitutes a possible iteration scheme, and one that is frequently used in practice. However, there are several others.

In the general case, the solution to an optimization problem must satisfy five conditions:

(1) the differential state equations

$$\dot{x} = f(x, t, u) \tag{2.1.28}$$

(2) the differential adjoint equations

$$\dot{P} = -\frac{\partial H}{\partial x} \tag{2.1.29}$$

(3) the optimization condition

$$H(x, t, P, u_{OPT}) \geq H(x, t, P, \bar{u}), \; u_{OPT}, \; \bar{u} \in U \tag{2.1.30}$$

(4) state boundary conditions

$$X = X_o \quad AT \quad t = t_o \; ; \quad \psi(x, t) = 0 \quad AT \quad t = t_f \tag{2.1.31}$$

(5) transversality conditions

$$P + \left(\frac{\partial \psi}{\partial X}\right)^T \mu + \left(\frac{\partial \phi}{\partial X}\right) \qquad AT \quad t = t_f$$

$$^* H = P^T f = 0 \tag{2.1.32}$$

To numerically construct a solution to an optimization problem, a solution is selected which satisfies only some of the above conditions; this estimate is then iteratively corrected in a direction which tends to bring in the other conditions. What distinguishes one iteration process from another is the conditions which are satisfied at the start as opposed to those which are brought in. The three iterative processes to be discussed in this report are characterized as follows:

a) Neighboring Extremal - each successive iteration satisfies conditions (1), (2), and (3) above, but does not satisfy (4) or (5). (This is essentially the technique in which unknown boundary conditions are guessed at one end of the trajectory.)

b) Steepest Descent - each successive iteration satisfies conditions (1), (2), (4) and (5), but does not satisfy (3).

c) Quasilinearization - each successive iteration satisfies conditions (4) and (5), but does not satisfy (1) or (2). Satisfaction of condition (3) is arbitrary.

Before turning to a detailed consideration of these three techniques, it is well to consider some other optimization problems which are of a considerably more elementary nature, but which require numerical solution by iterative techniques very similar to those above. The first such problem is the maxima-minima problem from the ordinary differential calculus; this problem is treated in the next section. The succeeding section deals with the one-dimensional Lagrange Problem. The numerical methods used in these elementary problems are identical in concept to those used in the Mayer problem; only the algebra and the actual numerical computations are more complex.

---

* This condition holds only if the final time is not specified explicitly.

2.2 NUMERICAL METHODS IN MAXIMA-MINIMA THEORY

2.2.1 One-Dimensional Problem

The one-dimensional minimization problem consists of determining the value of the scalar x which minimized a certain function, say

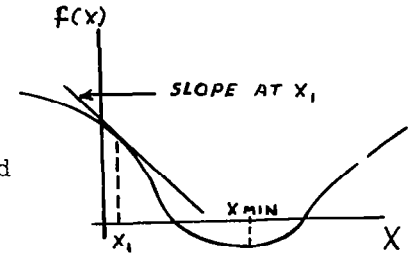
$$f(x) = MINIMUM \tag{2.2.1}$$

The term "one dimensional" refers to the condition that x is a scalar (a one-dimensional vector).

To solve such a problem either of two methods can be employed; both methods are most easily described by means of a graph.

2.2.1.1 Method #1 - Steepest Descent (Gradient Method)

Assume that the function f has the plot shown to the right and that the first guess (first iteration) at the minimizing point is x_1. The value $f(x_1)$ and the derivative (df/dx) are then evaluated. A second value, $x_2 = x_2 + \Delta x$, is selected where Δx is chosen in the direction in which $f(x)$ is decreasing, as determined from the sign of the derivative df/dx. In equation form,



Sketch (2)

$$X_2 = X_1 + \Delta X = X_1 - k_1 \frac{df}{dX} \tag{2.2.2}$$

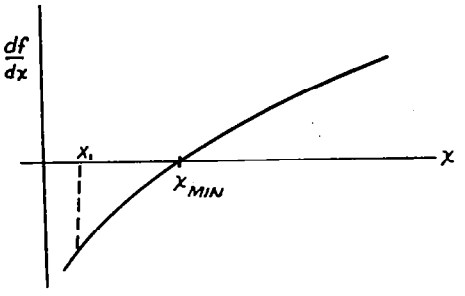
where k is some small positive constant which is arbitrarily determined. But to first order in Δx

$$f(X_2) \cong f(X_1) + \frac{df}{dX}(X - X_1) = f(X_1) - k_1\left(\frac{df}{dX_1}\right)^2 \tag{2.2.3}$$

Now, since k_1 is positive, $f(x_2)$ will be smaller than $f(x_1)$ provided k_1 is sufficiently small. Thus, by appropriate choice of the constant k_1, the process will converge towards the optimum point x_{MIN} characterized by the condition

$$\left(\frac{df}{dX}\right)_{X_{MIN}} = 0 \tag{2.2.4}$$

2.2.1.2 Method #2 - Newton's Method



Sketch (3)

The second numerical approach involves the iterative adjustment of X until the minimizing condition $df/dx = 0$ is satisfied. Assume the derivative df/dx has the plot shown to the right and that X_1 is again the first guess at the minimizing point. The functions $\left(\frac{df}{dX}\right)_{X_1}$ and $\left(\frac{d^2f}{dX^2}\right)_{X_1}$ are then evaluated, and to first order in $\Delta X = X_2 - X_1$

$$\left(\frac{df}{dX}\right)_{X_2} = \left(\frac{df}{dX}\right)_{X_1} + \left(\frac{d^2f}{dX^2}\right)_{X_1} (X_2 - X_1) \qquad (2.2.5)$$

For the second iteration, X_2 is selected as that value of X for which the first-order approximation to $\frac{df}{dX}$ given in Eq. (2.2.5) is zero. Hence,

$$X_2 = X_1 + \Delta X = X_1 - \frac{\left(\frac{df}{dX}\right)_{X_1}}{\left(\frac{d^2f}{dX^2}\right)_{X_1}} \qquad (2.2.6)$$

Thus, to second order in ΔX ,

$$f(X_2) \cong f(X_1) + \frac{df}{dX}(X - X_1) + \frac{1}{2}\left(\frac{d^2f}{dX^2}\right)(X - X_1)^2$$

$$= f(X_1) - \frac{1}{2}\frac{\left(\frac{df}{dX}\right)^2}{\frac{d^2f}{dX^2}} \qquad (2.2.7)$$

Since $\frac{d^2f}{dX^2} \geq 0$ at the minimizing point, the value of $f(X_2)$ will be less than the value of $f(X_1)$ provided ΔX in Eq. (2.2.6) is sufficiently small.

The steepest descent approach is a first-order technique in that only the first-order terms in a Taylor series are used to adjust the iterated quantity X (see Eq. (2.2.3)). In this regard only one addition quantity, the derivative df/dx, need be evaluated at each point. The constant k, which controls the magnitude of the correction ΔX (the slope df/dx controls the direction) is adjusted and readjusted at each step in the iteration process.

In contrast, Newton's method is second order and requires the calculation of both the first and second derivatives at each iteration. However, the correction (both magnitude and direction) is completely determined by the iteration process itself (see Eq. (2.2.7)) and no constant k_1 need be guessed.

As a general rule, the one-dimensional minimization problem is so easy to solve that little consideration need be given to the particular numerical method employed. Almost any method will suffice including a direct search procedure. This situation changes drastically as higher dimensional problems are considered. While the techniques employed are essentially the same, much more computation is involved and considerable attention must be paid to the rate and range of convergence in order to effect a solution.

2.2.2 n -Dimensional Problem

In the n -dimensional problem, the value of x which minimizes a certain function must once again be determined, i.e.,

$$f(X) = \text{MINIMUM} \tag{2.2.8}$$

However, in this case, x is an n -dimensional vector

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \tag{2.2.9}$$

with

$$f(X) = f(X_1, X_2, \cdots X_n) \tag{2.2.10}$$

Under the assumption that $f(X)$ is a sufficiently smooth function (at the very least, the third derivatives of f with respect to all its arguments must be bounded) it is well known that at the minimum point the n equations

$$\frac{\partial f}{\partial X} = \begin{bmatrix} \frac{\partial f}{\partial X_1} \\ \frac{\partial f}{\partial X_2} \\ \vdots \\ \frac{\partial f}{\partial X_n} \end{bmatrix} = 0 \tag{2.2.11}$$

are satisfied and that the matrix of second derivatives must be positive semi-definite, i.e.,

$$\frac{\partial^2 f}{\partial X^2} = \begin{bmatrix} \frac{\partial^2 f}{\partial X_1^2} & \frac{\partial^2 f}{\partial X_1 \partial X_2} & \cdots & \frac{\partial^2 f}{\partial X_1 \partial X_n} \\ \frac{\partial^2 f}{\partial X_1 \partial X_2} & \frac{\partial^2 f}{\partial X_2^2} & \cdots & \frac{\partial^2 f}{\partial X_2 \partial X_n} \\ \vdots & & & \\ \frac{\partial^2 f}{\partial X_n \partial X_1} & \frac{\partial^2 f}{\partial X_n \partial X_2} & \cdots & \frac{\partial^2 f}{\partial X_n^2} \end{bmatrix} = \text{POSITIVE SEMI-DEFINITE} \tag{2.2.12}$$

A direct solution to this problem can be achieved only if Eq. (2.2.11) can be solved directly (as opposed to iteratively). This, in turn, is possibly only in one very special case.

Assume that the function takes the form

$$f = x^T A x + x^T b + c \iff f = \sum_{i,j}^{n} a_{ij} x_i x_j + \sum_{i}^{n} b_i x_i + c \qquad (2.2.13)$$

where A is the $n \times n$ symmetric matrix,

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & & a_{nn} \end{bmatrix} \qquad (2.2.14)$$

b is the constant vector

$$b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \qquad (2.2.15)$$

and c is a scalar. In other words f is the sum of a quadratic function in x, a linear function in x and a constant. For convenience, such a function will be referred to as a "quadratic cost" function. In this case the minimizing conditions of Eqs. (2.2.11) and (2.2.12) become

$$Ax + b + o \iff \sum_{j=1}^{n} a_{ij} x_j + b_i = 0; \quad i = 1, n \qquad (2.2.16)$$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} = \text{POSITIVE DEFINITE} \qquad (2.2.17)$$

Note that Eq. (2.2.16) is linear in x and hence, it can be solved directly to yield the minimizing point

$$x_{MIN} = \begin{bmatrix} x_{1\,MIN} \\ x_{2\,MIN} \\ \vdots \\ x_{n\,MIN} \end{bmatrix} = -A^{-1} b \qquad (2.2.18)$$

In the general case in which f is not a quadratic cost function, the minimizing condition of Eq. (2.2.11) yields a set of nonlinear transendental equations to which no direct solution is possible. Hence, some iterative techniques must be employed to determine the minimizing solution. The basis of the two iterative techniques presented below consists in approximating the function to be minimized by a series of quadratic cost functions, in which case the minimum point is found by solving a series of linear problems of the form of Eq. (2.2.18).

2.2.2.1 Method #1 - Steepest Descent (Gradient Method)

Let x_1 denote the first guess at the minimizing point where

$$X_1 = \begin{vmatrix} X_{11} \\ X_{21} \\ \vdots \\ X_{n1} \end{vmatrix}$$

and let $f(X_1)$ denote the corresponding value of the function to be minimized. Now expand f in a Taylor series about x_1 and truncate the series after the first term; that is

$$f(X) \cong f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) \iff f(X) \cong f(X_1) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} (x_i - x_{i_1}) \tag{2.2.19}$$

where the superscript T denotes the transpose. The problem is now to find x_2, the second iterate, such that

$$f(X_2) < f(X_1) \tag{2.2.20}$$

Hence, if X_2 is required to be reasonably close to X_1, the approximation given in Eq. (2.2.19) can be used to develop a value of X_2 for which the inequality in (2.2.20) will be satisfied. Specifically, for $|X_2 - X_1|$ small, the best result (the smallest value for $f(X_2)$) would be achieved by selecting X_2 to minimize the approximate expression in (2.2.19).

Putting this idea into mathematical form, it follows that the quantity X_2 is to be selected to minimize the expression

$$J = f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) = f(X_1) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} (x_i - x_{i_1}) \tag{2.2.21}$$

subject to the constraint condition

$$(X - X_1)^T (X - X_1) = k_1^2 \iff \sum_{i=1}^{n} (x_i - x_{i_1})^2 = k_1^2 \tag{2.2.22}$$

where k_1 is some small positive constant. This new problem is easily solved by adjoining the constraint in (2.2.22) to the function to be minimized by means of a Lagrange multiplier λ ; that is, the function $\bar{J}$ is formed where

$$\bar{J} = f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T_{X_1} (X - X_1) + \lambda \left\{ (X - X_1)^T (X - X_1) - k_1^2 \right\} \qquad (2.2.23)$$

and where the multiplier λ is chosen so that (2.2.22) is satisfied. Minimizing $\bar{J}$ then yields the desired results; this minimization is accomplished by setting the derivative $\frac{\partial \bar{J}}{\partial X}$ to zero. That is,

$$\left(\frac{\partial f}{\partial X}\right) + 2\lambda (X - X_1) = 0 \qquad (2.2.24)$$

But, the second derivative condition requires

$$\left(\frac{\partial^2 \bar{J}}{\partial X^2}\right) = \begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots & \lambda \end{bmatrix} = \text{POSITIVE DEFINITE} \qquad (2.2.25)$$

from which it follows that λ is positive. Thus, substituting (2.2.24) into the constraint condition of (2.2.22) yields the value of λ as

$$\lambda = \frac{1}{2k_1} \sqrt{\left(\frac{\partial f}{\partial X}\right)^T \left(\frac{\partial f}{\partial X}\right)} = \frac{1}{2k_1} \sqrt{\sum_i^n \left(\frac{\partial f}{\partial X_i}\right)^2} \qquad (2.2.26)$$

and combining Eqs. (2.2.24) and (2.2.26) provides the new value of x as

$$X_2 = X_1 - \frac{k_1 \frac{\partial f}{\partial X}}{\sqrt{\left(\frac{\partial f}{\partial X}\right)^T \left(\frac{\partial f}{\partial X}\right)}} \qquad (2.2.27)$$

Several observations regarding this iterative process are in order. The function which was minimized to provide the value of x_2 (the quantity $\bar{J}$ is (2.2.23)) is a quadratic cost function and it is this condition which allows for the analytical determination of x_2 as given in Eq. (2.2.27). Similarly, each successive iterate of x is determined by minimizing a quadratic cost function. Note also, that the method is the n-dimensional analog of the 1-dimensional steepest descent technique; this fact can be seen by comparing Eq. (2.2.2) with Eq. (2.2.27). In both cases the magnitude of correction (the quantity k_1 in (2.2.27)) is not determined by the process, but must be selected by the user. However, substituting Eq. (2.2.27) into (2.2.19) provides, to first order,

$$f(X_2) = f(X_1) - k_1 \sqrt{\left(\frac{\partial f}{\partial X}\right)^T \left(\frac{\partial f}{\partial X}\right)} \qquad (2.2.28)$$

from which it follows that $f(X_2)$ will be less than $f(X_1)$ provided only that ℓ_1 is sufficiently small.

2.2.2.2 Method #2 - Newton-Raphson Method

As in the method of steepest descent, let x_1 denote the initial value of x and expand $f(x)$ in a Taylor series about x_1. This time, however, the second terms in the expansion are retained providing

$$f(X) \cong f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) + \frac{1}{2}(X-X_1)^T \left(\frac{\partial^2 f}{\partial X^2}\right)(X-X_1)$$

$$= f(X_1) + \sum_i \frac{\partial f}{\partial x_i}(x_i - x_{i_1}) + \frac{1}{2}\sum_{i,j} \frac{\partial^2 f}{\partial x_i \partial x_j}(x_i - x_{i_1})(x_j - x_{j_1}) \tag{2.2.29}$$

Once again the quantity x_2 is to be selected so that

$$f(X_2) < f(X_1) \tag{2.2.30}$$

If x_1 is sufficiently close to its minimum value, then the approximate expression in (2.2.29) can be used to determine an x_2 satisfying (2.2.30). Specifically, x_2 is selected to minimize

$$J = f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) + \frac{1}{2}(X-X_1)^T \left(\frac{\partial^2 f}{\partial X^2}\right)(X - X_1)$$

Since J is again a quadratic cost function, setting the derivative to zero provides the linear system

$$\left(\frac{\partial f}{\partial X}\right) + \left(\frac{\partial^2 f}{\partial X^2}\right)(X - X_1) = 0$$

which has the solution

$$X_2 = X_1 - \left(\frac{\partial^2 f}{\partial X^2}\right)^{-1} \frac{\partial f}{\partial X} \tag{2.2.31}$$

Thus, if x is a 1-dimension vector (a scalar) then Eq. (2.2.31) reduces to the 1-dimensional result given in Eq. (2.2.6). Unlike the steepest descent method, both first and second derivatives are needed at each point in the iteration. However, the magnitude of correction is not guessed, but comes directly from the iteration.

In regard to convergence of the Newton-Raphson method, it can be shown, as in the 1-dimensional case, that the error decreases quadratically with

$$\left|X_n - X_{n-1}\right| \sim \left|X_{n-1} - X_{n-2}\right|^2 \tag{2.2.32}$$

Also, sufficient conditions for convergence are easily developed (see Ref. (9)). In contrast, no general statements regarding the convergence of the steepest-descent technique can be made since the convergence depends on the manner in which the magnitude of correction (k_1 in Eq. (2.2.27)) is determined at each iteration.

The effectiveness of the two techniques depends on both the problem (the functional form of $f(x)$) and the starting point. If the first guess on x is "close" to the optimal point x_{MIN} then the Newton-Raphson approach will usually provide the most rapid convergence since it yields both the magnitude and direction of correction. If the starting point is not "close," then the second-order technique will diverge and it is necessary to resort to the first-order gradient method. In this sense, the term "close" might be arbitrarily defined as any point from which the second-order theory will converge. Hence, it appears that some combination of the two techniques will provide the best results.

In most problems in which the dimension n is three or larger, the selection of the constant k_1 in the steepest descent procedure proves very difficult. Inevitably, some estimation or calculation of the second derivatives, $\partial^2 f/\partial x^2$, are needed in order to appropriately adjust the correction magnitude. For this reason a modified version of the Newton-Raphson technique is frequently used in which the magnitude of correction is guessed, but second derivatives are used in the calculation of the direction of correction. This modification extends the range of convergence to points that are "not close" to the minimum point.

2.2.2.3 Modified Newton-Raphson Method

Let x_1 denote the first guess at the minimizing point and expand f in the truncated Taylor series as

$$f(X) \cong f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) + \frac{1}{2} (X - X_1)^T \left(\frac{\partial^2 f}{\partial X^2}\right) (X - X_1) \tag{2.2.33}$$

Now x_2 is to be selected so that this approximation of f is a minimum, but subject to a restriction on the magnitude of correction; that is, the quantity $|X_2 - X_1|$ is to be constrained. This constraint can take any one of several forms. For example,

$$\frac{(X - X_1)^T (X - X_1)}{2} = k_2^2 \tag{2.2.34A}$$

where k_2 is some small positive constant; or alternately

$$\frac{(X - X_1)^T A (X - X_1)}{2} = k_2^2 \tag{2.2.34B}$$

where A is any $n \times n$ positive definite symmetric matrix*. It is algebraically convenient to use the constraint in (2.2.34B) with A set equal to the second derivative matrix $\partial^2 f/\partial x_2$. If the point x_1 is not too far removed from the minimizing point, x_{MIN}, then $(\partial^2 f/\partial x^2)$ will be positive definite and Eq. (2.2.34B)

* The quantity $(X - X_1)^T A (X - X_1) > 0$ for $(X - X_1) \neq 0$

(with A replaced by $(\partial^2 f/\partial x^2)$) will serve as a proper magnitude constraint. Thus, the quantity

$$J = f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) + \frac{1}{2} (X - X_1)^T \left(\frac{\partial^2 f}{\partial X^2}\right) (X - X_1) \tag{2.2.35}$$

to be minimized subject to

$$\frac{1}{2} (X - X_1)^T \frac{\partial^2 f}{\partial X^2} (X - X_1) = \frac{k_2^2}{2} \tag{2.2.36}$$

Preceeding formally, the function $\bar{J}$ is formed where

$$\bar{J} = f(X_1) + \left(\frac{\partial f}{\partial X}\right)^T (X - X_1) + \frac{1}{2} (X - X_1)^T \frac{\partial^2 f}{\partial X^2} (X - X_1) + \lambda \left\{ \frac{1}{2} (X - X_1)^T \left(\frac{\partial^2 f}{\partial X^2}\right) (X - X_1) - \frac{k_2^2}{2} \right\}$$

and where λ is a Lagrange multiplier. Thus, setting $\frac{\partial \bar{J}}{\partial X}$ to zero provides

$$\frac{\partial f}{\partial X} + (1 + \lambda) \frac{\partial^2 f}{\partial X^2} (X - X_1) = 0 \tag{2.2.37}$$

Substituting (2.2.37) into (2.2.36) yields the value of λ as

$$(1 + \lambda)^2 = \frac{1}{k_2^2} \left(\frac{\partial f}{\partial X}\right)^T \left(\frac{\partial^2 f}{\partial X^2}\right)^{-1} \left(\frac{\partial f}{\partial X}\right)$$

and combining this expression with (2.2.37) provides the value x_2 as

$$X_2 = X_1 - \frac{k_2 \left(\frac{\partial^2 f}{\partial X^2}\right)^{-1} \frac{\partial f}{\partial X}}{\sqrt{\left(\frac{\partial f}{\partial X}\right)^T \left(\frac{\partial^2 f}{\partial X^2}\right)^{-1} \left(\frac{\partial f}{\partial X}\right)}} \tag{2.2.38}$$

For k_2 sufficiently small, the approximation in (2.2.33) can be used together with (2.2.38) to show that

$$f(X_2) < f(X_1)$$

The process is repeated until the minimum point is founded.

There are many other modified versions of the Newton-Raphson method, all of which vary only as the form of the constraint equation varies. (Eqs. (2.2.34)) The advantage of this modification is that by controlling the magnitude of correction (the quantity k_2). the radius of convergence can be substantially increased.

One of the disadvantages of the Newton-Raphson approach, as compared with the steepest descent method, is that considerably more computation is required

at each step. This extra computation consists primarily in the evaluation of the second derivative matrix $\partial^2 f/\partial x^2$. However, in many practical problems, it suffices to compute this matrix only occasionally (that is, not at every point) and to use some approximate procedure for updating the matrix. Three such approximations are illustrated in Ref. (10).

2.2.3 n-Dimensional Problem with Constraints

This section will demonstrate the application of the Newton-Raphson and steepest descent methods to the n-dimensional problem in which equality constraints have been included.

Again let $f(X) = f(X_1, X_2, \ldots X_n)$ denote the function to be minimized. This time, however, the minimization is to be accomplished subject to the M constraint condition

$$G(X) = \begin{bmatrix} G_1(X) \\ G_2(X) \\ \vdots \\ G_M(X) \end{bmatrix} = 0 \qquad (2.2.39)$$

where M is some integer which is less than n, the dimension of x. If f and G are sufficiently smooth it is a simple matter to show (see Ref. (5)) that the minimizing point must satisfy the conditions

$$\frac{\partial F}{\partial X} = \begin{bmatrix} \frac{\partial F}{\partial X_1} \\ \frac{\partial F}{\partial X_2} \\ \vdots \\ \frac{\partial F}{\partial X_n} \end{bmatrix} = 0 \; ; \qquad \frac{\partial F}{\partial \lambda} = \begin{bmatrix} \frac{\partial F}{\partial \lambda_1} \\ \frac{\partial F}{\partial \lambda_2} \\ \vdots \\ \frac{\partial F}{\partial \lambda_M} \end{bmatrix} = 0 \qquad (2.2.40A)$$

(2.2.40B)

where F is given by

$$F = F(x\lambda) = f + G^T\lambda = f + \sum_{i=1}^{M} \lambda_i G_i \qquad (2.2.41A)$$

and λ is a M-dimensional Lagrange multiplier vector. Using Eq. (2.2.41A), the necessary condition in (2.2.40A) becomes

$$\frac{\partial f}{\partial x} + \frac{\partial G^T}{\partial x}\lambda = 0 \Leftrightarrow \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} + \begin{bmatrix} \frac{\partial G_1}{\partial x_1} & \frac{\partial G_2}{\partial x_1} & \cdots & \frac{\partial G_M}{\partial x_1} \\ \frac{\partial G_1}{\partial x_2} & \frac{\partial G_2}{\partial x_2} & \cdots & \frac{\partial G_M}{\partial x_2} \\ \vdots & \vdots & & \vdots \\ \frac{\partial G_1}{\partial x_n} & \frac{\partial G_2}{\partial x_n} & \cdots & \frac{\partial G_M}{\partial x_n} \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_M \end{bmatrix} = 0 \tag{2.2.41B}$$

while (2.2.40B) takes the form

$$\frac{\partial F}{\partial \lambda} = \begin{bmatrix} \frac{\partial F}{\partial \lambda_1} \\ \frac{\partial F}{\partial \lambda_2} \\ \vdots \\ \frac{\partial F}{\partial \lambda_M} \end{bmatrix} = \begin{bmatrix} G_1 \\ G_2 \\ \\ G_M \end{bmatrix} = G = 0 \tag{2.2.41C}$$

Note that this last equation is simply a restatement of the constraint conditions, Eqs. (2.2.39). The M-dimensional multiplier vector λ is introduced for the explicity purpose of satisfying these constraints. Thus the $n+M$ quantities $x_i\,(i=1,n)$ and $\lambda_j\,(j=1,M)$ are to be selected to satisfy the conditions contained in (2.2.41B) and (2.2.41C). In addition, the second derivative condition takes the form

$$\delta x^T \left(\frac{\partial^2 F}{\partial x^2}\right) \delta x = \sum_{i,j}^{n} \frac{\partial^2 F}{\partial x_i \delta_j} \delta x_i \delta x_j \geq 0 \tag{2.2.42}$$

where

$$\delta x = x - x_{MIN} \tag{2.2.43}$$

and is constrained to satisfy the differential version of the constraint conditions

$$\begin{aligned} \left(\frac{\partial G_1}{\partial x}\right)^T \delta x &= 0 \\ \left(\frac{\partial G_2}{\partial x}\right)^T \delta x &= 0 \\ &\vdots \\ \left(\frac{\partial G_M}{\partial x}\right)^T \delta x &= 0 \end{aligned} \tag{2.2.44}$$

As in the unconstrained case, both iterative techniques for solving this problem consist in solving a succession of quadratic cost minimization problems but subject to linear constraint conditions.

2.2.3.1 Method #1 - Steepest Descent

Let x_1 denote the first guess at the minimizing point, and require that x_1 satisfy the M constraint conditions of Eq. (2.2.39); that is

$$G(X_1) = \begin{bmatrix} G_1(X_1) \\ G_2(X_1) \\ \vdots \\ G_M(X_1) \end{bmatrix} = 0 \tag{2.2.45}$$

Now, expanding the function $F = f + G^T\lambda$ in Taylor series about x_1 and truncating the series after the first-order terms yields

$$F(X,\lambda) \cong F(X_1,\lambda) + \left(\frac{\partial F}{\partial X}\right)^T (X - X_1) = f(X_1) + G^T(X_1)\lambda + \left(\frac{\partial}{\partial X}(f + G^T\lambda)\right)^T (X - X_1) \tag{2.2.46}$$

where λ is as yet an undetermined quantity. The second iterate, x_2, is determined by minimizing the approximation in (2.2.46) subject to a constraint on the magnitude of correction. This constraint takes the usual form

$$(X - X_1)^T (X - X_1) = \sum_{i=1}^{n} (X_i - X_{i_1})^2 = \ell_1^2 \tag{2.2.47}$$

where again ℓ_1 is some arbitrary but small positive constant.

Proceeding formally, the scalar multiplier μ is introduced and the function J is formed where

$$J = F(X_1,\lambda) + \left(\frac{\partial F}{\partial X}^T\right)(X - X_1) + \mu\left\{(X - X_1)^T (X - X_1) - \ell_1^2\right\} \tag{2.2.48}$$

Note that J is a function of both x and λ (by virtue of the fact that $F = F(X,\lambda)$. Hence to minimize J, the derivatives $\partial J/\partial x$ and $\partial J/\partial\lambda$ are set equal to zero providing

$$\frac{\partial J}{\partial X} = 0;\quad X - X_1 = \frac{-1}{2\mu}\frac{\partial F}{\partial X} \iff \begin{bmatrix} X_1 - X_{1_1} \\ X_2 - X_{2_1} \\ \vdots \\ X_n - X_{n_1} \end{bmatrix} = \frac{-1}{2\mu}\begin{bmatrix} \frac{\partial F}{\partial X_1} \\ \frac{\partial F}{\partial X_2} \\ \vdots \\ \frac{\partial F}{\partial X_n} \end{bmatrix} \tag{2.2.49A}$$

$$\frac{\partial J}{\partial \lambda} = 0;\ G(X) + \left(\frac{\partial G}{\partial X}\right)(X - X_1) = 0 \iff \begin{bmatrix} G_1(X_1) \\ G_2(X_1) \\ \vdots \\ G_M(X_1) \end{bmatrix} + \begin{bmatrix} \frac{\partial G_1}{\partial X_1} & \frac{\partial G_1}{\partial X_2} \cdots \frac{\partial G_1}{\partial X_n} \\ \frac{\partial G_2}{\partial X_1} & \frac{\partial G_2}{\partial X_2} \cdots \frac{\partial G_2}{\partial X_n} \\ \vdots \\ \frac{\partial G_M}{\partial X_1} & \frac{\partial G_n}{\partial X_2} \cdots \frac{\partial G_M}{\partial X_n} \end{bmatrix} \begin{bmatrix} X_1 - X_{1_1} \\ X_2 - X_{2_1} \\ X_n - X_{n_1} \end{bmatrix} = 0 \tag{2.2.49B}$$

Eq. (2.2.49A) is essentially the same as that encountered in the unconstrained problem. Also, since $G(X_1) = 0$, Eq. (2.2.49B) reduced to the M conditions

$$\frac{\partial G}{\partial X}(X - X_1) = 0 \iff \left(\frac{\partial G}{\partial X}^{j}\right)^{T}(X - X_1); \quad j = 1, M \tag{2.2.50}$$

which is simply a requirement that the second iterate, x_2, satisfy the first-order approximation to the constraint condition $G(x) = 0$. The M-dimensional multiplier λ is to be selected so that these equations (the linearized version of the actual constraints) are satisfied. This selection is denomstrated next.

Substituting (2.2.49A) into (2.2.50) yields

$$-\frac{1}{2\mu}\ \frac{\partial G}{\partial X}\ \frac{\partial F}{\partial X} = 0$$

which, from (2.2.41A) reduces to

$$\frac{\partial G}{\partial X}\ \frac{\partial f}{\partial X} + \frac{\partial G}{\partial X}\left(\frac{\partial G}{\partial X}\right)^{T}\lambda = 0 \tag{2.2.51}$$

or

$$\lambda = -\left[\left(\frac{\partial G}{\partial X}\right)\left(\frac{\partial G}{\partial X}\right)^{T}\right]^{-1}\frac{\partial G}{\partial X}\ \frac{\partial f}{\partial X} \tag{2.2.52}$$

In scalar form, this equation becomes

$$\begin{bmatrix}\lambda_1\\ \lambda_2\\ \\ \lambda_M\end{bmatrix} = -\left\{\begin{bmatrix}\frac{\partial G_1}{\partial X_1} & \cdots & \frac{\partial G_1}{\partial X_n}\\ \frac{\partial G_2}{\partial X_1} & \cdots & \frac{\partial G_2}{\partial X_n}\\ \vdots & & \\ \frac{\partial G_M}{\partial X_1} & \cdots & \frac{\partial G_M}{\partial X_n}\end{bmatrix}\begin{bmatrix}\frac{\partial G_1}{\partial X_1} & \cdots & \frac{\partial G_M}{\partial X_1}\\ \frac{\partial G_1}{\partial X_2} & \cdots & \frac{\partial G_M}{\partial X_2}\\ \vdots & & \\ \frac{\partial G_1}{\partial X_n} & \cdots & \frac{\partial G_M}{\partial X_n}\end{bmatrix}\right\}^{-1}\begin{bmatrix}\frac{\partial G_1}{\partial X_1} & \cdots & \frac{\partial G_1}{\partial X_n}\\ \frac{\partial G_2}{\partial X_1} & \cdots & \frac{\partial G_2}{\partial X_n}\\ \vdots & & \\ \frac{\partial G_M}{\partial X_1} & \cdots & \frac{\partial G_M}{\partial X_n}\end{bmatrix}\begin{bmatrix}\frac{\partial f}{\partial X_1}\\ \frac{\partial f}{\partial X_2}\\ \vdots\\ \frac{\partial f}{\partial X_n}\end{bmatrix} \tag{2.2.53}$$

Thus, with λ selected to satisfy (2.2.53), the linearized version of the constraint equation (Eq. (2.2.50)) will be satisfied and to first order

$$G(X_2) = G(X_1) + \frac{\partial G}{\partial X}(X - X_1) = 0$$

Once λ is known, the function $F(x) = f(x) + G(x)^{T}\lambda$ and the derivative $(\partial F/\partial x)_{x_1}$ can be evaluated at the point $x - x_1$. With this result, the quantity μ can be determined by substituting (2.2.49A) into (2.2.47) to yield

$$\mu = \frac{1}{2k_1}\sqrt{\left(\frac{\partial F}{\partial X}\right)^{T}\frac{\partial F}{\partial X}} \tag{2.2.54}$$

Hence, the second iterate satisfies the equation

$$X_2 = X_1 - \frac{\ell_1 \frac{\partial F}{\partial X}}{\sqrt{\left(\frac{\partial F}{\partial X}\right)^T \frac{\partial F}{\partial X}}} \qquad (2.2.55)$$

The similarity between this result and Eq. (2.2.27) (the unconstrained result) should be noted.

Summarizing the results for the steepest descent technique, the second iterate x_2 is selected using Eq. (2.2.55) where ℓ_1 is a positive small constant which regulates the magnitude of correction as indicated in Eq. (2.2.47). To evaluate the derivative $\frac{\partial F}{\partial x} = \frac{\partial f}{\partial x} + \frac{\partial G^T}{\partial x}\lambda$, the λ vector must be determined by means of Eq. (2.2.53) to ensure that the first-order approximation to the constraint equations is satisfied. The process is repeated until the minimum point is found.

2.2.3.2 Method #2 - Newton-Raphson Technique

Let x_1 denote the first guess at the minimum point and let λ_1 denote the first guess at the correct value of λ. Expanding the function $F(x,\lambda) = f + G^T\lambda$ about the point x_1, λ_1 in a Taylor series truncated after the second-order terms provides

$$F(X, \lambda) \cong F(X_1, \lambda_1) + \left(\frac{\partial F}{\partial X}\right)^T (X - X_1) + \left(\frac{\partial F}{\partial \lambda}\right)^T (\lambda - \lambda_1)$$
$$+ \frac{1}{2}\left\{(X-X_1)^T \frac{\partial^2 F}{\partial X^2}(X-X_1) + 2(X-X_1)^T \frac{\partial^2 F}{\partial X \partial \lambda}(\lambda-\lambda_1) + (\lambda-\lambda_1)^T \frac{\partial^2 F}{\partial \lambda^2}(\lambda-\lambda_1)\right\} \qquad (2.2.56)$$

If the point x_1, λ_1 is sufficiently close to the correct point, them selecting x_2 and λ_2 so as to minimize the second-order approximation to $F(x, \lambda)$ should provide improved values of x and λ.

Proceeding with the minimization, let J be defined by

$$J = F(X_1, \lambda_1) + \left(\frac{\partial F}{\partial X}\right)^T (X - X_1) + \left(\frac{\partial F}{\partial \lambda}\right)^T (\lambda - \lambda_1) + \frac{1}{2}\left\{(X-X_1)^T \frac{\partial^2 F}{\partial X^2}(X-X_1)\right.$$
$$\left. + 2(X-X_1)^T \frac{\partial^2 F}{\partial X \partial \lambda}(\lambda-\lambda_1) + (\lambda-\lambda_1)^T \frac{\partial^2 F}{\partial \lambda^2}(\lambda-\lambda_1)\right\} \qquad (2.2.57)$$

Now forming $\frac{\partial J}{\partial X}$ and $\frac{\partial J}{\partial \lambda}$ and equating the results to zero yields

$$\frac{\partial F}{\partial X} + \frac{\partial^2 F}{\partial X^2}(X - X_1) + \left(\frac{\partial G}{\partial X}\right)^T (\lambda - \lambda_1) = 0$$

$$\frac{\partial F}{\partial \lambda} + \frac{\partial^2 F}{\partial \lambda^2}(\lambda - \lambda_1) + \frac{\partial^2 F}{\partial X \partial \lambda}(X - X_1) = 0$$

In view of the definition of the quantity F , these two equations reduce to

$$\frac{\partial F}{\partial X} + \left(\frac{\partial^2 F}{\partial X \partial \lambda}\right)^T (\lambda - \lambda_1) + \frac{\partial^2 F}{\partial X^2} (X - X_1) = 0 \qquad (2.2.58A)$$

$$G + \frac{\partial G}{\partial X} (X - X_1) = 0 \qquad (2.2.58B)$$

or in the scalar notation

$$\begin{vmatrix} \frac{\partial F}{\partial X_1} \\ \frac{\partial F}{\partial X_2} \\ \vdots \\ \frac{\partial F}{\partial X_n} \end{vmatrix} + \begin{vmatrix} \frac{\partial^2 F}{\partial X_1^2} & \frac{\partial^2 F}{\partial X_1 \partial X_2} & \cdots & \frac{\partial^2 F}{\partial X_1 \partial X_n} \\ \vdots & & & \\ \frac{\partial^2 F}{\partial X_n \partial X_1} & \frac{\partial^2 F}{\partial X_n \partial X_2} & \cdots & \frac{\partial^2 F}{\partial X_n^2} \end{vmatrix} \begin{vmatrix} X_1 - X_{1_1} \\ X_2 - X_{2_1} \\ \vdots \\ X_n - X_{n_1} \end{vmatrix} + \begin{vmatrix} \frac{\partial G_1}{\partial X_1} & \frac{\partial G_2}{\partial X_1} & \cdots & \frac{\partial G_M}{\partial X_1} \\ \vdots & & & \\ \frac{\partial G_1}{\partial X_n} & \frac{\partial G_2}{\partial X_n} & \cdots & \frac{\partial G_M}{\partial X_n} \end{vmatrix} \begin{vmatrix} \lambda_1 - \lambda_{1_1} \\ \lambda_2 - \lambda_{2_1} \\ \vdots \\ \lambda_M - \lambda_{M_1} \end{vmatrix} = 0 \qquad (2.2.59A)$$

$$\begin{bmatrix} G_1(X_1) \\ G_2(X_1) \\ \vdots \\ G_M(X_1) \end{bmatrix} + \begin{bmatrix} \frac{\partial G_1}{\partial X_1} & \frac{\partial G_1}{\partial X_2} & \cdots & \frac{\partial G_1}{\partial X_n} \\ \vdots & & & \\ \frac{\partial G_M}{\partial X_1} & \frac{\partial G_M}{\partial X_2} & \cdots & \frac{\partial G_M}{\partial X_n} \end{bmatrix} \begin{bmatrix} X_1 - X_{1_1} \\ X_2 - X_{2_1} \\ \vdots \\ X_n - X_{n_1} \end{bmatrix} = 0 \qquad (2.2.59B)$$

Eqs. (2.2.59A) and (2.2.59B) constitute a set of $n+M$ linear equations in the $n+M$ unknowns, $(x_i - x_{i_1})$, $i = 1, n$ and $(\lambda_j - \lambda_{j_1})$, $j = 1, M$. This system possesses the solution

$$\begin{bmatrix} X - X_1 \\ \lambda - \lambda_1 \end{bmatrix} = \begin{bmatrix} X_2 - X_1 \\ \lambda_2 - \lambda_1 \end{bmatrix} = -B^{-1} \begin{bmatrix} \frac{\partial F}{\partial X} \\ G \end{bmatrix} = -B^{-1} \begin{bmatrix} \frac{\partial F}{\partial X_1} \\ \vdots \\ \frac{\partial F}{\partial X_n} \\ G_1 \\ G_2 \\ \vdots \\ G_M \end{bmatrix} \qquad (2.2.60)$$

where B is the $(n+M) \times (n+M)$ matrix

$$B = \begin{vmatrix} \frac{\partial^2 F}{\partial X^2} & \left(\frac{\partial G}{\partial X}\right)^T \\ (n \times n) & (n \times M) \\ \frac{\partial G}{\partial X} & 0 \\ (M \times n) & (M \times M) \end{vmatrix} \tag{2.2.61}$$

The process is repeated until the minimum point(x_{MIN}) and the correct λ are reached.

As in the unconstrained problem, a modified version of the Newton-Raphson technique can be developed in which a magnitude constraint is placed on the size of the step to be taken. As in Eq. (2.2.34B) this constraint would take the form

$$\frac{1}{2}\left[(X-X_1)^T, (\lambda-\lambda_1)^T\right] A \begin{bmatrix} X-X_1 \\ \lambda-\lambda_1 \end{bmatrix} = k_2^2 \tag{2.2.62}$$

where A is some positive definite symmetric matrix. The minimization of in (2.2.57) subject to this constraint would lead to an iteration equation similar to that resulting in the unconstrained case (see Eq. (2.2.38)). By such a modification, the range of convergence can be considerably extended.

2.2.4 Discussion

The fact that only two numerical procedures have been presented for solving maxima-minima problems is not meant to imply that the methods of steepest descent and Newton-Raphson are the only two methods available. Indeed, there are many other techniques for numerically finding an extremum value of a function. However, all of these techniques are local in character; that is, they depend completely on the behavior of the function in the vicinity of current estimate of the minimizing point (or in the vicinity of the current estimate and the preceding estimates of the minimizing point). This behavior of the function at the current estimate of x_{MIN} is determined through the evaluation of the first few terms in a series expansion of the function about the current point. (Recall that in the steepest descent procedure this behavior was evaluated through the computation of the derivative $\partial f/\partial x$ while in the Newton-Raphson technique, first and second derivatives were used.) Thus, while there are many numerical procedures, they are conceptually very similar to the two presented here - although the differences may be very important computationally for certain problems.

As has been shown, only "quadratic cost" functions can be minimized directly since in this case the first derivative condition leads to a set of linear equations. Hence, the basis of both iterative techniques consists in approximating the function to be minimized by a quadratic cost function, and then solving a succession of linear problems.

A similar situation occurs in variational problems (for example, the problem of Mayer in Section 2.1). Only a certain type of quadratic variational problem can be solved directly (the variational analog of the quadratic cost problem in maxima-minima theory). In this case, the first variation condition (the Euler equations) leads to a set of linear differential equations which have a closed-form solution. Hence the basic approach of the numerical iteration techniques used in variational problems is to approximate the functional to be minimized by a quadratic cost functional and then to solve a succession of linear problems. If only the first variation about some current estimate of the minimizing solution is used to compute the next estimate, the numerical method is referred to as the gradient or steepest descent procedure and is essentially a function space extrapolation of the gradient technique in maxima-minima theory. If both first and second variations are used in the calculation of the next estimate, then either of two numerical procedures can result, neighboring extremal or quasilinearization. Both constitute the variational analog of the Newton-Raphson technique.

In the next section, the use of these three techniques, neighboring extremal, quasilinearization and steepest descent, are illustrated in connection with the one-dimensional Lagrange problem. As in the one-dimensional maxima-minima problem, the one-dimensional Lagrange problem is rather easy to solve and hence, does not warrant extensive numerical treatment. However, this elementary problem clearly demonstrates what is involved in the application of each of the three numerical procedures. The complexity in the extension to these techniques to the problem of Mayer is then a matter of algebra and computation, rather than concept.

2.3 NUMERICAL SOLUTION OF THE ONE-DIMENSIONAL LAGRANGE PROBLEM

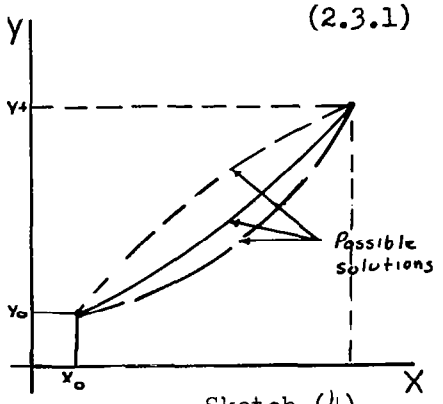
The one-dimensional Lagrange problem consists of determining the function for which the integral

$$J = \int_{x_0}^{x_f} f(x, y, y') \, dx \tag{2.3.1}$$

is a minimum. The minimizing arc is required to satisfy the boundary conditions

$$y = y_0 \quad \text{AT} \quad x = x_0$$
$$y = y_f \quad \text{AT} \quad x = x_f \tag{2.3.2}$$

The physical situation is pictured on the graph to the right. Typical problems of this type are the shortest distance and brachistocrone problems. (see Ref. (7)).



Sketch (4)

Setting the first variation of J to zero provides the first necessary condition (the Euler equation)

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) - \frac{\partial f}{\partial y} = 0 \tag{2.3.3}$$

while the second variation requires, in part, that

$$\frac{\partial^2 f}{\partial y'^2} \geq 0 \tag{2.3.4}$$

In addition, three other conditions must be satisfied by the minimizing solution $y_{MIN}(x)$; the Jacobi condition (see Ref. (7)), the Weierstrass condition, which takes the form

$$f(x,y,Y') - f(x,y,y') - (Y'-y')\frac{\partial f}{\partial y'}(x,y,y') \geq 0$$

$$y = y_{MIN} \; ; \; y' = y'_{MIN} \tag{2.3.5}$$

and the corner condition, which states that across a slope discontinuity

$$\frac{\partial f}{\partial y'}(x,y,y'^{(+)}) = \frac{\partial f}{\partial y'}(x,y,y'^{(-)}) \tag{2.3.6}$$

In the analysis which follows, use will be made only of the Euler and Lagrange conditions (with occasional reference to the Jacobi condition). Also, it will be assumed that the minimizing solution $y_{MIN}(x)$, has a continuous first derivative (i.e., $y'_{MIN}(x)$ is continuous). This last assumption is rather weak and is made only to simplify the analysis.

The minimizing solution is to satisfy the Euler condition (2.3.3). This is a second-order differential equation and can be written as

$$\frac{d}{dX}\left(\frac{\partial f}{\partial y'}\right) - \frac{\partial f}{\partial y} = 0 \iff y''\frac{\partial^2 f}{\partial y'^2} + y'\frac{\partial^2 f}{\partial y \partial y'} + \frac{\partial^2 f}{\partial y' \partial x} - \frac{\partial f}{\partial y} = 0 \tag{2.3.7}$$

In addition, the boundary conditions

$$y = y_o \quad AT \quad X = X_o$$
$$y = y_f \quad AT \quad X = X_f \tag{2.3.8}$$

must hold. Thus, equations (2.3.7) and (2.3.8) constitute a two-point boundary problem and, as indicated in Section (2.1), such problems must be solved iteratively, except in one special case.

Suppose the functional to be minimized takes the form

$$J = \frac{1}{2}\int_{X_o}^{X_f} (a y'^2 + 2b y' y + c y^2 + 2d y' + 2e y)\, dX \tag{2.3.9}$$

where a, b, c, d and e denote functions of the independent variable x; that is

$$\begin{aligned} a &= a(X) \\ b &= b(X) \\ c &= c(X) \\ d &= d(X) \\ e &= e(X) \end{aligned} \tag{2.3.10}$$

Analogous to the maxima-minima problem, functionals of this form, will be referred to as "quadratic cost" functional. Again the boundary conditions are of the two point type with

$$y = y_o \quad AT \quad X = X_o$$
$$y = y_f \quad AT \quad X = X_f \tag{2.3.11}$$

In this case however, the Euler condition of (2.3.7) takes the linear form

$$y'' a(x) + y' (a'(x)) + y (b'(x) - c(x)) = e(x) - d'(x) \tag{2.3.12}$$

At this point, letting

$$y_1 = y \qquad y_2 = y' \tag{2.3.13}$$

allows it to be noted that Eq. (2.3.12) is equivalent to the two first-order equations

$$\begin{bmatrix} y_1' \\ y_2' \end{bmatrix} = G \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} + \begin{bmatrix} 0 \\ \frac{e-d'}{a} \end{bmatrix} \tag{2.3.14}$$

where the matrix G is given by

$$^{*}\ G = \begin{bmatrix} g_{11}(x) & g_{12}(x) \\ g_{21}(x) & g_{22}(x) \end{bmatrix} = \begin{bmatrix} 0 & 1 \\ -\left(\frac{b'(x)-c(x)}{a(x)}\right) & -\frac{a'(x)}{a(x)} \end{bmatrix} \tag{2.3.15}$$

This system has the solution

$$\begin{bmatrix} y_1(x) \\ y_2(x) \end{bmatrix} = \phi(x) \left\{ \begin{bmatrix} y_{1_o} \\ y_{2_o} \end{bmatrix} + \int_{x_o}^{x} \phi^{-1} \begin{bmatrix} 0 \\ \frac{e-d'}{a} \end{bmatrix} dx \right\} \tag{2.3.16}$$

---

* From the Lagrange condition in (2.3.4), $a(x) \geq 0$ all $x \in (x_o, x_f)$ for the minimization problem to make sense. In writing the G matrix, it is assumed that the strong version of this condition holds with $a(x) > 0$.

where

$$y_{1_0} = y_1(x_0) = y_0$$
$$y_{2_0} = y_2(x_0) = y_0' \tag{2.3.17}$$

and where ϕ is the fundamental matrix

$$\dot{\phi} = G\phi \iff \begin{bmatrix} \dot{\phi}_{11} & \dot{\phi}_{12} \\ \dot{\phi}_{21} & \phi_{22} \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix} \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$$
$$\phi(x_0) = I \qquad \begin{bmatrix} \phi_{11}(x_0) & \phi_{12}(x_0) \\ \phi_{21}(x_0) & \phi_{22}(x_0) \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \tag{2.3.18}$$

Note that the development of the ϕ matrix is an initial value problem (all boundary data is available at $x = x_0$). Hence, the solution to (2.3.18) can be achieved in one run (one integration) on a digital computer.

Since the minimizing arc has the general form indicated in (2.3.16), all that is involved in completing the solution is the selection of the initial slope $y_{2_0} = y_0' = y'(x_0)$ so that the terminal condition $y_{1_f} = y_f$ at x_f is satisfied. But from (2.3.16)

$$y_{1_f} = \phi_{11}(x_f)\left\{ y_{1_0} - \int_{x_0}^{x_f} \left[\frac{\phi_{12}}{\phi_{11}\phi_{22} - \phi_{12}\phi_{21}}\right]\left[\frac{e - d'}{a}\right] dx \right\}$$
$$+ \phi_{12}(x_f)\left\{ y_{2_0} + \int_{x_0}^{x_f} \left[\frac{\phi_{11}}{\phi_{11}\phi_{22} - \phi_{12}\phi_{21}}\right]\left[\frac{e - d'}{a}\right] dx \right\}$$

Thus

$$* y_{2_0} = \frac{1}{\phi_{12}(x_f)} \left\{ y_{1_f} - \phi_{11}(x_f)\left(y_{1_0} - \int_{x_0}^{x_f} \left[\frac{\phi_{12}}{\phi_{11}\phi_{22} - \phi_{12}\phi_{21}}\right]\left[\frac{e - d'}{a}\right] dx \right) \right\}$$
$$- \int_{x_0}^{x_f} \left[\frac{\phi_{11}}{\phi_{11}\phi_{22} - \phi_{12}\phi_{21}}\right]\left[\frac{e - d'}{a}\right] dx \tag{2.3.19}$$

* It is assumed that $\phi_{12}(x_f) \neq 0$, a condition which will be satisfied provided the strong version of the Jacobi condition holds (see Ref. 11).

Substitution of this expression into (2.3.16) now provides the minimizing solution for the quadratic cost functional J in Eq. (2.3.9).

The quadratic cost problem of Eq. (2.3.9) can be solved directly (without iteration) because the Euler equation is linear. In the general case, the Euler equation is nonlinear and the resulting two-point boundary value problem must be solved iteratively on a computer. As in the maxima-minima case, the iterative procedures presented below consist in approximating the functional to be minimized by a quadratic cost functional and then solving a succession of linear problems.

2.3.1 Steepest Descent (Gradient) Technique

The problem under consideration is the determination of the arc which minimizes the functional

$$J = \int_{x_0}^{x_f} f(x, y, y')\, dx \tag{2.3.20}$$

and goes through the boundary points

$$\begin{aligned} y &= y_0 \quad AT \quad X = X_0 \\ y &= y_f \quad AT \quad X = X_f \end{aligned} \tag{2.3.21}$$

Let $y_1(x)$ denote the first guess of the minimizing arc and require that y_1 satisfy the boundary conditions of (2.3.21) with

$$\begin{aligned} y_1(X_0) &= y_0 \\ y_1(X_f) &= y_f \end{aligned} \tag{2.3.22}$$

In addition, let $J(y_1)$ denote the corresponding value of the functional J. Then, if the second iterate $y_2(x)$ is to be more nearly minimizing, it must be selected so that the inequality

$$J(y_2) < J(y_1) \tag{2.3.23}$$

holds.

Expanding the functional J in a Taylor series about the arc $y_1(x)$ and truncating the series after the first-order terms provides

$$\begin{aligned} J(y) &\cong J(y_1) + \delta J(y - y_1) \\ &= \int_{x_0}^{x_f} f(X, y_1, y_1')\, dX + \int_{x_0}^{x_f} \left[\frac{\partial f}{\partial y}\, \delta y + \frac{\partial f}{\partial y'}\, \delta y' \right] dx \end{aligned} \tag{2.3.24}$$

where

$$\delta y = \delta y(X) = y(X) - y_1(X)$$
$$\delta y' = \delta y'(X) = y'(X) - y_1'(X) \tag{2.3.25}$$

Integrating the term $\frac{\partial f}{\partial y'}\,\delta y'$ in (2.3.24) by parts and noting that $\delta y(X_0)$ and $\delta y(X_f)$ must be zero for the new solution to satisfy the boundary conditions, it follows that

$$J(y) \cong J(y_1) + \delta J$$
$$= \int_{X_0}^{X_f} f(X, y_1, y_1')\,dX + \int_{X_0}^{X_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dX}\frac{\partial f}{\partial y'}\right] \delta y\,dX \tag{2.3.26}$$

Now, if $y_2(x)$ is "sufficiently close" to $y_1(x)$ { i.e., $\delta y(x)$ is small }, then the first-order approximation in (2.3.26) will be reasonably good. Hence, selecting $y_2(x)$ to minimize the expression in (2.3.26) subject to a magnitude constraint in the separation $|\delta y(x)| = |y(x) - y_1(x)|$ should provide an improved second iterate, $y_2(x)$ (i.e., one which satisfies the inequality in (2.3.23)). There are two different types of magnitude constraints which can be imposed. Both lead to slightly different versions of the steepest descent technique.

Method #1

In this method, the quantity $\delta y(x)$ is selected to minimize the expression

$$J(y_1) + \delta J$$
$$= J(y_1) + \int_{X_0}^{X_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dX}\frac{\partial f}{\partial y'}\right] \delta y\,dX \tag{2.3.27}$$

subject to a magnitude constraint

$$\int_{X_0}^{X_f} \left(\delta y(X)\right)^2\,dX = k_1^2 \tag{2.3.28}$$

where k_1 is some positive small constant. Such a constraint insures that $\delta y(x)$ will be sufficiently small at all values of x in the interval (x_0, x_f).

To minimize the expression in (2.3.27) subject to the magnitude constraint, the quantity $\bar{J}$ is formed where

$$\bar{J} = J(y) + \int_{X_0}^{X_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dX} \frac{\partial f}{\partial y'} \right] \delta y \, dX + \lambda \left\{ \int_{X_0}^{X_f} [\delta y(X)]^2 \, dX - k_1^2 \right\} \tag{2.3.29}$$

and where λ is a constant multiplier to be determined so that the magnitude constraint in (2.3.28) is satisfied. The quantity $\delta y(x)$ is now chosen to minimize $\bar{J}$ which can be rewritten as

$$\bar{J}(\delta y) = J(y_1) - \lambda k_1^2 + \int_{X_0}^{X_f} F(X, \delta y) dX \tag{2.3.30}$$

where

$$F = \left[\frac{\partial f}{\partial y} - \frac{d}{dX} \frac{\partial f}{\partial y'} \right] \delta y + \lambda \delta y^2$$

Hence, δy must satisfy the Euler equation

$$\frac{\partial F}{\partial (\delta y)} - \frac{d}{dX} \left(\frac{\partial F}{\partial (\delta y')} \right) = 0 \tag{2.3.31}$$

Since F is not a function of $\delta y'$, it follows that

$$\frac{\partial F}{\partial \delta y} = 0 \iff \delta y = -\frac{1}{2\lambda} \left[\frac{\partial f}{\partial y} - \frac{d}{dX} \frac{\partial f}{\partial y'} \right] \tag{2.3.32}$$

This final equation allows the multiplier λ to be determined by substituting Eq. (2.3.32) into (2.3.28) to yield

$$^{*}\lambda = \frac{1}{2k_1} \left\{ \int_{X_0}^{X_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dX} \frac{\partial f}{\partial y'} \right]^2 dX \right\}^{\frac{1}{2}} \tag{2.3.33}$$

Collecting results for this method, the quantity $\delta y(x)$ is selected to minimize the first-order approximation of the minimizing functional J subject to the magnitude constraint given in (2.3.28). This leads to the expression

---

* Note that the positive square root is taken in this equation. This choice follows from a second variation test which requires λ to be greater than zero.

$$\delta y(x) = \frac{-k_1\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right)}{\left\{\int_{x_0}^{x_f}\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right]^2 dx\right\}^{\frac{1}{2}}} \tag{2.3.34}$$

at points x interior to the interval (x_0, x_f). At the boundary points, $\delta y(x)$ must satisfy

$$\delta y(x_0) = 0$$

$$\delta y(x_f) = 0 \tag{2.3.35}$$

The second iterate, $y_2(x)$, is then determined from the expression

$$y_2(x) = y_1(x) + \delta y(x) \tag{2.3.36}$$

Substitution of (2.3.34) into (2.3.26) leads to the first-order approximation

$$J(y_2) \cong J(y_1) - k_1\left\{\int_{x_0}^{x_f}\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right]^2 dx\right\}^{\frac{1}{2}} \tag{2.3.37}$$

where the quantity $\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}$ is evaluated along the first iterate $y_1(x)$. Hence, $J(y_2) < J(y_1)$ and $y_2(x)$ will be an improved approximation to the minimizing solution provided the constant k_1 is sufficiently small.

Some comments on the form of the solution are in order. First, the optimizing condition for the problem of Lagrange (the Euler condition)

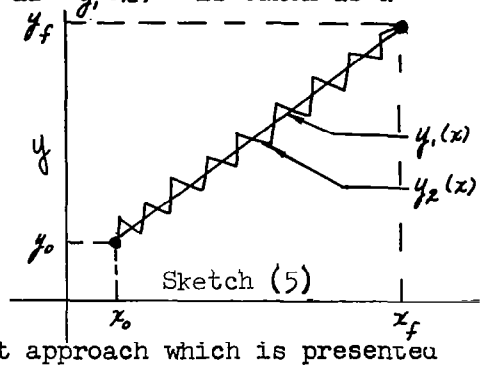
$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = 0$$

is the analog of the first derivative condition

$$\frac{\partial f}{\partial x} = 0$$

in maxima-minima theory. In fact, the Euler operator $\left(\frac{\partial}{\partial y} - \frac{d}{dx}\frac{\partial}{\partial y'}\right)$, may be considered the function space equivalent of the derivative operator $\partial/\partial x$ which operates in the space of the vector x. From this point of view, the iteration process incorporated in Eq. (2.3.34) is essentially the same as that used in Eq. (2.2.27) for locating the minimum of a function.

The magnitude constraint of Eq. (2.3.28) requires only that the solution to $y_2(x)$ be sufficiently close to $y_1(x)$. If $y_1(x)$ is taken as a straight line connecting the points (x_0, y_0) and (x_f, y_f), then it is not unlikely that $y_2(x)$ will take the form shown in the sketch to the right; that is, starting with a smooth iterate $y_1(x)$, it is entirely possible that the second iterate may be highly unsmooth and contain several corners. The reason for this behavior is that the magnitude constraint in (2.3.28) requires only that $y_2(x)$ be close to $y_1(x)$, but allows the derivatives $y_2'(x)$ and $y_1'(x)$ to differ considerably. The second steepest descent approach which is presented next avoids this difficulty and leads to iterated arcs which are as smooth as the first arc $y_1(x)$



Sketch (5)

Method #2

As in the previous method, the quantity $\delta y(x)$ is to be selected to minimize the first-order approximation to be functional J which takes the form

$$J(y) \cong J(y_1) + \delta J(y - y_1)$$
$$= J(y_1) + \int_{X_0}^{X_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dX}\frac{\partial f}{\partial y'}\right] \delta y \, dX \tag{2.3.38}$$

In this case, however, a magnitude constraint of the form

$$\int_{X_0}^{X_f} \left(\delta y'(X)\right)^2 dX = k_2^2 \tag{2.3.39}$$

is imposed where k_2 is some positive small constant. Since $\delta y(x)$ must satisfy the boundary conditions

$$\delta y(X_0) = 0$$
$$\delta y(X_f) = 0 \tag{2.3.40}$$

and, since $\delta y'$ must be small in order to satisfy (2.3.39), it follows that δy will also be small. Hence, the problem encountered in the previous method in which δy was small but $\delta y'$ was not (that is, $y_2(x)$ was not smooth), is not present here.

To minimize the expression in (2.3.38) subject to the constraint in (2.3.39), the functional $\bar{J}$ is formed where

$$\bar{J} = J(y_1) + \int_{x_0}^{x_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right] \delta y \, dx + \lambda \left\{ \int_{x_0}^{x_f} (\delta y'(x))^2 \, dx - k_2^2 \right\} \tag{2.3.41}$$

and where λ is a constant multiplier. Rewriting this expression as

$$\bar{J} = J(y_1) - \lambda k^2 + \int_{x_0}^{x_f} F(x, \delta y, \delta y') \, dx \tag{2.3.42}$$

with

$$F = \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right] \delta y + \lambda \delta y'^2 \tag{2.3.43}$$

it follows that the minimizing arc $\delta y(x)$ must satisfy the Euler equation

$$\frac{\partial F}{\partial(\delta y)} - \frac{d}{dx}\left(\frac{\partial F}{\partial(\delta y')}\right) = 0 \implies \delta y'' = +\frac{1}{2\lambda}\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right) \tag{2.3.44}$$

together with the boundary conditions

$$\begin{aligned} \delta y(x_0) &= 0 \\ \delta y(x_f) &= 0 \end{aligned} \tag{2.3.45}$$

Equation (2.3.44) has the solution

$$\delta y(x) = \cancel{\delta y}^{\,0}(x_0) + (x - x_0)\,\delta y'(x_0) + \frac{1}{2\lambda}\left\{ \int_{x_0}^{x} (x_0 - x)\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right] dx + (x - x_0) \int_{x_0}^{x} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right] dx \right\} \tag{2.3.46}$$

Thus, using the second condition in (2.3.45), it follows that $\delta y'(x_o)$ must satisfy

$$\delta y'(X_o) = \frac{-1}{2\lambda(X_f - X_o)} \left\{ \int_{X_o}^{X_f} (X_o - X) \left[\frac{\partial f}{\partial y} - \frac{d}{dX} \frac{\partial f}{\partial y'} \right] dX + (X_f - X_o) \int_{X_o}^{X_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dX} \frac{\partial f}{\partial y'} \right] dX \right\} \tag{2.3.47}$$

Finally, substitution of this expression back into (2.3.46) provides the desired solution. The constant multiplier λ is then evaluated using Eq. (2.3.46) and the magnitude condition of Eq. (2.3.39). While these calculations are algebraically involved, they are easily accomplished on a digital computer once $y_1(X)$ has been specified. The second iterate satisfied

$$y_2(X) = y_1(X) + \delta y(X) \tag{2.3.48}$$

where δy is evaluated using (2.3.46).

2.3.2 Quasilinearization

Again, the problem is to determine the arc $y_{MIN}(x)$ which minimizes the functional

$$J(y) = \int_{X_o}^{X_f} f(X, y, y')\, dX = MIN \tag{2.3.51}$$

subject to the boundary conditions that

$$y = y_o \quad AT \quad X = X_o$$
$$y = y_f \quad AT \quad X = X_f \tag{2.3.52}$$

Let $y_1(X)$ denote the first guess at the minimizing arc (with $y_1(x)$ satisfying the boundary conditions in (2.3.52)) and expand $J(y)$ in a Taylor series about $y_1(x)$. Truncating the series after the second-order terms provides

$$J(y) \cong J(y_1) + \delta J(y-y_1) + \frac{\delta^2 J(y-y_1)}{2}$$

$$= \int_{x_0}^{x_f} f(x, y_1, y_1')\,dx + \int_{x_0}^{x_f} \left[\frac{\partial f}{\partial y}\,\delta y + \frac{\partial f}{\partial y'}\,\delta y'\right] dx$$

$$+ \frac{1}{2}\int_{x_0}^{x_f} \left[\frac{\partial^2 f}{\partial y^2}\,\delta y^2 + \frac{2\partial^2 f}{\partial y \partial y'}\,\delta y \delta y' + \frac{\partial^2 f}{\partial y'^2}\,\delta y'^2\right] dx \qquad (2.3.53)$$

Thus, for $y_1(x)$ sufficiently close to $y_{MIN}(x)$, selecting $y_2(x)$ to minimize this approximation should provide an improved arc with

$$J(y_2) < J(y_1) \qquad (2.3.54)$$

Before proceeding with the minimization, rewrite Eq. (2.3.53) in the form

$$J(y) \cong J(y_1) + \int_{x_0}^{x_f} F(x, \delta y, \delta y')\,dx \qquad (2.3.55)$$

where

$$F = \frac{\partial f}{\partial y}\,\delta y + \frac{\partial f}{\partial y'}\,\delta y' + \frac{1}{2}\left[\frac{\partial^2 f}{\partial y^2}\,\delta y^2 + \frac{\partial^2 f}{\partial y \partial y'}\,\delta y \delta y' + \frac{\partial^2 f}{\partial y'^2}\,\delta y'^2\right] \qquad (2.3.56)$$

Hence, for $\delta y(x)$ to be minimizing, the Euler condition

$$\frac{\partial F}{\partial(\delta y)} - \frac{d}{dx}\,\frac{\partial F}{\partial(\delta y')} = 0 \qquad (2.3.57)$$

must hold. This equation, in turn, reduces to the second-order equation

$$\delta y''\,a(x) + \delta y'(a'(x)) + \delta y\,(b'(x) - c(x)) = e(x) - d'(x) \qquad (2.3.58)$$

where the coefficients a, b, c, d and e are given by

$$a = \frac{\partial^2 f}{\partial y'^2} \qquad d = \frac{\partial f}{\partial y'}$$

$$b = \frac{\partial^2 f}{\partial y \partial y'} \qquad e = \frac{\partial f}{\partial y}$$

$$c = \frac{\partial^2 f}{\partial y^2} \tag{2.3.59}$$

Note that this differential equation is linear and takes the same form as that given in (2.3.12). Hence, its solution is identical to that developed in Eqs. (2.3.13) to (2.3.19) but with the specific boundary conditions

$$\delta y(x_0) = \delta y(x_f) = 0 \tag{2.3.60}$$

Thus, the correction $\delta y(x)$ and the second iterate

$$y_2(x) = y_1(x) + \delta y(x) \tag{2.3.61}$$

are easily determined. The process is repeated until the minimizing solution is found.

One disadvantage of this technique is that its range of convergence may be small in comparison with the steepest descent procedure. At any one point the correction magnitude is determined which will remove all of the error based on the quadratic approximation to the minimizing functional J. This approximation will be poor if the first guess $y_1(x)$ is not close to the minimizing arc and in such a case the iterative procedure will diverge.

To extend the range of convergence, that is, to allow convergence for the case in which $y_1(x)$ is not close to $y_{min}(x)$, a modification can be used which is essentially the same as the modification introduced in Section 2.2.2 to extend the range of convergence of the Newton-Raphson method. This extension consists of minimizing the expression in (2.3.55), but subject to a magnitude constraint on $\delta y(x)$ of the form

$$\frac{1}{2}\int_{x_0}^{x_f} (\delta y', \delta y)\,(A)\begin{pmatrix}\delta y \\ \delta y'\end{pmatrix} dx = k_3^2 \tag{2.3.62}$$

where A is the positive semi-definite matrix.

$$A = \begin{bmatrix} a_{11}(X) & a_{12}(X) \\ a_{21}(X) & a_{22}(X) \end{bmatrix} \tag{2.3.63}$$

For example, letting $a_{11} = 1$ and $a_{12} = a_{21} = a_{22} = 0$ reduces (2.3.62) to

$$\frac{1}{2}\int_{X_0}^{X_f} (\delta y')^2 \, dX = k_3^2 \tag{2.3.64}$$

which is essentially the magnitude constraint used in the steepest descent approach of Eq. (2.3.39). The problem of minimizing the expression in Eq. (2.3.55) subject to such a constraint is still a "quadratic cost" type problem and hence possesses a closed-form solution. As the minimizing arc is approached, the magnitude constraint in (2.3.62) can be relaxed and the unmodified quasilinear technique used.

2.3.3 Neighboring Extremal

Neighboring extremal, like quasilinearization, is a second-order iterative process in the sense that second-order terms in a Taylor series expansion are used to develop each succeeding estimate of the minimizing arc. Thus, $y_2(X)$ is selected to minimize a quadratic approximation of the functional J. However, unlike quasilinearization, it does not follow that y_2 provides a smaller value for J than y_1. In fact, it is equally likely that $J(y_2) > J(y_1)$. The reason for this is that both $y_1(X)$ and $y_2(X)$ are chosen to satisfy the optimizing condition (the Euler equation)

$$\frac{\partial f}{\partial y} - \frac{d}{dX}\frac{\partial f}{\partial y'} = 0$$

and the initial condition

$$y(X_0) = y_0 \tag{2.3.65}$$

The iteration consists of "bringing in" the terminal condition

$$y(X_f) = y_f \tag{2.3.66}$$

Hence, the terminal values, $y_1(X_f)$ and $y_2(X_f)$, may be such that $J(y_2) > J(y_1)$; but this is of no consequence. What is important is that the process of minimizing a quadratic approximation to $J(y)$ leads to

successive iterates which more nearly satisfy the terminal constraint of Eq. (2.3.66). The name "neighboring extremal" comes from the fact that each iterate satisfies the extremizing Euler condition.

Let $y_1(x)$ denote the first iterate which is generated as follows: Require that $y_1(x_0)$ satisfy the initial condition of (2.3.65) and guess some value for the initial slope $\frac{dy(x_0)}{dx}$. This guessed slope will be denoted by y'_{0_1}. Integrate the Euler equation

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'} = 0 \tag{2.3.67}$$

from x_0 to x_f with the initial conditions

$$y_1(x_0) = y_0$$
$$\frac{d}{dx} y_1(x_0) = y'_{0_1} \tag{2.3.68}$$

and record the value of y_1 at x_f as

$$y_1(x_f) = y_{f_1} \tag{2.3.69}$$

If y_{f_1} satisfies the correct terminal condition indicated in (2.3.66), then $y_1(x)$ is in fact the optimizing arc (since it satisfies both the Euler equation and the boundary conditions). If y_{f_1} does not satisfy (2.3.66), then a second iterate is generated with the new initial conditions

$$y_2(x_0) = y_0$$

$$\frac{d}{dx} y_2(x_0) = y'_{0_2} = y'_{0_1} + \delta y'_0 \tag{2.3.70a}$$

The change in initial slope $\delta y'_0$ is to be determined so that $y_{f_2} = y_2(x_f)$ more nearly satisfies the terminal constraint.

Note that $y_1(x)$ depends on the choice of the initial slope y'_{0_1} as indicated in (2.3.68). Thus, $y_1(x_f)$ is a function of y'_{0_1} and this is written as

$$y_{f_1} = y_{f_1}(y'_0) \tag{2.3.70b}$$

Similarly, y_f is a function of y_o'. Expanding y_f in a Taylor series about y_{f_1} and retaining only first-order terms provides

$$y_f \cong y_f(y_o') \cong = y_{f_1} + \frac{dy_f}{dy_o'}(y_o' - y_{o_1}')$$

$$= y_{f_1} + \frac{dy_f}{dy_o'}\delta y_o' \tag{2.3.71}$$

where the derivative dy_f/dy_o' denotes the rate of the change in terminal value to the change in initial slope evaluated at y_{f_1}. Hence, to first order, the correction in the initial slope should satisfy

$$\delta y_o' = \frac{y_f - y_{f_1}}{\left(\frac{dy_f}{dy_o'}\right)} \tag{2.3.72}$$

and it is this value which is to be used in (2.3.70a) for the second iterate.

In the slope correction process just outlined, the derivative $\frac{dy_f}{dy_o'}$ is required. As will be shown next, the neighboring extremal technique is essentially a method for generating this derivative.

Let $y_1(x)$ denote the initial iterate generated by means of Eqs. (2.3.67) and (2.3.68) and let $J(y_1)$ denote the corresponding value of the functional J. Expanding J in a Taylor series about $J(y_1)$ provides (to the second-order terms)

$$J(y) \cong J(y_1) + \int_{x_o}^{x_f}\left[\frac{\partial f}{\partial y}\delta y + \frac{\partial f}{\partial y'}\delta y'\right]dx$$

$$+ \frac{1}{2}\int_{x_o}^{x_f}\left[\frac{\partial^2 f}{\partial y^2}\delta y^2 + \frac{2\partial^2 f}{\partial y \partial y'}\delta y \delta y' + \frac{\partial^2 f}{\partial y'^2}\delta y'^2\right]dx \tag{2.3.73}$$

where

$$\delta y = y(x) - y_1(x)$$
$$\delta y' = y'(x) - y_1'(x) \tag{2.3.74}$$

Now, integrating the second term under the integral by parts and noting that $\delta y(x_o) = 0$ provides

$$J(y) = J(y_1) + \int_{x_0}^{x_f} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right] \delta y\, dx + \left[\frac{\partial f}{\partial y'}\right] \delta y_f$$

$$+ \frac{1}{2} \int_{x_0}^{x_f} \left[\frac{\partial^2 f}{\partial y^2}\, \delta y^2 + 2\frac{\partial^2 f}{\partial y \partial y'}\, \delta y \delta y' + \frac{\partial^2 f}{\partial y'^2}\, \delta y'^2\right] dx \qquad (2.3.75)$$

But, since $y_1(x)$ satisfies the Euler condition, the first integral expression is zero leaving

$$J(y) = J(y_1) + \left(\frac{\partial f}{\partial y'}\right)_{x_f} \delta y_f + \int_{x_0}^{x_f} F(x, \delta y, \delta y')\, dx \qquad (2.3.76)$$

where

$$F = \frac{1}{2}\left[\frac{\partial^2 f}{\partial y^2}\, \delta y^2 + 2\frac{\partial^2 f}{\partial y \partial y'}\, \delta y \delta y' + \frac{\partial^2 f}{\partial y'^2}\, \delta y'^2\right] \qquad (2.3.77)$$

Now, δy_f is to satisfy

$$\delta y_f = y_f - y_{f_1} \qquad (2.3.78)$$

that is, δy_f is equal to the difference between the specified terminal value of y and that value use on the first iterate, y_{f_1}. Under this condition, the minimization of the expression in (2.3.76) is accomplished by requiring $\delta y(x)$ to satisfy the Euler condition

$$\frac{\partial F}{\partial(\delta y)} - \frac{d}{dx}\left(\frac{\partial F}{\partial(\delta y')}\right) = 0 \qquad (2.3.79)$$

together with the boundary conditions

$$\begin{aligned} \delta y(x_0) &= \delta y_0 = 0 \\ \delta y(x_f) &= \delta y_f = y_f - y_{f_1} \end{aligned} \qquad (2.3.80)$$

Substituting (2.3.77) into (2.3.79) provides

$$\delta y'' a(x) + \delta y' (a'(x)) + \delta y (b'(x) - c(x)) = 0 \qquad (2.3.81)$$

where

$$a = \frac{\partial^2 f}{\partial y'^2} \qquad b = \frac{\partial^2 f}{\partial y \partial y'} \qquad c = \frac{\partial^2 f}{\partial y^2} \tag{2.3.82}$$

At this point, comparison of this expression with Eqs. (2.3.12) to (2.3.18) indicates the solution

$$\delta y(x) = \phi_{11}(x)\delta y_0 + \phi_{12}(x)\delta y_0' \tag{2.3.83}$$

where the calculation of ϕ_{11} and ϕ_{12} is given in Eq. (2.3.18). Finally, evaluating (2.3.83) at x_f and using the boundary conditions of (2.3.80) yields the result

$$\delta y_0' = \frac{y_f - y_{f_1}}{\phi_{12}(x_f)} \tag{2.3.84}$$

Hence, the second iterate $y_2(x)$ is chosen to satisfy the Euler condition of (2.3.67) but with the initial values

$$\begin{aligned} y_2(x_0) &= y_0 \\ y_2'(x_0) &= y_{0_2}' = y_{0_1}' + \delta y_0' \end{aligned} \tag{2.3.85}$$

where $\delta y_0'$ is given in (2.3.84). A comparison of (2.3.84) with (2.3.72) indicates the identify

$$\left(\frac{dy_f}{dy_0'}\right)_{y_{f_1}} = \phi_{12}(x_f) \tag{2.3.86}$$

2.3.4 Discussion of Results

The steepest descent technique is a first-order technique and, like the equivalent procedure in maxima-minima theory, requires that the user guess the magnitude of correction at each step (the quantity k_1 in Eq. (2.3.28)). The process has a relatively wide range of convergence in that it can be made to converge from starting solutions, that are far removed from the minimizing solution. However, the technique breaks down in the vicinity of the optimal solution since the correction mechanism, the Euler operator $\left(\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y'}\right)$

approaches zero, thereby, making the selection of the constant k_1 extremely difficult.

In contrast, both quasilinearization and neighboring extremal are second-order methods and converge rapidly in the vicinity of the minimizing solution. However, both techniques tend to diverge if the starting condition is not sufficiently close to the minimizing solution since both techniques attempt to remove all the error in one step - a procedure which works well if the error is small, but fails if the error is large. Fortunately, by the inclusion of a constraint on the magnitude of correction, the range of convergence of both techniques can be extended. Such a procedure for the quasilinear approach was outlined in Section 2.3.2 (see Eq. (2.3.62)). An equivalent condition for neighboring extremal would simply replace the terminal constraint of Eq. (2.3.80).

$$\delta y_f = y_f - y_{f_1} \tag{2.3.87}$$

by

$$\delta y_f = k_1 (y_f - y_{f_1}) \tag{2.3.88}$$

where k_1 is some number between zero and unity. By this device only a part of the terminal error would be removed on each iteration and an enlargement of the range of convergence would result.

In the neighboring extremal technique it was noted that the second iterate $y_2(x)$ did not necessarily provide a smaller value for the functional J than $y_1(x)$. The reason for this observation is that the value of the functional J is not a proper measure of closeness of a particular iterate to its optimum value. A better measure is the difference between the specified terminal condition and the terminal value provided by the particular iterate.

A similar situation occurs when both quasilinearization and neighboring extremal are used to solve the generalized Bolza problem which was formulated in Section 2.1; that is, successive iterates do not necessarily decrease the functional which is to be minimized. The reason for this is that in neither method do successive iterates satisfy all the constraint conditions (just as the terminal constraint was violated in the neighboring extremal approach of Section 2.3.3). Thus, here also, the value of the functional to be minimized is not necessarily a proper measure of the closeness of a particular iterated solution to the desired extremum. However, the corrections are still chosen to minimize a quadratic approximation to the minimizing functional, since by this device the errors in satisfying the specified constraints are automatically "brought in."

2.4 CLASSICAL MAYER PROBLEM

The Mayer Problem consists of minimizing the functional

$$J = \phi(x_f) \tag{2.4.1}$$

subject to the differential constraints

$$\dot{x} = f(x, u, t) \Longleftrightarrow \dot{x}_i = f_i(x, u, t); \quad i = 1, n \tag{2.4.2}$$

the specified initial state

$$x = x_o \quad AT \quad t = t_o \tag{2.4.3}$$

and the M terminal conditions

$$\psi(x) = \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \\ \cdot \\ \cdot \\ \psi_M(x) \end{bmatrix} = 0 \quad AT \quad t = t_f \tag{2.4.4}$$

This minimization is accomplished through the appropriate selection of the r dimensional control vector u which is required to lie in the control set U.

As indicated in Section 2.1, this problem is solved once a solution has been generated to the 2n equations

$$\begin{aligned} \dot{x} &= f(x, u, t) \Longrightarrow \dot{x}_i = f_i(x, u, t) \\ \dot{p} &= -\frac{\partial H}{\partial x} \Longrightarrow \dot{p}_i = -\frac{\partial H}{\partial x_i} \\ H &= p^T f \Longrightarrow H = \sum_{i=1}^{n} p_i f_i \end{aligned} \tag{2.4.5}$$

which satisfies the boundary conditions

$$x = x_o \ AT \ t = t_o \tag{2.4.6}$$

$$\psi(x) = 0 \iff \psi_j(x) = 0 \; ; \; j = 1, M$$

$$P + \left(\frac{\partial \psi}{\partial x}\right)^T \mu + \frac{\partial \phi}{\partial x} = 0 \iff P_i + \sum_{j=1}^{M} \mu_j \frac{\partial \psi_j}{\partial x_i} + \frac{\partial \phi}{\partial x_i} = 0; \; i = 1, n \tag{2.4.7}$$

If the final time t_f is not explicitly specified, then the additional condition

$$H = P^T f = 0 \quad AT \quad t = t_f \tag{2.4.8}$$

must also hold. The optimal control is chosen as that control in the set $\mathcal{U}$ which maximizes the Hamiltonian; that is

$$H(x, u_{opt}, P, t) \geq H(x, \hat{u}, P, t) \tag{2.4.9}$$

where u_{OPT} denotes the optimal control and $\hat{u}$ is any other control both of which must lie in $\mathcal{U}$.

In this section the three numerical procedures, steepest descent, neighboring extremal and quasilinearization, will be used to solve this problem but under two additional assumptions.

1. It will be assumed that the control set $\mathcal{U}$ is the entire r dimensional control space; that is, no bound is placed on the control and each component u_i $(i = 1, r)$ can take on any value between $-\infty$ and $+\infty$. Under this assumption the optimizing condition in (2.4.9) reduces to the r equations

$$\frac{\partial H}{\partial u} = \begin{bmatrix} \frac{\partial H}{\partial u_1} \\ \frac{\partial H}{\partial u_2} \\ \vdots \\ \frac{\partial H}{\partial u_r} \end{bmatrix} = 0 \tag{2.4.10}$$

These equations are the classical Euler equations for the control action (u) and are just sufficient in number to determine the r control components at each point along the optimal trajectory.

2. It will be assumed that the final time t_f is explicitly specified. An extension to the case in which the final time is allowed to vary will be indicated in Section 2.5.

As in both the maxima-minima and Lagrange problems, only a very special form of the problem of Mayer can be solved directly. This class is referred to as the "quadratic cost" problem and will be treated next.

2.4.1 Quadratic Cost Problem

Let the state equations and boundary conditions take the convenient linear form

$$\dot{X} = A(t)X + G(t)u \Longrightarrow \begin{bmatrix} \dot{x}_1 \\ \dot{x}_2 \\ \vdots \\ \dot{x}_n \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \cdots a_{1n} \\ a_{21} & a_{22} \quad a_{2n} \\ \vdots \\ a_{n1} & a_{n2} \quad a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} g_{11} & g_{12} \cdots g_{1r} \\ g_{21} & g_{22} \quad g_{2r} \\ \vdots \\ g_{n1} & g_{n2} \quad g_{nr} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{bmatrix} \tag{2.4.11A}$$

$$X = X_0 \quad At \quad t = t_0$$

$$\psi = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_m \end{bmatrix} = Q = CX - d = 0 \Longrightarrow \begin{bmatrix} c_{11} & c_{12} \cdots c_{1n} \\ c_{21} & c_{22} \cdots c_{2n} \\ \vdots \\ c_{M1} & c_{M2} \cdots c_{Mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} - \begin{bmatrix} d_1 \\ d_2 \\ \vdots \\ d_M \end{bmatrix} = 0 \tag{2.4.11B}$$

Note that A and G are $n \times n$ and $n \times r$ time dependent matrices, respectively, while C and d are $M \times n$ and $M \times 1$ constant matrices. The admissible control set $\mathcal{U}$ is the entire control space with each component u_i $(i = 1, r)$ allowed to take any value between $-\infty$ and $+\infty$. The problem is to select the control history which minimizes the functional

$$^{*}J = b^T x_f + \int_{t_0}^{t_f} (x^T Q_1 x + u^T Q_2 u)\, dt$$

$$= \sum_{i=1}^{n} b_i x_{if} + \int_{t_0}^{t_f} \left[\sum_{i,j=1}^{n} g^{(1)}_{ij} x_i x_j + \sum_{i,j=1}^{r} g^{(2)}_{ij} u_i u_j \right] dt \tag{2.4.11C}$$

* It should be noted that linear terms in x and u can also be included in the integrand of this performance index without destroying the "quadratic cost" property of the problem.

where Q_1 is a symmetric $n \times n$ positive semi-definite matrix (i.e., $x^T Q_1 x \geq 0$) and Q_2 is an $r \times r$ positive definite matrix. The quantity b is an n dimensional constant vector. This problem is frequently referred to in the literature as the "linear dynamics - quadratic cost" problem. Note that the performance index J is a linear function of the terminal state (x_f) and a quadratic function of x and u of the interval (t_0, t_f). As will be shown next, the solution to this problem can be developed without iteration.

The problem stated to this point can be put in the standard Mayer form by introducing the new state variable z defined as

$$z = \int_{t_0}^{t_f} (x^T Q_1 x + u^T Q_2 u)\, dt \tag{2.4.11D}$$

Hence,

$$\dot{z} = x^T Q_1 x + u^T Q_2 u \tag{2.4.12A}$$

$$z(t_0) = 0 \tag{2.4.12B}$$

Thus, using the new state variables $\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \\ z \end{bmatrix}$

which satisfy the differential constraints of Eqs. (2.4.11A) and (2.4.12A) together with the boundary conditions of (2.4.11B), the problem becomes one of selecting the control u to minimize the functional J where

$$J = \phi(x_f, z_f) = z_f + \sum_{i=1}^{n} b_i x_{if} \tag{2.4.12C}$$

This problem is now in the standard Mayer form to which the Maximum Principle conditions of Eqs. (2.4.5) to (2.4.9) can be applied.

This solution is facilitated by introducing the $n+1$ dimensional vector $\begin{pmatrix} p \\ p_z \end{pmatrix}$

$$\begin{pmatrix} p \\ p_z \end{pmatrix} = \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \\ p_z \end{bmatrix}$$

where the first n components of p are the adjoint variables associated with the state $x_i\ (i=1,n)$ and the $n+1^{st}$ component, p_z, is associated with the state variable z. Thus, the Hamiltonian becomes

$$H = p^T \dot{x} + p_z \dot{z} = \sum_{i=1}^{n} p_i \dot{x}_i + p_z \dot{z}$$

$$= p^T (Ax + Gu) + p_z (x^T Q_1 x + u^T Q_2 u) \qquad (2.4.13A)$$

Using Eq. (2.4.7) with ψ and ϕ given by (2.4.11B) and (2.4.12C), it follows that the $\begin{pmatrix} p \\ p_z \end{pmatrix}$ vector must satisfy the boundary conditions

$$P + C^T \mu + b = 0 \Longleftrightarrow \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} + \begin{bmatrix} c_{11} & c_{21} & \cdots & c_{M1} \\ c_{12} & c_{22} & & c_{M2} \\ \vdots & & & \\ c_{1n} & c_{2n} & \cdots & c_{Mn} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix} + \begin{bmatrix} b_1 \\ b_2 \\ \\ b_n \end{bmatrix} = 0 \qquad (2.4.13B)$$

$$AT \quad t = t_f$$

$$p_z + 1 = 0 \quad ; \quad AT \quad t = t_f \qquad (2.4.13C)$$

Also, the vector $\begin{pmatrix} p \\ p_z \end{pmatrix}$ must satisfy the differential equations

$$\dot{p}_i = -\frac{\partial H}{\partial x_i} \qquad i = 1, n \qquad (2.4.14)$$

Further, since from (2.4.13A) H is not an explicit function of the state z, it follows that

$$\dot{p}_z = -\frac{\partial H}{\partial z} = 0$$

Combining this last result with (2.4.13C) yields the first of the multipliers as

$$p_z = -1$$

Under this condition, Eq. (2.4.14) reduces to

$$\dot{P} = -A^T P + 2Q_1 X \Longleftrightarrow \begin{bmatrix} \dot{P}_1 \\ \dot{P}_2 \\ \vdots \\ \dot{P}_n \end{bmatrix} = -\begin{bmatrix} a_{11} & a_{21} \cdots a_{n1} \\ a_{12} & a_{22} \quad a_{n2} \\ \vdots & \\ a_{1n} & a_{2n} \quad a_{nn} \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} + 2 \begin{bmatrix} q_{11}^{(i)} & q_{12}^{(i)} \cdots q_{1n}^{(i)} \\ q_{21}^{(i)} & q_{22}^{(i)} \cdots q_{2n}^{(i)} \\ \vdots & \vdots \\ q_{n1}^{(i)} & q_{n2}^{(i)} \cdots q_{nn}^{(i)} \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{bmatrix}$$

(2.4.15A)

Finally, to maximize the Hamiltonian, the control must be selected so that

$$\frac{\partial H}{\partial U} = 0 \Longleftrightarrow \frac{\partial H}{\partial u_i} = 0 \; ; \; i = 1, r$$

This condition leads to the optimizing control

$$U = \frac{Q_2^{-1} G^T P}{2} \Longleftrightarrow \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_r \end{bmatrix} = \frac{(Q_2^{-1})(G)^T}{2} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix}$$

(2.4.15B)

Combining results for the quadratic cost problem, the state and adjoint vectors satisfy the differential equations

$$\dot{X} = AX + GU$$
$$\dot{P} = -A^T P + 2Q_1 X$$

(2.4.16A)

and the boundary conditions

$$X = X_o \quad AT \quad t = t_o$$

$$\left.\begin{aligned} Cx - d &= 0 \\ P + C^T \mu + b &= 0 \end{aligned}\right\} \quad AT \; t = t_f$$

(2.4.16B)

The optimal control satisfies

$$u = \frac{Q_2^{-1} G^T P}{2} \tag{2.4.16C}$$

Substituting (2.4.16C) into (2.4.16A) leads to the $2n$ linear system

$$\begin{aligned} \dot{X} &= AX + \frac{G Q_2^{-1} G^T P}{2} \\ \dot{P} &= -A^T P + 2Q_1 X \end{aligned} \tag{2.4.16D}$$

and since the boundary conditions of (2.4.16B) are linear, this system has a closed-form solution.

Let $\Lambda(t)$ denote the $2n \times 2n$ matrix solution to the differential set

$$\dot{\Lambda} = \begin{bmatrix} A & \frac{G Q_2^{-1} G^T}{2} \\ 2Q_1 & -A^T \end{bmatrix} \Lambda \; ; \quad \Lambda(t_0) = I \tag{2.4.17A}$$

and let Λ be partitioned into the four $n \times n$ matrices

$$\Lambda(t) = \begin{bmatrix} \Lambda^{(1)} & \Lambda^{(2)} \\ \Lambda^{(3)} & \Lambda^{(4)} \end{bmatrix}$$

Then (2.4.16D) has the solution representation

$$\begin{bmatrix} X(t) \\ P(t) \end{bmatrix} = \Lambda(t) \begin{bmatrix} X_0 \\ P_0 \end{bmatrix} \tag{2.4.17B}$$

Thus, the terminal point can be evaluated as

$$\begin{bmatrix} X_f \\ P_f \end{bmatrix} = \Lambda(t_f) \begin{bmatrix} X_o \\ P_o \end{bmatrix}$$

or using the partitioned form of Λ this equation becomes

$$\begin{aligned} X_f &= \Lambda^{①}(t_f) X_o + \Lambda^{②}(t_f) P_o \\ P_f &= \Lambda^{③}(t_f) X_o + \Lambda^{④}(t_f) P_o \end{aligned}$$

(2.4.18A)

Combining this expression with the terminal constraints of (2.4.16B) provides

$$\begin{bmatrix} C\Lambda^{②}(t_f) & 0 \\ \Lambda^{④}(t_f) & C^T \end{bmatrix} \begin{bmatrix} P_o \\ \mu \end{bmatrix} = \begin{bmatrix} d - C\Lambda^{①}(t_f) X_o \\ -\Lambda^{③}(t_f) X_o - b \end{bmatrix}$$

(2.4.18B)

Now, since X_o is known, this equation can be used to determine P_o. Substitution of P_o back into Eq. (2.4.17B) then provides the solution to the quadratic cost optimization problem.

As in the maxima-minima and one-dimensional Lagrange case, the quadratic problem yields to the solution in one run on a digital computer (no iteration is required). Hence, the numerical procedures used in the solution of non-linear problems consist of approximating the problem by a series of quadratic cost type problems. The particular iterative technique used is a function only of the type of approximation which is made. In all cases, the procedures used are essentially the same as those used in connection with the maxima-minima and Lagrange problems.

2.4.2 Steepest Descent

The starting iterate in the steepest descent technique, which is denoted by $u_1(t)$, must satisfy all the problem constraints; that is, it must drive the system

$$\dot{x} = f(x, u, t) \qquad (2.4.19A)$$

from the initial point

$$X = X_o \quad AT \quad t = t_o \tag{2.4.19B}$$

to the specified terminal set

$$\psi = \begin{vmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_M \end{vmatrix} = 0 \quad AT \quad t = t_f \tag{2.4.19C}$$

However, $u_1(t)$ is not the optimal solution in that it does not minimize the functional

$$J = \phi(X_f)$$

The problem is to select the second iterate

$$u_2(t) = u_1(t) + \delta u(t)$$

so that it will also satisfy the constraint conditions of Eqs. (2.4.19A) to (2.4.19C) and more nearly minimizes the functional J.

Rewrite the functional J as

$$\bar{J} = \phi(X_f) + \psi^T \nu + \int_{t_o}^{t_f} P^T(\dot{x} - f(x,u,t))\,dt \equiv J \tag{2.4.20A}$$

where P and ν are multipliers to be selected so that the constraints of Eqs. (2.4.19A) and (2.4.19C) are satisfied. Since each solution will start at the point x_o at time t_o, there is no need to explicitly include the constraint condition of Eq. (2.4.19B). Note that since $\bar{J}$ is equal to J (only terms equal to zero have been added on), minimizing the functional $\bar{J}$ in (2.4.20A) is equivalent to minimizing the functional J, and in what follows, attention will be focused exclusively on the functional $\bar{J}$.

Now, $u_1(t)$ denotes the first iterate which drives the system from the fixed-initial point to the terminal set $\psi = 0$ <u>at the fixed-terminal time t_f</u>. Hence, let $\bar{J}(u_1)$ denote the corresponding value of the functional in (2.4.20A). Expanding $\bar{J}(u)$ about $\bar{J}(u_1)$ in a Taylor series, and truncating after the first-order terms provides

$$\bar{J}(u) - \bar{J}(u_1) \cong \delta \bar{J}$$

$$= \left[\frac{\partial \phi}{\partial x} + \frac{\partial \psi^T}{\partial x}\mu\right]^T \delta x + \int_{t_0}^{t_f} p^T \left[\delta \dot{x} - \frac{\partial f}{\partial x}\delta x - \frac{\partial f}{\partial u}\delta u\right] dt$$

But the first term under the integral can be integrated by parts and combined with the definition of the Hamiltonian

$$H = p^T f = \sum_{j=1}^{n} p_i f_i$$

to yield

$$\delta \bar{J} = \left\{\frac{\partial \phi}{\partial x} + \frac{\partial \psi^T}{\partial x}\mu + p\right\}^T \delta x - \int_{t_0}^{t_f} \left[\left(\dot{p} + \frac{\partial H}{\partial x}\right)^T \delta x + \left(\frac{\partial H}{\partial u}\right)^T \delta u\right] dt \qquad (2.4.20B)$$

Now, noting that the n dimensional vector p satisfies the differential equations

$$\dot{p} = -\frac{\partial H}{\partial x} \Longleftrightarrow \dot{p}_i = \frac{-\partial H}{\partial x_i} = -\sum_{j=1}^{n} p_j \frac{\partial f_j}{\partial x_i} \qquad (2.4.21A)$$

and the boundary conditions

$$p + \frac{\partial \phi}{\partial x} + \frac{\partial \psi}{\partial x}^T \mu = 0 \Longleftrightarrow p_i + \frac{\partial \phi}{\partial x_i} + \sum_{j=1}^{M} \mu_j \frac{\partial \psi_j}{\partial x_i} = 0 \qquad (2.4.21B)$$
$$i = 1, n$$

allows Eq. (2.4.20B) to be written as

$$\bar{J}(u) - \bar{J}(u_1) \cong \delta \bar{J} = \int_{t_0}^{t_f} -\left(\frac{\partial H}{\partial u}\right)^T \delta u \, dt \qquad (2.4.22)$$

Now, minimizing $\bar{J}$ is equivalent to minimizing J ; thus, if the amount by which the control can change is limited, then minimizing the first-order approximation to $\bar{J}$ should provide a second iterate $u_2 = u_1 + \delta u$ which is more nearly optimal. Hence, require that

$$\int_{t_0}^{t_f} \delta u^2 \, dt = \epsilon_1^2 \tag{2.4.23}$$

where ϵ_1 is some small positive quantity, and that $\delta \bar{J}$ is to be minimized subject to this condition.

Proceeding formally, the scalar multiplier λ is introduced and the functional $\delta \bar{\bar{J}}$ formed where

$$\delta \bar{\bar{J}} = \delta \bar{J} + \lambda \left\{ \int_{t_0}^{t_f} \delta u^2 \, dt - \epsilon_1^2 \right\} = \int_{t_0}^{t_f} \left\{ \lambda \delta u^2 - \left(\frac{\partial H}{\partial u} \right)^T \delta u \right\} dt - \lambda \epsilon_1^2$$

Again, requiring the first variation of $\delta \bar{\bar{J}}$ to vanish provides

$$\int_{t_0}^{t_f} \left\{ 2\lambda \, \delta u - \frac{\partial H}{\partial u} \right\}^T \delta u \, dt = 0 \tag{2.4.24}$$

which is equivalent to

$$2\lambda \, \delta u - \frac{\partial H}{\partial u} = 0 \iff \delta u_i = \frac{1}{2\lambda} \frac{\partial H}{\partial u_i} = \frac{1}{2\lambda} \sum_{j=1}^{n} p_j \frac{\partial f_j}{\partial u_i} \qquad i = 1, r \tag{2.4.25}$$

Now, the adjoint vector p must satisfy the differential and boundary conditions of Eqs. (2.4.21A) and (2.4.21B). For an integration to begin, however, the M dimensional vector μ must be known. Recall that μ is selected so that

$$\psi_r(x_f) = \begin{bmatrix} \psi_1 \\ \psi_2 \\ \vdots \\ \psi_M \end{bmatrix} = 0$$

or to the first order these constraints are

$$\left[\left(\frac{\partial \psi_r}{\partial x} \right)^T \delta x = 0 \right]^{t=t_f} \iff \begin{bmatrix} \frac{\partial \psi_1}{\partial x_1} & \frac{\partial \psi_1}{\partial x_2} & \cdots & \frac{\partial \psi_1}{\partial x_n} \\ \frac{\partial \psi_2}{\partial x_1} & \frac{\partial \psi_2}{\partial x_2} & \cdots & \frac{\partial \psi_2}{\partial x_n} \\ \vdots & & & \vdots \\ \frac{\partial \psi_M}{\partial x_1} & \frac{\partial \psi_M}{\partial x_2} & \cdots & \frac{\partial \psi_M}{\partial x_n} \end{bmatrix} \begin{bmatrix} \delta x_1 \\ \delta x_2 \\ \vdots \\ \delta x_n \end{bmatrix} = 0 \tag{2.4.26}$$

The selection of $\mathcal{M}$ to satisfy Eq. (2.4.26) is accomplished by integrating the system

$$\dot{p} = - \frac{\partial H}{\partial x}$$

from t_f to t_o $M+1$ times with the $M+1$ different terminal conditions

$$\left.\begin{aligned}
p^{(1)} &= - \frac{\partial \psi_1}{\partial x} \iff p_i^{(1)} = - \frac{\partial \psi_1}{\partial x_i} \qquad i = 1, n \\
p^{(2)} &= - \frac{\partial \psi_2}{\partial x} \iff p_i^{(2)} = - \frac{\partial \psi_2}{\partial x_i} \\
&\vdots \\
p^{(M)} &= - \frac{\partial \psi_M}{\partial x} \iff p_i^{(M)} = - \frac{\partial \psi_M}{\partial x_i} \\
p^{(M+1)} &= - \frac{\partial \phi}{\partial x} \iff p^{(M+1)} = - \frac{\partial \phi}{\partial x_i}
\end{aligned}\right\} \quad At \quad t = t_f \qquad (2.4.27)$$

Now, let

$$p(t) = p^{(M+1)}(t) + \sum_{j=1}^{M} \mu_j \, p^{(j)}(t) \qquad (2.4.28A)$$

Since the linearized form of the state equation is given by

$$\delta \dot{x} = \frac{\partial f}{\partial x} \delta x + \frac{\partial f}{\partial u} \delta u \qquad (2.4.28B)$$

it is a simple matter to demonstrate the identity

$$\frac{d}{dt}\left\{ p^T \delta x \right\} = \left(\frac{\partial H}{\partial u} \right)^T \delta u = \left(p^T \frac{\partial f}{\partial u} \right)^T \delta u \qquad (2.4.29)$$

for any vector p satisfying the differential condition in (2.4.21A).

Hence, integrating this expression between t_o and t_f and noting that $\delta x_o = 0$ provides

$$p_f^T \delta x_f = \int_{t_o}^{t_f} \left\{ p^T \frac{\partial f}{\partial u} \right\}^T \delta u \, dt \qquad (2.4.30)$$

Now, substituting each of the vectors $p^{(i)}$ $(i = 1, M)$ into (2.4.30) and using the boundary conditions of (2.4.26) provides the M equations

$$-\left(\frac{\partial \psi_i}{\partial x}\right)^T \delta x \Big\}_{t_0}^{t_f} = 0 = \int_{t_0}^{t_f} \left\{ p^{(i)T} \frac{\partial f}{\partial u} \right\}^T \delta u \, dt \; ; \; i = 1, M \qquad (2.4.31)$$

Combination of this result with the expression for δu in Eq. (2.4.25) yields

$$-\left(\frac{\partial \psi_i}{\partial x}\right)^T \delta x \Big]_{t=t_f} = 0 \Longrightarrow \int_{t_0}^{t_f} \left\{ p^{(i)} \frac{\partial f}{\partial u} \right\}^T \left(p^{(M+1)} + \sum_{j=1}^{M} \mu_j p^{(j)} \right)^T \frac{\partial f}{\partial u} \, dt = 0 \qquad (2.4.32A)$$

This set can be used to determine the appropriate values for the μ_j $(j = 1, M)$. Specifically, these equations are equivalent to the M linear equations

$$A\mu = b \Longleftrightarrow \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1M} \\ a_{21} & a_{22} & \cdots & a_{2M} \\ \vdots & & & \vdots \\ a_M & a_{M2} & \cdots & a_{MM} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{bmatrix} \qquad (2.4.32B)$$

where

$$a_{ij} = \left(p^{(i)T} \frac{\partial f}{\partial u} \right)^T \left(p^{(j)T} \frac{\partial f}{\partial u} \right) \qquad (2.4.33A)$$

$$b_i = -\left(p^{(i)T} \frac{\partial f}{\partial u} \right) \left(p^{(M+1)T} \frac{\partial f}{\partial u} \right) \qquad (2.4.33B)$$

With the multiplier μ determined from these equations, the exact value of the p vector can now be computed using Eq. (2.4.28A). The scalar multiplier λ is then evaluated using Eq. (2.4.25) and the constraint condition of (2.4.23). This operation yields

$$\lambda = \frac{1}{2k_1}\left\{\int_{t_0}^{t_f}\left(p^T\frac{\partial f}{\partial u}\right)^2 dt\right\}^{1/2} \tag{2.4.34}$$

Finally, combining Eq. (2.4.30) with Eqs. (2.4.27), (2.4.32) and (2.4.34) yields

$$p^T\delta x\Big]_{t_f} = -\left(\frac{\partial\phi}{\partial x} + \frac{\partial\psi^T}{\partial x}\mu\right)^T_{t_f}\delta x = -\left(\frac{\partial\phi}{\partial x}\right)^T_{t_f}\delta x = -d\phi = \frac{1}{2\lambda}\int_{t_0}^{t_f}\left(p^T\frac{\partial f}{\partial u}\right)^2 dt$$

Thus

$$d\phi = -k_1\left\{\int_{t_0}^{t_f}\left(p^T\frac{\partial f}{\partial u}\right)^2 dt\right\}^{1/2} \tag{2.4.35}$$

and for k_1 sufficiently small, the new value of the performance index $\phi(x_f)$ will be smaller than the previous value.

The step by step calculation procedure to be used in the steepest descent method is as follows:

(1) Select an initial control program, $u_1(t)$, which will drive the system
$\dot{x} = f(x, u, t)$
from the initial point x_0 to the terminal set $\psi_i(x_f) = 0 \; (i = 1, m)$.

(2) Integrate the state system
$\dot{x} = f(x, u, t)$
forward from t_0 to t_f

(3) Integrate the adjoint system
$\dot{p} = -\frac{\partial H}{\partial x}$
backwards $m+1$ times with the terminal conditions

$$p_f^{(1)} = -\frac{\partial \psi_1}{\partial x}$$

$$p_f^{(2)} = -\frac{\partial \psi_2}{\partial x}$$

$$\vdots$$

$$p_f^{(M)} = -\frac{\partial \psi_M}{\partial x}$$

$$p_f^{(M+1)} = -\frac{\partial \phi}{\partial x}$$

(4) Compute μ from Eqs. (2.4.32B) where the quantities a_{ij} and b_i are evaluated from the results of Steps (1) to (3).

(5) Compute ρ and λ from Eqs. (2.4.28A) and (2.4.34).

(6) Use Eqs. (2.4.25) to develop the control correction $\delta u = \frac{1}{2\lambda} \rho^T \frac{\partial f}{\partial u}$

(7) Set $u_2(t) = u_1(t) + \delta u(t)$

(8) Go to Step (2)

The iteration continues until additional changes in the control program produce no improvement in the performance index. At this point the optimizing condition

$$\frac{\partial H}{\partial u} = 0 \iff \frac{\partial H}{\partial u_i} = 0 \; ; \qquad i = 1, r$$

has essentially been reached and the process is terminated. Additional details of the steepest descent procedure are given in Refs. (12) and (13).

In this iteration process the second iterate is chosen to minimize the first-order approximation to the modified functional $\bar{J}$ (defined in (2.4.20A)) subject to the magnitude constraint of Eq. (2.4.33). This first approximation is a quadratic cost problem of the type discussed in Section (2.4.1). However, in the approximation and minimization processes conducted in Eqs. (2.4.20B) to (2.4.25) inclusive, the quadratic nature of the problem is somewhat obscure. Hence, it is worthwhile to give an alternate derivation of Eqs. (2.4.20B) to (2.4.25) in which the quadratic nature of the problem is more apparent, and the solution developed in Section 2.4.1 can be applied directly.

The quantity

$$J = \phi(x_f) \tag{2.4.36}$$

is to be minimized subject to the differential constraints

$$\dot{x} = f(x, u, t) \tag{2.4.37A}$$

and boundary conditions

$$x = x_o \quad at \quad t = t_o$$

$$\psi_i(x) = 0 \quad at \quad t = t_f \; ; \qquad i = 1, M \tag{2.4.37B}$$

The first iterate satisfies the constraint conditions but does not minimize the functional J in (2.4.36). Hence, to first-order terms, the second iterate is to be selected to minimize

$$J(u) - J(u_1) \cong \delta J = \left(\frac{\partial \phi}{\partial x}\right)^T \delta x_f \tag{2.4.38}$$

subject to the linearized differential constraints

$$\delta \dot{x} = \left(\frac{\partial f}{\partial x}\right)^T \delta x + \left(\frac{\partial f}{\partial u}\right)^T \delta u \tag{2.4.39A}$$

the linearized boundary conditions

$$\delta x_o = 0 \qquad at \quad t = t_o$$

$$\left(\frac{\partial \psi}{\partial x}\right)\delta x_f = 0 \iff \begin{bmatrix} \frac{\partial \psi_1}{\partial x_1} & \frac{\partial \psi_1}{\partial x_2} & \cdots & \frac{\partial \psi_1}{\partial x_n} \\ \frac{\partial \psi_2}{\partial x_1} & \frac{\partial \psi_2}{\partial x_2} & \cdots & \frac{\partial \psi_2}{\partial x_n} \\ \vdots & & & \\ \frac{\partial \psi_M}{\partial x_1} & \frac{\partial \psi_M}{\partial x_2} & \cdots & \frac{\partial \psi_M}{\partial x_n} \end{bmatrix} \begin{bmatrix} \delta x_{1_f} \\ \delta x_{2_f} \\ \vdots \\ \delta x_{n_f} \end{bmatrix} = 0 \tag{2.4.39B}$$

and the magnitude constraint

$$\int_{t_o}^{t_f} (\delta u)^2 \, dt = k_1^2 \tag{2.4.39C}$$

If the constraint in (2.4.39C) is adjoined to the linearized functional in (2.4.38) through a constant multiplier λ, then the problem becomes one of minimizing the quantity

$$\delta\hat{J} = \left(\frac{\partial \phi}{\partial x}\right)^T \delta x_f + \lambda \left\{ \int_{t_0}^{t_f} \delta u^2 \, dt - \ell_1^2 \right\} \tag{2.4.40}$$

subject to the differential constraints of (2.4.39A) and the boundary conditions of (2.4.39B).

This problem is the quadratic cost type problem of Eqs. (2.4.11A), (2.4.11B) and (2.4.11C), but with

$$b = \frac{\partial \phi}{\partial x} \; ; \quad A = \left(\frac{\partial f}{\partial x}\right)^T \; ; \quad G = \left(\frac{\partial f}{\partial u}\right)^T$$

$$c = \frac{\partial \psi}{\partial x} \; ; \quad d = 0 \; ; \quad Q_1 = 0 \; , \quad Q_2 = \lambda I \tag{2.4.41}$$

Hence, the adjoint equations and boundary conditions (see Eqs. (2.4.13B) and (2.4.15A)) take the form

$$\dot{p} = - \frac{\partial H}{\partial x}$$

$$p + \frac{\partial \phi}{\partial x} + \left(\frac{\partial \psi^T}{\partial x}\right)\mu = 0 \tag{2.4.42}$$

with the optimal control satisfying

$$\delta u = \frac{1}{2\lambda} \left(p^T \frac{\partial f}{\partial u} \right) \tag{2.4.43}$$

The methods of Section 2.4.1 can now be used to develop the solution.

Note that the required conditions in (2.4.42) and (2.4.43) are identical to those used in Eqs. (2.4.21A), (2.4.21B) and (2.4.25). (These equations were used in the calculation of the steepest descent correction.) This discussion establishes the quadratic nature of the auxiliary minimization problem used in the steepest descent procedure. More important, it also establishes that minimizing the first-order approximation to the modified functional $\bar{J}$

$$\bar{J} = \phi(x_f) + \psi^T \mu + \int_{t_o}^{t_f} p^T (\dot{x} - f)\, dt \qquad (2.4.43A)$$

is equivalent to minimizing the first-order approximation to $J = \phi(x_f)$ subject to linearized versions of the boundary and differential constraints.

In the quasilinear and neighboring extremal techniques to be discussed subsequently, a similar procedure will be used. That is, second iterates will be developed by minimizing some approximation to the functional $\bar{J}$ in Eq. (2.4.43A). This approach is algebraically cleaner than minimizing an approximation to $\bar{J}$ subject to linearized constraints. However, in taking this approach, the quadratic nature of the auxiliary minimization problem is obscured.

2.4.3 Neighboring Extremal

The optimization problem is solved once a solution has been generated to the $2n$ system of equations

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

$$\dot{p} = -\frac{\partial H}{\partial x} \qquad (2.4.44)$$

which satisfies the $2n$ boundary conditions

$$x = x_o \quad at \quad t = t_o \qquad (2.4.45)$$

$$\frac{\partial \phi}{\partial x} + \frac{\partial \psi^T}{\partial x} \mu + p = 0 \quad at \quad t = t_f \qquad (2.4.46)$$

where the multiplier μ is determined so that

$$\psi(x_f) = \begin{bmatrix} \psi_1(x_f) \\ \psi_2(x_f) \\ \vdots \\ \psi_M(x_f) \end{bmatrix} = 0 \qquad (2.4.47)$$

This optimal control is determined from the condition

$$\frac{\partial H}{\partial u} = \begin{bmatrix} \frac{\partial H}{\partial u_1} \\ \frac{\partial H}{\partial u_2} \\ \vdots \\ \frac{\partial H}{\partial u_r} \end{bmatrix} = 0 \qquad (2.4.48)$$

Note that Eqs. (2.4.46) and (2.4.47) constitute a set of $n+M$ terminal conditions. If the matrix $(\partial\psi_i / \partial x_j)$ has maximum rank (i.e., if the constraints $\psi_i = 0$ are not redundant) then the vector μ can be eliminated from this set providing a system of n equations in the variables p_f and x_f. For example, Eq. (2.4.46) can be rewritten as

$$\begin{bmatrix} \frac{\partial \phi}{\partial x_1} \\ \frac{\partial \phi}{\partial x_2} \\ \vdots \\ \frac{\partial \phi}{\partial x_n} \end{bmatrix} + \begin{bmatrix} \frac{\partial \psi_1}{\partial x_1} & \frac{\partial \psi_2}{\partial x_1} & \cdots & \frac{\partial \psi_M}{\partial x_1} \\ \frac{\partial \psi_1}{\partial x_2} & \frac{\partial \psi_2}{\partial x_2} & \cdots & \frac{\partial \psi_M}{\partial x_2} \\ \vdots & & & \vdots \\ \frac{\partial \psi_1}{\partial x_n} & \frac{\partial \psi_2}{\partial x_n} & \cdots & \frac{\partial \psi_M}{\partial x_n} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix} + \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_n \end{bmatrix} = 0$$

which is equivalent to the two sets

$$\begin{bmatrix} \frac{\partial \phi}{\partial x_1} \\ \vdots \\ \frac{\partial \phi}{\partial x_M} \end{bmatrix} + \begin{bmatrix} \frac{\partial \psi_1}{\partial x_1} & \frac{\partial \psi_2}{\partial x_1} & \cdots & \frac{\partial \psi_M}{\partial x_1} \\ \vdots & & & \\ \frac{\partial \psi_1}{\partial x_M} & \frac{\partial \psi_2}{\partial x_M} & \cdots & \frac{\partial \psi_M}{\partial x_M} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix} + \begin{bmatrix} p_1 \\ p_2 \\ \vdots \\ p_M \end{bmatrix} = 0 \qquad (2.4.49A)$$

$$\begin{bmatrix} \frac{\partial \phi}{\partial x_{M+1}} \\ \vdots \\ \frac{\partial \phi}{\partial x_n} \end{bmatrix} + \begin{bmatrix} \frac{\partial \psi_1}{\partial x_{M+1}} & \frac{\partial \psi_2}{\partial x_{M+1}} & \cdots & \frac{\partial \psi_M}{\partial x_{M+1}} \\ \vdots & & & \\ \frac{\partial \psi_1}{\partial x_n} & \frac{\partial \psi_2}{\partial x_n} & \cdots & \frac{\partial \psi_M}{\partial x_n} \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_M \end{bmatrix} + \begin{bmatrix} p_{M+1} \\ p_{M+2} \\ \vdots \\ p_n \end{bmatrix} = 0 \qquad (2.4.49B)$$

Solving the first set for the vector μ and substituting into the second set leaves a system of $n - M$ equations in the variables p_f and x_f. Combining these equations with Eqs. (2.4.47) then provides a system of n equations in the $2n$ unknowns, p_f and x_f. In what follows, it will be assumed that such a procedure has been followed with the multiplier μ eliminated from the $n+M$ constraints of Eqs. (2.4.46) and (2.4.47). The resulting n equations in x_f and p_f will be denoted by the expression

$$\eta(x_f, p_f) = \begin{bmatrix} \eta_1(x_f, p_f) \\ \eta_2(x_f, p_f) \\ \vdots \\ \eta_n(x_f, p_f) \end{bmatrix} = 0 \quad at \quad t = t_f \tag{2.4.50}$$

Thus, the terminal constraints of Eqs. (2.4.46) and (2.4.47) have been replaced by the equivalent representation given in Eq. (2.4.50).

In the neighboring extremal technique, each iterate satisfies the optimizing condition of Eq. (2.4.48) along with Eqs. (2.4.44) and (2.4.45), but does not satisfy the required terminal conditions in Eq. (2.4.50). The iteration consists of correcting and recorrecting the initial value of the vector, p_o, so that Eq. (2.4.50) are satisfied. The method by which this is accomplished is as follows: Since the initial value p_o completely determines a solution (a particular iterate) in the neighboring extremal case, both x_f and p_f are functions of p_o. Hence, the terminal constraint of (2.4.50) takes the form

$$\eta(x_f, p_f) = \eta(p_o) = \begin{bmatrix} \eta_1(p_o) \\ \eta_2(p_o) \\ \vdots \\ \eta_n(p_o) \end{bmatrix} = 0 \tag{2.4.51}$$

Let p_{o_1} denote the first iterate (i.e., the first guess of the initial vector) and expand η in the truncated Taylor series

$$\eta(p_o) \cong \eta(p_{o_1}) + \left(\frac{\partial \eta}{\partial p_o}\right)(p_o - p_{o_1})$$

The second iterate is taken as that value of p_o for which the above approximation is zero. Hence,

$$p_{o_2} = p_{o_1} - \left(\frac{\partial \eta}{\partial p_o}\right)^{-1} \eta(p_{o_1}) \tag{2.4.52}$$

In the initial application of this method (see for example Ref. (14)) the derivative matrix $(\partial\eta_i/\partial p_{j_o})$ was approximated by the finite difference matrix $(\Delta\eta_i/\Delta p_{j_o})$ with the latter generated by numerically integrating the original system but with a slightly perturbed set of p_o vectors. Later on, a method was suggested (see Refs. (15), (16), and (17)) for a more precise evaluation of this derivative using the theory of the second variation in the Calculus of Variations. This method is now generally referred to as the neighboring extremal technique. An account of this method is given next.

Minimizing $\phi(x_f)$ subject to the state equations

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

and the terminal conditions

$$\psi(x_f) = 0$$

is equivalent to minimizing the functional $\bar{J}$ where

$$\bar{J} = \phi + \psi^T\mu + \int_{t_o}^{t_f} p^T(\dot{x} - f)\,dt \equiv J \qquad (2.4.53)$$

Let p_{o_1} denote the first iterate and $\bar{J}_1$ the corresponding value of $\bar{J}$. Expanding $\bar{J}$ in a Taylor series about p_{o_1} and truncating the series after the second-order terms provides

$$\bar{J} \cong \bar{J}_1 + \delta\bar{J} + \frac{1}{2}\delta^2\bar{J}$$

where

$$\delta\bar{J} = \left(\frac{\partial\phi}{\partial x} + \frac{\partial\psi^T}{\partial x}\mu\right)^T\delta x_f + \psi^T\delta\mu + \int_{t_o}^{t_f} p^T\left(\delta\dot{x} - \frac{\partial f}{\partial x}\delta x - \frac{\partial f}{\partial u}\delta u\right)dt + \int_{t_o}^{t_f}\delta p^T(\dot{x} - f)\,dt \qquad (2.4.54)$$

$$\frac{\delta^2\bar{J}}{2} = \left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\psi^T}{\partial x^2}\mu\right)\frac{\delta x_f^2}{2} + \delta x_f\,\frac{\partial\psi^T}{\partial x}\delta\mu + \int_{t_o}^{t_f}\delta p^T\left(\delta\dot{x} - \frac{\partial f}{\partial x}\delta x - \frac{\partial f}{\partial u}\delta u\right)dt + \int_{t_o}^{t_f} -p^T\left(\frac{\partial^2 f}{\partial x^2}\frac{\delta x^2}{2} + \frac{\partial^2 f}{\partial x\,\partial u}\delta x\,\delta u + \frac{\partial^2 f}{\partial u^2}\frac{\delta u^2}{2}\right)dt \qquad (2.4.55)$$

Now, the nominal trajectory (the first iterate) will satisfy all the optimizing equations but the terminal condition of Eq. (2.4.50); that is, Eqs. (2.4.44), (2.4.45) and (2.4.48) are satisfied by the nominal trajectory. Hence, Eqs. (2.4.54) and (2.4.55) reduce to

$$\begin{aligned}\bar{J} - \bar{J}_1 \cong \delta\bar{J} + \frac{\delta^2\bar{J}}{2} &= \left(\frac{\partial\phi}{\partial x} + \frac{\partial\psi^T}{\partial x}\mu\right)\delta x_f + \psi^T\delta\mu + \left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\psi^T}{\partial x^2}\mu\right)\frac{\delta x_f^2}{2} \\ &+ \delta x_f \frac{\partial\psi^T}{\partial x}\delta\mu + p\,\delta x_f \\ &- \int_{t_o}^{t_f} p^T\left(\frac{\partial^2 f}{\partial x^2}\frac{\delta x^2}{2} + \frac{\partial^2 f}{\partial x\partial u}\delta x\delta u + \frac{\partial^2 f}{\partial u^2}\frac{\delta u^2}{2}\right)dt \\ &+ \int_{t_o}^{t_f}\delta p^T\left(\delta\dot{x} - \frac{\partial f}{\partial x}\delta x - \frac{\partial f}{\partial u}\delta u\right)dt \qquad (2.4.56)\end{aligned}$$

Now the second iterate is to be selected to minimize the second-order approximation to $\bar{J}$ as given in Eq. (2.4.56). Proceeding formally, the first variation of the expression in (2.4.56) with respect to the variables $\delta x(t)$, $\delta u(t)$, $\delta p(t)$, δx_f and $\delta\mu$ is set equal to zero. This leads to the equations

$$\delta\dot{x} = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial u}\delta u \iff \delta\dot{x}_i = \sum_{j=1}^{n}\frac{\partial f_i}{\partial x_j}\delta x_j + \sum_{k=1}^{r}\frac{\partial f_i}{\partial u_k}\delta u_k \qquad (2.4.57)$$

$$\delta\dot{p} = -\frac{\partial^2 H}{\partial p\partial x}\delta p - \frac{\partial^2 H}{\partial x^2}\delta x - \frac{\partial^2 H}{\partial x\partial u}\delta u$$

$$\delta\dot{p}_i = -\sum_{j=1}^{n}\frac{\partial^2 H}{\partial x_i\partial x_j}\delta p_j - \sum_{j=1}^{n}\frac{\partial^2 H}{\partial x_i\partial x_j}\delta x_j - \sum_{k=1}^{r}\frac{\partial^2 H}{\partial x_i\partial u_k}\delta u_k \qquad (2.4.58)$$

$$\frac{\partial^2 H}{\partial u\,\partial p}\delta p + \frac{\partial^2 H}{\partial u^2}\delta u + \frac{\partial^2 H}{\partial u\partial x}\delta x = 0 \iff \sum_{j=1}^{n}\frac{\partial^2 H}{\partial u_k\,\partial p_j}\delta p_j + \sum_{i=1}^{r}\frac{\partial^2 H}{\partial u_k\,\partial u_i}\delta u_i + \sum_{j=1}^{n}\frac{\partial^2 H}{\partial u_k\partial x_j}\delta x_j = 0 \qquad ; k = 1, r \qquad (2.4.59)$$

$$\left.\begin{aligned} \psi + \frac{\partial \phi}{\partial x} + \frac{\partial \psi^T}{\partial x}\mu + \left\{\delta\psi + \frac{\partial^2 \phi}{\partial x^2}\,\delta x + \frac{\partial^2 \psi^T}{\partial x^2}\mu\,\delta x + \frac{\partial \psi^T}{\partial x}\,\delta\mu\right\} = 0 \\ \psi + \frac{\partial \psi}{\partial x}\,\delta x = 0 \end{aligned}\right\}_{t=t_f} \qquad (2.4.60)$$

Eqs. (2.4.57) to (2.4.59) are the linearized approximations to the state, adjoint and optimizing equations given in (2.4.44) and (2.4.48); the boundary condition of (2.4.60) is the linearized version of Eqs. (2.4.46) and (2.4.47). In view of the assumption that the multiplier μ has been eliminated and Eqs. (2.4.46) and (2.4.47) represented in the form of Eq. (2.4.50), it can be shown that the linearized result in (2.4.60) is equivalent to

$$\eta(x_{f_1}, p_{f_1}) + \frac{\partial \eta}{\partial x_f}\,\delta x_f + \frac{\partial \eta}{\partial p_f}\,\delta p_f = 0 \qquad (2.4.61)$$

Thus, the second iterate is generated from

$$p_{0_2} = p_{0_1} + \delta p_0$$

where δp_0 is to be chosen so that Eqs. (2.4.57), (2.4.58) and (2.4.59) are satisfied, subject to the boundary conditions of (2.4.61).

Under the assumption that $\left|\partial^2 H/\partial u^2\right| \neq 0$ along the first iterate, Eq. (2.4.59) can be solved for δu yielding

$$\delta u = -\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}\left\{\frac{\partial^2 H}{\partial u \partial p}\,\delta p + \frac{\partial^2 H}{\partial u \partial x}\,\delta x\right\} \qquad (2.4.62)$$

Substitution of this result into (2.4.57) and (2.4.58) provides the linear system in δx and δp

$$\begin{pmatrix} \delta \dot{x} \\ \delta \dot{p} \end{pmatrix} = \begin{pmatrix} C^{①} & C^{②} \\ C^{③} & -C^{①T} \end{pmatrix}\begin{pmatrix} \delta x \\ \delta p \end{pmatrix} \qquad (2.4.63)$$

where the $C^{(k)}$ $(k=1,3)$ are $n \times n$ matrices with the elements given by

$$C^{①} = (C^{①}_{i,j}) = \left\{ \frac{\partial f_i}{\partial x_j} - \sum_{k,l=1}^{r} \left(\frac{\partial^2 f_i}{\partial u_k} \frac{\partial^2 H}{\partial u_l \partial x_j} \left[\frac{\partial^2 H}{\partial u^2} \right]^{-1}_{l,k} \right) \right\}$$

$$C^{②} = (C^{②}_{ij}) = \left\{ -\sum_{k,l=1}^{r} \left(\frac{\partial f_i}{\partial u_k} \frac{\partial^2 H}{\partial u_l \partial p_j} \left[\frac{\partial^2 H}{\partial u^2} \right]^{-1}_{l,k} \right) \right\}$$

$$C^{③} = (C^{③}_{ij}) = \left\{ -\frac{\partial^2 H}{\partial x_i \partial x_j} + \sum_{k,l=1}^{r} \left(\frac{\partial^2 H}{\partial x_i \partial u_k} \frac{\partial^2 H}{\partial u_l \partial x_j} \left[\frac{\partial^2 H}{\partial u^2} \right]^{-1}_{l,k} \right) \right\} \quad (2.4.64)$$

The solution to this system can be represented by

$$\begin{pmatrix} \delta x(t) \\ \delta p(t) \end{pmatrix} = \Lambda(t) \begin{pmatrix} \delta x_o \\ \delta p_o \end{pmatrix} \quad (2.4.65)$$

where Λ is the 2n by 2n fundamental matrix solutions

$$\dot{\Lambda} = \begin{pmatrix} C^{①} & C^{②} \\ C^{③} & -C^{①T} \end{pmatrix} \Lambda \quad ; \qquad \Lambda(0) = I \quad (2.4.66)$$

Let Λ be partitioned into four n by n matrices of the form

$$\Lambda = \begin{pmatrix} \Lambda_1 & \Lambda_2 \\ \Lambda_3 & \Lambda_4 \end{pmatrix}$$

Then, since δx_o is zero, it follows that

$$\delta x_f = \Lambda_2(t_f)\, \delta p_o$$

$$\delta p_f = \Lambda_4(t_f)\, \delta p_o \quad (2.4.67)$$

$$\delta p_o = -\left(\frac{\partial \eta}{\partial x_f}\Lambda_2(t_f) + \frac{\partial \eta}{\partial p_f}\Lambda_4(t_f)\right)^{-1} \eta(x_{f_1}, p_{f_1}) \qquad (2.4.68)$$

The second iterate is now generated from the initial p vector as

$$p_{20} = p_{10} + \delta p_o$$

where δp_o is evaluated using (2.4.68). If the first solution comes reasonably close to the correct terminal set

$$\eta(x_f, p_f) = 0$$

then the second solution should be even closer. In fact, if the method converges to the correct terminal set, the convergence will be quadratic with

$$|p_{o_{n+1}} - p_{o_n}| \sim |p_{o_n} - p_{o_{n-1}}|^2$$

It is again emphasized that the neighboring extremal technique is simply a device for solving the transcendental equation in (2.4.51) for the unknown vector p_o

A step by step account of the calculations to be performed at each iteration is given below:

(1) Guess a p_o vector

(2) Integrate the $2n$ system

$$\dot{x} = f(x, u, t) \quad ; \quad \dot{p} = -\frac{\partial H}{\partial x}$$

from t_o to t_f with u satisfying

$$\frac{\partial H}{\partial u} = 0$$

(3) Test to see if the n terminal conditions

$$\eta(x_f, p_f) = 0$$

are satisfied. If they are not, continue.

(4) Determine the $2n \times 2n$ fundamental matrix solution

$$\dot{\Lambda} = C\Lambda \quad ; \quad \Lambda(t_o) = I$$

where the C matrix is given in Eq. (2.4.64). Note that since only Λ_2 and Λ_4 are used (see Eq. (2.4.67)), only half of the

calculation required in evaluation Λ need be performed.

(5) Evaluate δp_o using Eq. (2.4.68) and set

$$p_{o_2} = p_{o_1} + \delta p_o$$

(6) Return to Step (2)

The process is continued until the required terminal conditions

$$\eta(x_f, p_f) = 0$$

are satisfied.

2.4.4 Quasilinearization

Stating the problem again, the solution to the system

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

$$\dot{p} = -\frac{\partial H}{\partial x} \tag{2.4.68}$$

is to be determined which satisfies the initial conditions

$$x = x_o \quad at \quad t = t_o \tag{2.4.69}$$

and the terminal conditions

$$\left.\begin{aligned} \psi(x_f) &= 0 \\ \frac{\partial \phi}{\partial x} + \frac{\partial \psi^T}{\partial x}\mu + p &= 0 \end{aligned}\right\} \quad at \quad t = t_f \tag{2.4.70}$$

and where u is selected so that

$$\frac{\partial H}{\partial u} = 0 \tag{2.4.71}$$

It is also assumed that a procedure similar to that indicated in Eqs. (2.4.49) to (2.4.50) has been used to eliminate the multiplier μ from the $n+M$ terminal conditions of Eq. (2.4.70). This approach results in a set of n equations in the $2n$ variables x_f and p_f which will be represented by

$$\eta(x_f, p_f) = \begin{bmatrix} \eta_1(x_f, p_f) \\ \eta_2(x_f, p_f) \\ \vdots \\ \eta_n(x_f, p_f) \end{bmatrix} = 0 \qquad (2.4.72)$$

Thus, the multiplier μ no longer appears in the boundary conditions that are to be satisfied.

The method of quasilinearization is frequently referred to in the literature as the generalized Newton-Raphson technique. As indicated in Refs. (18) to (20) the technique is conceptually the same as the Newton-Raphson procedure employed in maxima-minima theory. For this reason, the latter name may be more appropriate. The use of the technique in trajectory optimization problems is demonstrated in Refs. (21) and (22).

Like neighboring extremal, quasilinearization is a second-order iterative process in that second-order terms in the series expansion are used. In contrast, each successive iterate satisfies the boundary conditions of (2.4.69) and (2.4.72), but does not satisfy the governing differential equations of (2.4.68). The iterative process consists of correcting and recorrecting the starting solution in a direction which tends to "bring in" the differential constraints subject to the constraint that the optimizing condition of (2.4.71) is satisfied. This condition can either be satisfied exactly by each iterate (the usual procedure in the literature) or can be "brought in" by the iteration process.

Again, let the functional to be minimized by written as

$$\bar{J} = \phi + \psi^T \mu + \int_{t_0}^{t_f} p^T (\dot{x} - f)\, dt \qquad (2.4.73)$$

and let the first iteration be denoted by

$$\left.\begin{aligned} x &= x_1(t) \\ p &= p_1(t) \\ u &= u_1(t) \end{aligned}\right\} \quad \text{first iteration} \qquad (2.4.74)$$

where x_1 and p_1 are required to satisfy the boundary constraints of Eqs. (2.4.69) and (2.4.72) but not the differential constraints of (2.4.68). Now, expand $\bar{J}$ in a truncated Taylor series about the first iteration as

$$\begin{aligned}\bar{J} - \bar{J}_1 \cong\ & \delta\bar{J} + \frac{\delta^2\bar{J}}{2} \\ =\ & \left(\frac{\partial\phi}{\partial x}\right)^T \delta x_f + \left(\frac{\partial\psi^T}{\partial x}\mu\right)^T \delta x_f + \left(\frac{\partial^2\phi}{\partial x^2} + \frac{\partial^2\psi^T}{\partial x^2}\mu\right)\frac{\delta x_f^2}{2} + \delta x_f \frac{\partial\psi^T}{\partial x}\delta\mu \\ & + \int_{t_o}^{t_f} \delta p^T (\dot{x}_1 - f_1)\, dt + \int_{t_o}^{t_f} p^T\left(\delta\dot{x} - \frac{\partial f}{\partial x}\delta x - \frac{\partial f}{\partial u}\delta u\right) dt \\ & + \int_{t_o}^{t_f} \delta p^T\left\{\delta\dot{x} - \frac{\partial f}{\partial x}\delta x - \frac{\partial f}{\partial u}\delta u\right\} dt - \int_{t_o}^{t_f} p^T\left\{\frac{\partial^2 f}{\partial x^2}\frac{\delta x^2}{2} + \frac{\partial^2 f}{\partial x \partial u}\delta x\delta u + \frac{\partial^2 f}{\partial u^2}\frac{\delta u^2}{2}\right\}\end{aligned} \tag{2.4.76}$$

and select the second iteration so as to minimize this second-order approximation. Proceeding formally, the first variation with respect to the quantities $\delta x(t)$, $\delta u(t)$, $\delta p(t)$, δx_f, and $\delta\mu$ is set equal to zero providing the differential expressions

$$\delta\dot{x} = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial u}\delta u - \left\{\dot{x}_1 - f(x_1, u_1, t)\right\} \tag{2.4.77}$$

$$\delta\dot{p} = -\frac{\partial^2 H}{\partial x \partial p}\delta p - \frac{\partial^2 H}{\partial x \partial u}\delta u - \frac{\partial^2 H}{\partial x^2}\delta x - \left\{\dot{p}_1 + \left(\frac{\partial H}{\partial x}\right)\right\} \tag{2.4.78}$$

$$\frac{\partial H}{\partial u} + \frac{\partial^2 H}{\partial u \partial x}\delta x + \frac{\partial^2 H}{\partial u \partial p}\delta p + \frac{\partial^2 H}{\partial u^2}\delta u = 0 \tag{2.4.79}$$

and the boundary conditions

$$\begin{aligned}& \frac{\partial\phi}{\partial x} + \frac{\partial\psi^T}{\partial x}\mu + p_f + \frac{\partial^2\phi}{\partial x^2}\delta x_f + \frac{\partial^2\psi^T}{\partial x^2}\mu\,\delta x_f + \frac{\partial\psi^T}{\partial x}\delta\mu + \delta p_f = 0 \\ & \psi + \frac{\partial\psi}{\partial x}\delta x_f = 0\end{aligned} \tag{2.4.80}$$

Since the first iterate must satisfy the terminal conditions of Eq. (2.4.72) (which is equivalent to (2.4.70)), the boundary equations can be shown to reduce to the n conditions

$$\frac{\partial \eta}{\partial x_f}\,\delta x_f + \frac{\partial \eta}{\partial p_f}\,\delta p_f = 0 \tag{2.4.81}$$

In addition, the initial condition

$$\delta x_o = 0 \tag{2.4.82}$$

must hold since the first iteration satisfies Eq. (2.4.69).

The required boundary conditions of Eqs. (2.4.81) and (2.4.82) are the linearized from of the actual boundary conditions in (2.4.69) and (2.4.70). Also, Eqs. (2.4.77) to (2.4.79) are simply the linearized from of the constraint conditions in (2.4.68) and (2.4.71). Note the appearance of the terms $(\dot{x}_1 - f_1)$ and $(\dot{p}_1 + (\frac{\partial H}{\partial x})_1)$ in the Eqs. (2.4.77) and (2.4.78). These terms are present since the first solution does not satisfy the differential constraints. For example, define z by

$$z = \dot{x} - f(x, u, t) = z(\dot{x}, x, u, t) = 0$$

and expand z about some first solution $z_1 = z(\dot{x}_1, x_1, u_1, t)$, which does not satisfy the zero condition. This procedure yields

$$z = 0 \cong z_1 + \frac{\partial z}{\partial \dot{x}}\,\delta\dot{x} + \frac{\partial z}{\partial x}\,\delta x + \frac{\partial z}{\partial u}\,\delta u$$

which is exactly the same as Eq. (2.4.77).

Once the quantities δ_x, δ_p and δ_u have been computed using Eqs. (2.4.77) to (2.4.79) and the boundary conditions of (2.4.81) and (2.4.82), the second iterate can be determined from

$$\begin{aligned} x_2 &= x_1 + \delta x \\ p_2 &= p_1 + \delta p \\ u_2 &= u_1 + \delta u \end{aligned} \tag{2.4.83}$$

However, it is computationally more convenient to substitute (2.4.83) into the governing equations, Eqs. (2.4.77) to (2.4.79) and (2.4.81) to (2.4.82), and then solve for the second iterate directly. This approach avoids any trouble which the presence of the derivatives, $\dot{x}_1$ and $\dot{p}_1$ in (2.4.77) and (2.4.78), may cause. Thus, substituting Eq. (2.4.83) into Eqs. (2.4.77) to (2.4.79) provides

$$\dot{x}_2 = \left(\frac{\partial f}{\partial x}\right)_1 x_2 + \left(\frac{\partial f}{\partial u}\right)_1 u_2 + f(x_1, u_1, t) - \left(\frac{\partial f}{\partial x}\right)_1 x_1 - \left(\frac{\partial f}{\partial u}\right)_1 u_1$$

(2.4.84)

$$\dot{p}_2 = -\left(\frac{\partial^2 H}{\partial x \partial p}\right)_1 p_2 - \left(\frac{\partial^2 H}{\partial x \partial u}\right) u_2 - \left(\frac{\partial^2 H}{\partial x^2}\right)_1 x_2 - \left(\frac{\partial H}{\partial x}\right)_1 + \left(\frac{\partial^2 H}{\partial x \partial p}\right)_1 p_1 + \left(\frac{\partial^2 H}{\partial x \partial u}\right)_1 u_1 + \left(\frac{\partial^2 H}{\partial x^2}\right)_1 x_1$$

(2.4.85)

$$\left(\frac{\partial H}{\partial u}\right) + \left(\frac{\partial^2 H}{\partial u \partial x}\right)_1 (x_2 - x_1) + \left(\frac{\partial^2 H}{\partial x \partial p}\right)_1 (p_2 - p_1) + \left(\frac{\partial^2 H}{\partial u^2}\right) (u_2 - u_1) = 0 \quad (2.4.86)$$

Similarly, the boundary conditions of (2.4.82) and (2.4.81) become

$$x_2 = x_0 \quad \text{at} \quad t = t_0$$

$$\left(\frac{\partial \eta}{\partial x_f}\right)_1 (x_2 - x_1) + \left(\frac{\partial \eta}{\partial p_f}\right)_1 (p_2 - p_1) = 0 \quad \text{at} \quad t = t_f$$

(2.4.87)

Thus, under the assumption that $\left|\frac{\partial^2 H}{\partial u^2}\right|$ does not vanish, Eq. (2.4.86) can be solved for u and the result substituted into Eqs. (2.4.84) and (2.4.85). This substitution provides the linear system

$$\begin{pmatrix} \dot{x}_2 \\ \dot{p}_2 \end{pmatrix} = \begin{pmatrix} C^{(1)} & C^{(2)} \\ C^{(3)} & -C^{(1)T} \end{pmatrix} \begin{pmatrix} x_2 \\ p_2 \end{pmatrix} + \begin{pmatrix} d^{(1)} \\ d^{(2)} \end{pmatrix}.$$

(2.4.88)

where the $C^{(1)}$ are given by Eq. (2.4.64) and where $d^{(1)}$ and $d^{(2)}$ are n vectors given by

$$d^{(1)} = f(x_1, u_1, t) - \frac{\partial f}{\partial x} x_1 + \frac{\partial f}{\partial u}\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}\left\{\frac{\partial^2 H}{\partial u \partial x} x_1 + \frac{\partial^2 H}{\partial u \partial p} p_1 - \frac{\partial H}{\partial u}\right\}$$

$$d^{(2)} = -\left(\frac{\partial H}{\partial x}\right) + \frac{\partial^2 H}{\partial x^2} x_1 - \frac{\partial^2 H}{\partial x \partial u}\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}\left\{\frac{\partial^2 H}{\partial u \partial x} x_1 + \frac{\partial^2 H}{\partial u \partial p} p_1 - \frac{\partial H}{\partial u}\right\}$$

(2.4.89)

This system has the solution

$$\begin{bmatrix} x_2(t) \\ p_2(t) \end{bmatrix} = \Lambda(t)\left\{\begin{bmatrix} x_2(t_0) \\ p_2(t_0) \end{bmatrix} + \int_{t_0}^{t} \Lambda^{-1}(\tau) \begin{bmatrix} d^{(1)} \\ d^{(2)} \end{bmatrix} d\tau\right\} \qquad (2.4.90)$$

where the quantity $x_2(t_0) = x_0$ and $p_2(t_0)$ is to be chosen so that Eq. (2.4.87) is satisfied. The matrix Λ is the fundamental matrix solution given in (2.4.66).

The step by step iterative process takes the following form:

(1) Select the time histories of the three functions $x_1(t)$, $p_1(t)$ and $u_1(t)$ so that the boundary conditions

$$x(t_0) = x_0 \quad at \quad t = t_0$$

$$\eta(x_f, t_f) = 0 \quad at \quad t = t_f$$

(2) Compute the matrices $C^{(1)}$, $C^{(2)}$, $C^{(3)}$, $d^{(1)}$ and $d^{(2)}$ using Eqs. (2.4.64) and (2.4.89). From these matrices determine the fundamental matrix Λ where

$$\dot{\Lambda} = C\Lambda; \quad \Lambda(t_0) = I$$

(3) Select $p_2(t_0)$ so that the terminal state given by

$$\begin{bmatrix} x_2(t_f) \\ p_2(t_f) \end{bmatrix} = \Lambda(t_f)\left\{\begin{bmatrix} x_0 \\ p_2(t_0) \end{bmatrix} + \int_{t_0}^{t_f} \Lambda^{-1}(\tau) \begin{bmatrix} d^{(1)} \\ d^{(2)} \end{bmatrix} d\tau\right\}$$

satisfies the linearized terminal conditions

$$\frac{\partial \eta}{\partial x_f}(x_2 - x_1) - \frac{\partial \eta}{\partial p_f}(p_2 - p_1) = 0 \quad at \quad t = t_f$$

(4) Use Eq. (2.4.90) to compute the second iterate x_2 and p_2.

(5) Use Eq. (2.4.86) to evaluate $u_2(t)$.

(6) Go to Step (2).

The process continues until additional iterations produce essentially no change in $x(t)$, $p(t)$ or $u(t)$. At this point, the solution has been achieved.

2.4.5 Discussion

The three techniques which have been presented in the preceeding sections are those most commonly used to solve optimization problems of the Classical Mayer type (i.e., a Mayer Problem in which no constraints are imposed on the control action u). It is to be emphasized that these are not the only techniques that are available. However, all other techniques are simply minor variations on the above three. While these variations may be of considerable importance when it comes to solving specialized problems, they are of limited interest as far as general methods of solution are conserned.

Each of the techniques has certain advantages and disadvantages when compared with the other two. An assessment of the relative merits and utility of each method is offered in the following paragraphs.

A. Amount of Computation Required for Each Iteration

As a general rule, the steepest descent method requires less computation for any one iteration than either neighboring extremal or quasilinearization. This advantage is the result of the fact that steepest descent is a first-order method and only the first-order terms in a series expansion need be evaluated to compute each succeeding iterate. The evaluation of second-order terms, which are needed is neighboring extremal and quasilinearization, causes a slight to moderate increase in the amount of computation required.

The fact that steepest descent requires less computation at any one step does not mean that the overall amount of computation is minimized by using this method. Possibly more steps (more iterates) may be required to generate a solution by steepest descent; or possibly, the method itself may diverge. This, of course, would depend on the particular problem being solved and the exact value of the starting iteration.

B. Storage on the Computer

In general, the neighboring extremal technique requires the least amount of storage and quasilinearization, the most.

The computer storage is directly proportional to the amount of information required to completely specify a particular iterate. In neighboring extremal, this information consists of simply n number, the initial value of the p vector. In the steepest descent method, the required information is the time history of the control vector u, while in quasilinearization, the time history of the x, u and p vectors are needed.

C. Difficulty to Program for the Computer

As in the case of storage, the programming difficulty is proportional to

the amount of information that must be retained in order to specify a particular iterate. For this reason, neighboring extremal is considerably simpler to program than steepest descent (considerably less information is needed to specify the iterate) and quasilinearization is significantly more difficult.

D. Selection of a Starting Iterate

To begin the numerical process, a starting iterate must be selected. The difficulty in making this selection is a function of the number of conditions that the starting iterate must satisfy. In neighboring extremal, the starting iterate is specified by the value of the initial P vector and this vector is not required to satisfy any condition whatsoever. Hence, the difficulty here is zero. In quasilinearization, the starting iterate, $x_1(t)$, $P_1(t)$ and $u_1(t)$, must satisfy certain initial and terminal conditions (Eqs. (2.4.69) and (2.4.72)). However, since these functions need not satisfy the differential constraints, it is an easy matter to make this selection. (The difficulty here is very close to zero.) On the other hand, the starting iterate in steepest descent must satisfy both the state boundary conditions and the state differential constraints (Eqs. (2.4.19A) to (2.4.19C)). For this reason, the selection of a starting iterate in steepest descent is an order of magnitude more difficult than that encountered in either of the other two methods. In fact, the program must generally have a special subroutine (which is itself iterative) for calculating a satisfactory starting iterate.

E. Rate and Radius of Convergence

As a general rule, second-order iterative techniques converge rapidly, if the starting point is close to the optimal point, and diverge rapidly elsewhere. On the other hand, first-order methods can be made to converge from points far removed from the optimal point but the rate of convergence tends to zero as the optimal point is approached. For this reason it appears that some combination of both the methods might produce the best results. Of course, the radius of convergence of both neighboring extremal and quasilinearization can be extended by using a procedure similar to that employed in Eq. (2.2.34) for extending the radius of convergence of the Newton-Raphson method. By this device the second-order techniques attempt to remove only a part of the error (rather than the total error) during any one iteration. In papers dealing with optimization problems which have been solved using a second-order method, such a procedure is almost always used.

2.5 NONCLASSICAL MAYER PROBLEM

The previous section dealt strictly with the Classical Mayer Problem in which the admissible control set U is the entire r dimensional control space and the individual control components satisfy the Euler equation

$$\frac{\partial H}{\partial u} = \begin{bmatrix} \frac{\partial H}{\partial u_1} \\ \frac{\partial H}{\partial u_2} \\ \vdots \\ \frac{\partial H}{\partial u_r} \end{bmatrix} = 0 \qquad (2.5.1)$$

In many applications, these conditions are not satisfied. For example, in the vehicle problem of Section 2.1, the control component u_2 (corresponding to the throttle setting(had to satisfy the inequality condition of Eq. (2.1.13) with the control set U limited by Eq. (2.1.14). In this case, the optimal value of u_2 (the value which maximizes the Hamiltonian H) does not satisfy Eq. (2.5.1) as is indicated by Eqs. (2.1.25) and (2.1.26).

The extension of the three iterative procedures to nonclassical problems is a relatively straight forward matter. Unfortunately the exact form which this extension takes depends strongly on the exact form of the admissible control set U and the number of control components which are constrained. Hence, it is not possible to develop a set of equations which completely describe the iterative process and which are also generally applicable. For this reason, the attention will be focused on one particular type of nonclassical problem; namely, the "bang-bang" problem. For convenience, it will also be assumed that only one control component is constrained. The approach does not change if additional constraints are included. However, the algebra gets unwieldly.

The state equations for the problem under consideration are

$$\dot{x} = f(x, u, t) \Longrightarrow \dot{x}_i = f_i(x, u, t) \; ; \quad i = 1, n \qquad (2.5.1)$$

where u is an r dimensional control vector which is required to lie in the set U . It is assumed that the control components have been numbered so that U is described by

$$u \in U \Longleftrightarrow \begin{cases} -\infty \leq u_1 \leq +\infty \\ -\infty \leq u_2 \leq +\infty \\ \vdots \\ -\infty \leq u_{r-1} \leq +\infty \\ u_{r_{MIN}} \leq u_r \leq u_{r_{MAX}} \end{cases} \qquad (2.5.2)$$

that is, only the r^{th} control component, u_r, is constrained; this component is to lie in the region $[U_{r_{MAX}}, U_{r_{MAX}}]$. The other components can take on any values whatsoever. In what follows, it will be convenient to let v denote the constrained component; that is

$$v = u_r$$

$$u_{r_{MIN}} = v_{MIN} \leq v \leq v_{MAX} = u_{r_{MAX}} \tag{2.5.3}$$

with the control vector given by

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{r-1} \\ v \end{bmatrix} \tag{2.5.4}$$

In addition to the form of U, it is required that v appear linearly in the governing state equations with (2.5.1) taking the explicit form

$$\begin{aligned} \dot{x} &= f(x, u, v, t) = F(x, u, t) + v\,G(x, u, t) \\ \dot{x}_i &= F_i(x, u, t) + v\,G_i(x, u, t) \; ; \quad i = 1, n \end{aligned} \tag{2.5.5}$$

It is this latter requirement that causes the problem to be "bang-bang" - a term which indicates that v is either on its upper or lower bound and that transitions from one bound to the other are made discontinuously. Note that the vehicle problem of Section 2.1 (see Eqs. (2.1.9)) took just this from with v equal to u_2 and with v satisfying

$$0 \leq v \leq 1$$

as indicated by (2.1.13).

While this problem is not completely general, it does occur rather frequently in trajectory applications. More important, the manner in which the iterative techniques are modified to handle this particular problem should serve to indicate similar modification procedures for other nonclassical problems.

2.5.1 Problem Statement and Necessary Conditions

The state of the system is given by

$$\dot{x}_i = f_i(x, u, \nu, t) = F_i(x, u, t) + \nu G_i(x, u, t) \qquad (2.5.6)$$

$$i = 1, n$$

where u is an $r-1$ dimensional control vector

$$u = \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{r-1} \end{bmatrix}$$

and where ν is a one-dimensional control variable required to satisfy the inequality

$$\nu_{MIN} \leq \nu \leq \nu_{MAX} \qquad (2.5.7)$$

Again, the control action, u and ν, is to be determined so that the terminal constraints

$$\psi(x) = \begin{bmatrix} \psi_1(x) \\ \psi_2(x) \\ \vdots \\ \psi_M(x) \end{bmatrix} = 0 \quad at \quad t = t_f \qquad (2.5.8)$$

are satisfied and so that a function of the terminal state is minimized; that is

$$J = \phi(x_f) \quad = \text{minimum} \qquad (2.5.9)$$

Applying the maximum principle of Section 2.1 (which requires that the control maximize the Hamiltonian) leads to the condition that u be determined so that

$$\frac{\partial H}{\partial u} = \begin{bmatrix} \frac{\partial H}{\partial u_1} \\ \frac{\partial H}{\partial u_2} \\ \vdots \\ \frac{\partial H}{\partial u_{r-1}} \end{bmatrix} = 0 \qquad (2.5.10)$$

where

$$H = p^T f = \sum_{i=1}^{n} p_i f_i = \sum_{i=1}^{n} p_i \left\{ F_i + \nu G_i \right\}$$

and that ν be selected so that

$$\begin{aligned} \nu &= \nu_{MAX} , \quad \theta > 0 \\ \nu &= \nu_{MIN} , \quad \theta < 0 \end{aligned} \tag{2.5.11}$$

where

$$\theta = p^T G = \sum_{i=1}^{n} p_i G_i (x, u, t) \tag{2.5.12}$$

The adjoint variables for this problem satisfy the differential system

$$\dot{p}_i = - \frac{\partial H}{\partial x_i} ; \qquad i = 1, n \tag{2.5.13}$$

and the transversality conditions

$$p + \left(\frac{\partial \psi}{\partial x} \right)^T \mu + \frac{\partial \phi}{\partial x} = 0 \iff p_i + \sum_{j=1}^{m} \mu_j \frac{\partial \psi_j}{\partial x_i} + \frac{\partial \phi}{\partial x_i} = 0 \quad \text{at} \quad t = t_f \tag{2.5.14}$$

where μ is selected so that Eq. (2.5.8) holds. Note that the multiplier μ can be eliminated from the terminal conditions of Eqs. (2.5.8) and (2.5.14). This elimination leads to a system of n independent equations in the $2n$ variables p_f and x_f. As in the preceeding section, this system will be represented by

$$\eta (x_f, p_f) = 0 \iff \begin{bmatrix} \eta_1 (x_f, p_f) \\ \eta_2 (x_f, p_f) \\ \vdots \\ \eta_n (x_f, p_f) \end{bmatrix} = 0 \quad \text{at} \quad t = t_f \tag{2.5.15}$$

Once again, it is assumed that the final time is explicitly specified. An extension to the "final time open" case will be given in Section 2.5.5.

2.5.2 Neighboring Extremal Technique

Assume that an initial p vector, p_{0_1}, has been selected and that an optimal trajectory has been computed which does not satisfy the terminal constraints of Eq. (2.5.15). As in the classical case of Section 2.4.3, the neighboring extremal technique seeks to correct the initial value of p so that the terminal conditions are met. The process of correction is the Newton-Raphson technique contained in the iterative equation

$$\eta(p_0) = \eta(p_{0_1}) + \left(\frac{\partial \eta}{\partial p_0}\right) \delta p_0 = 0 \tag{2.5.16}$$

where δp_0 is the difference between the first and second values of the initial p vector; i.e.,

$$p_{0_2} = p_{0_1} + \delta p_0$$

As pointed out in the previous section, the rate of convergence of Eq. (2.5.16) is quadratic if the partial derivative matrix $\partial \eta / \partial p_0$ is calculated exactly. It has been this investigator's experience that finite difference approximations to this matrix are sometimes inadequate unless a great deal of care is taken in the approximation process such as performing the calculations in double precision on the computer. The advantage of the neighboring extremal technique is that it allows for a precise evaluation of the derivative matrix without resorting to the double precision calculations.

Denote by the subscript $_1$ the trajectory and control action resulting from a certain choice of p_0, say p_{0_1}; that is

$$\begin{aligned} x(t) &= x_1(t) \\ u(t) &= u_1(t) \\ w(t) &= w_1(t) \\ p(t) &= p_1(t) \end{aligned} \tag{2.5.17}$$

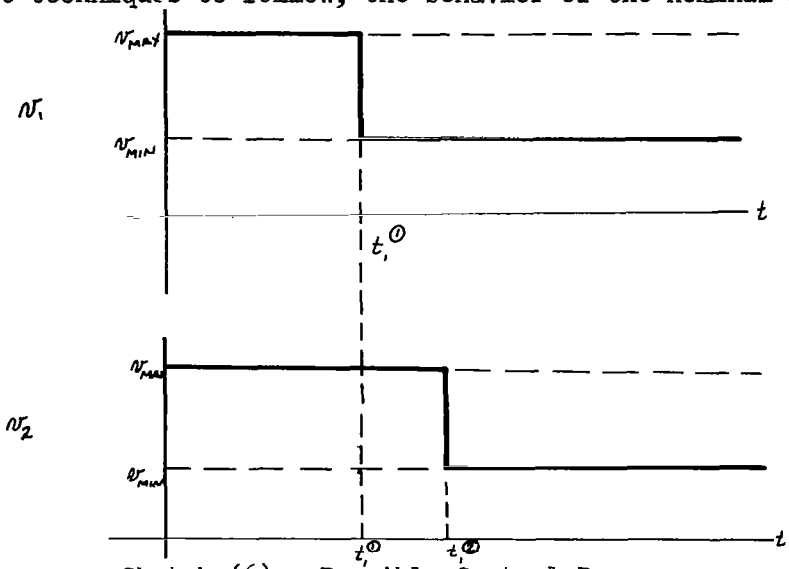
Now, write the perturbed state equation in the form

$$\delta \dot{x} = \left(\frac{\partial f}{\partial x}\right)^T \delta x + \left(\frac{\partial f}{\partial u}\right)^T \delta u + \left(\frac{\partial f}{\partial w}\right)^T \delta w \tag{2.5.18}$$

where

$$\delta x = x_2(t) - x_1(t)$$

and where the trajectory $x_2(t)$ is that resulting from a new choice of p_0, say p_{0_2}. It is tacitly assumed in Eq. (2.5.18) that δx, δu and δv are small (infinitesimal). This, however, is not the case. If the two control programs, $v_1(t)$ and $v_2(t)$ shown in Sketch (6) are compared, then it follows that δv is finite on the interval $(t_1^{①}, t_1^{②})$ and the validity of Eq. (2.5.18) on this interval becomes questionable. Now, in the quasilinear and gradient techniques to follow, the behavior of the nominal and neighboring



Sketch (6) - Possible Control Programs

solutions specifically on this interval, $(t_1^{①}, t_1^{②})$, will be of prime interest; thus, the description contained in Eq. (2.5.18) is not correct. In the neighboring extremal technique, however, the derivative matrix is calculated by considering the limiting case in which the interval goes to zero. In this case the description contained in Eq. (2.5.18) is adequate since, though δv is finite, its effect becomes infinitesimal as the interval shrinks to zero. However, to be consistent with the quasilinearization and gradient techniques which follow, a different perturbation method will be used from that employed in Eq. (2.5.18) even though Eq. (2.5.18) is adequate for the neighboring extremal case.

Suppose the perturbations δx, δp and δu are calculated by comparing the nominal set and neighboring solutions at different values of time. For example, let

$$\begin{aligned}
\delta x(t) &= x_2(t+\delta t) - x_1(t) \\
\delta u(t) &= u_2(t+\delta t) - u_1(t) \\
\delta p(t) &= p_2(t+\delta t) - p_1(t) \\
\delta \nu(t) &= \nu_2(t+\delta t) - \nu_1(t)
\end{aligned} \tag{2.5.19}$$

where δt is a small time variation which in itself can vary with time; i.e.,

$$\delta t = \delta t(t) \tag{2.5.20}$$

Further, require that $\delta t(t)$ be selected so that at the switching times, ($t_i^{(2)}$ along the neighboring or second solution and $t_i^{(1)}$ along the nominal or first solution) the equation

$$t_i^{(2)} = t_i^{(1)} + \delta t(t_i^{(1)}) \tag{2.5.21}$$

holds. In this case,

$$\delta \nu(t_i^{(1)}) = \nu_2(t_i^{(2)}) - \nu_1(t_i^{(1)}) = 0$$

and since $\delta \nu$ differs from zero only in the vicinity of the switching times t_i, it follows that

$$\delta \nu \equiv 0 \tag{2.5.22}$$

By this device, the difficulty arising from the finiteness of $\delta \nu$ in Eq. (2.5.18) is eliminated.

From Eq. (2.5.19), the perturbation equations take the form

$$\begin{aligned}
\delta \dot{x} &= \left(\frac{\partial f}{\partial x}\right)^T \delta x + \left(\frac{\partial f}{\partial u}\right)^T \delta u + f\, \delta \dot{t} \\
\delta \dot{p} &= -\frac{\partial^2 H}{\partial x\, \partial p}\, \delta p - \frac{\partial^2 H}{\partial x^2}\, \delta x - \frac{\partial^2 H}{\partial x\, \partial u}\, \delta u - \frac{\partial H}{\partial x}\, \delta \dot{t}
\end{aligned} \tag{2.5.23}$$

where the quantity δu is computed from Eq. (2.4.59) to be

$$\delta u = -\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}\left\{\frac{\partial^2 H}{\partial u\,\partial x}\,\delta x + \frac{\partial^2 H}{\partial u\,\partial p}\,\delta p\right\} \tag{2.5.24}$$

Note that the matrix $(\partial^2 H/\partial u^2)$ is again assumed to have nonvanishing determinant.

Substitution of (2.5.24) into (2.5.23) provides the system

$$\begin{bmatrix}\delta\dot{x}\\ \delta\dot{p}\end{bmatrix} = \begin{bmatrix}C^{①}(t) & C^{②}(t)\\ C^{③}(t) & -C^{①T}(t)\end{bmatrix}\begin{bmatrix}\delta x\\ \delta p\end{bmatrix} + \begin{bmatrix}\dot{x}_1\\ \dot{p}_1\end{bmatrix}\delta\dot{t} \tag{2.5.25}$$

where the $C^{①}$ are n by n matrices given by

$$C^{①} = (C^{①}_{i,j}) = \left[\frac{\partial f_i}{\partial x_j} - \sum_{k,l}\frac{\partial f_i}{\partial u_k}\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}_{l,k}\frac{\partial^2 H}{\partial u_l\,\partial x_j}\right]$$

$$C^{②} = (C^{②}_{i,j}) = -\left[\sum_{k,l}\frac{\partial f_i}{\partial u_k}\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}_{l,k}\frac{\partial^2 H}{\partial u_l\,\partial p_j}\right]$$

$$C^{③} = (C^{③}_{i,j}) = \left[-\frac{\partial^2 H}{\partial x_i\,\partial x_j} + \sum_{k,l}\frac{\partial^2 H}{\partial x_i\,\partial u_k}\left(\frac{\partial^2 H}{\partial u^2}\right)^{-1}_{l,k}\frac{\partial^2 H}{\partial u_l\,\partial x_j}\right] \tag{2.5.26}$$

The solution to (2.5.25) takes the form

$$\begin{bmatrix}\delta x_f\\ \delta p_f\end{bmatrix} = \Lambda(t_f)\left\{\begin{bmatrix}\delta x_0\\ \delta p_0\end{bmatrix} + \sum_{i=1}^{K}\Lambda^{-1}(t_i)\,\delta t_i\begin{bmatrix}\dot{x}_1(t_i^-) - \dot{x}_1(t_i^+)\\ \dot{p}_1(t_i^-) - \dot{p}_1(t_i^+)\end{bmatrix}\right\} \tag{2.5.27}$$

where $\Lambda(t)$ is the fundamental $2n$ by $2n$ matrix solution to the system

$$\dot{\Lambda} = C(t)\Lambda \; ; \quad \Lambda(t_0) = I$$

where $C(t)$ has the components given by Eq. (2.5.26), where the summation in Eq. (2.5.27) is over the number of switches and where δt_i indicates the difference in the i^{th} switch time between the neighboring (second) and nominal (first) solutions. It is assumed that the nominal solution contains K switches. For any time t between the j^{th} and $j+1^{th}$ switch, the solution to (2.5.25) takes the form

$$\begin{bmatrix} \delta x(t) \\ \delta p(t) \end{bmatrix} = \Lambda(t) \left\{ \begin{bmatrix} \delta x_0 \\ \delta p_0 \end{bmatrix} + \sum_{i=1}^{j} \Lambda^{-1}(t_i) \delta t_i \begin{bmatrix} \dot{x}_i(t_i^-) - \dot{x}_i(t_i^+) \\ \dot{p}_i(t_i^-) - \dot{p}_i(t_i^+) \end{bmatrix} \right\} + \delta t(t) \begin{bmatrix} \dot{x}_i(t) \\ \dot{p}_i(t) \end{bmatrix} \tag{2.5.28}$$

Now, the values of δt_i can be calculated from the condition contained in Eq. (2.5.11) that the switching function vanish at the switching times; that is

$$\theta\left[x(t_i), p(t_i), u(t_i)\right] = 0 \quad ; \; i = 1, K$$

Expanding θ about the nominal solution at time $t_i^{\circledcirc}$ provides

$$\delta\theta \sum_{j=1}^{n} \left(\frac{\partial \theta}{\partial x_j} \delta x_j + \frac{\partial \theta}{\partial p_j} \delta p_j \right) + \sum_{k=1}^{r-1} \frac{\partial \theta}{\partial u_k} \delta u_k = 0 \qquad at \; t = t_i^{\circledcirc} \; ; \; i = 1, K \tag{2.5.29}$$

and substituting Eq. (2.5.24) into (2.5.29) yields the K conditions

$$\left(\sum_{j=1}^{n} \alpha_j \delta x_j(t_i^{\circledcirc}) + \beta_j \delta p_j(t_i^{\circledcirc}) = 0 \; ; \; i = 1, K \right.$$
$$\left. \alpha \cdot \delta x(t_i^{\circledcirc}) + \beta \cdot \delta p(t_i^{\circledcirc}) = 0 \right. \tag{2.5.30}$$

where α and β are 1 by n row matrices given by

$$\alpha = (\alpha_j) = \left(\frac{\partial \theta}{\partial x_j} - \sum_{k,l} \frac{\partial \theta}{\partial u_k} \left[\frac{\partial^2 H}{\partial u^2} \right]^{-1}_{k,l} \frac{\partial^2 H}{\partial u_l \partial x_j} \right)$$

$$\beta = (\beta_j) = \left(\frac{\partial \theta}{\partial p_j} - \sum_{k,l} \frac{\partial \theta}{\partial u_k} \left[\frac{\partial^2 H}{\partial u^2} \right]^{-1}_{k,l} \frac{\partial^2 H}{\partial u_l \partial p_j} \right) \qquad (2.5.31)$$

which are evaluated at the K switching points $t_i^{\oplus}$.

Now, at the switch point $t_i^{\oplus}$ it follows from Eq. (2.5.28) that

$$\begin{bmatrix} \delta x(t_i^{\oplus}) \\ \delta p(t_i^{\oplus}) \end{bmatrix} = \Lambda(t_i^{\oplus}) \left\{ \begin{bmatrix} \delta x_0 \\ \delta p_0 \end{bmatrix} + \sum_{j=1}^{i-1} \delta t_j \Lambda(t_j^{\oplus}) \begin{bmatrix} \dot{x}_1(t_j^-) - \dot{x}_1(t_j^+) \\ \dot{p}_1(t_j^-) - \dot{p}_1(t_j^+) \end{bmatrix} \right\}$$
$$+ \begin{bmatrix} \dot{x}_1(t_i^-) \\ \dot{p}_1(t_i^-) \end{bmatrix} \delta t_i \qquad (2.5.32)$$

Thus, Eqs. (2.5.30) and (2.5.32) can be used to evaluate the changes in the switching times, δt_i . By straightforward algebraic manipulation, it can be shown that

$$\delta t_i = - \frac{1}{\dot{\theta}(t_i^{\oplus})} \left[\alpha(t_i^{\oplus}), \beta(t_i^{\oplus}) \right] \Lambda(t_i^{\oplus}) \left\{ \prod_{j=0}^{i=1} (I + \gamma(t_j^{\oplus})) \begin{pmatrix} \delta x_0 \\ \delta p_0 \end{pmatrix} \right\} \qquad (2.5.33)$$

where I is the $2n$ by $2n$ unit matrix, $\gamma(t_j^{\oplus})$ is the $2n$ by $2n$ matrix given by

$$\gamma(t_j^{\oplus}) = 0 \qquad j = 0$$

$$\gamma(t_j^{\oplus}) = \frac{1}{\dot{\theta}(t_j^{\oplus})} \Lambda^{-1}(t_j^{\oplus}) \begin{bmatrix} \dot{x}_1(t_j^+) - \dot{x}_1(t_j^-) \\ \dot{p}_1(t_j^+) - \dot{p}_1(t_j^-) \end{bmatrix} \left[\alpha(t_j^{\oplus}), \beta(t_j^{\oplus}) \right] \Lambda(t_j^{\oplus}) \qquad (2.5.34)$$

and with $\dot{\theta}$ a scalar given by *

$$\dot{\theta}(t_i^{\ominus}) = \left[\alpha(t_i^{\ominus}), \beta(t_i^{\ominus}) \right] \begin{pmatrix} \dot{x}_1(t_i^{\ominus}) \\ \dot{p}_1(t_i^{\ominus}) \end{pmatrix} \qquad (2.3.35)$$

* Note the additional requirements on the nominal solution that $\dot{\theta}(t_i^{\circ}) \neq 0$

Substitution of (2.5.33) into (2.5.27) yields the desired result

$$\begin{pmatrix} \delta x_f \\ \delta p_f \end{pmatrix} = \Lambda(t_f) \left\{ \prod_{i=1}^{K} \left(I + \gamma(t_i^{(1)}) \right) \right\} \begin{pmatrix} \delta x_o \\ \delta p_o \end{pmatrix} \qquad (2.5.36)$$

Recall that the objective of the neighboring extremal technique is to correct the initial p vector, p_{o_1}, by an amount δp_o so that the terminal conditions of Eq. (2.5.15) are satisfied. The correction procedure of Eq. (2.5.16) requires the evaluation of the derivative matrix $(\partial \eta / \partial p_o)$ and the last few paragraphs have been concerned with the development of Eq. (2.5.36) which is needed in this evaluation. The derivative matrix $(\partial \eta / \partial p_o)$ can be represented by

$$\left(\frac{\partial \eta}{\partial p_o} \right) = \left(\frac{\partial \eta}{\partial p_f} \right) \left(\frac{\partial p_f}{\partial p_o} \right) + \left(\frac{\partial \eta}{\partial x_f} \right) \left(\frac{\partial x_f}{\partial p_o} \right) \qquad (2.5.37)$$

Note that the matrices $(\partial \eta / \partial p_f)$ and $(\partial \eta / \partial x_f)$ can be evaluated analytically from Eq. (2.5.15). Further, the matrices $(\partial x_f / \partial p_o)$ and $(\partial p_f / \partial p_o)$ are easily computed once the $2n$ by $2n$ matrix $(\Lambda(t_f) \prod_{i=1}^{K} \{ I + \gamma(t_i^{(1)}) \})$ has been found. This step is accomplished by partitioning this $2n$ by $2n$ matrix into four n by n matrices as

$$\Omega = \Lambda(t_f) \prod_{i=1}^{K} \left\{ I + \gamma(t_i^{(1)}) \right\} = \begin{pmatrix} \Omega^{(1)} & \Omega^{(2)} \\ \Omega^{(3)} & \Omega^{(4)} \end{pmatrix} \qquad (2.5.38)$$

Now, since δx_o in (2.5.36) is zero, it follows that Eq. (2.5.37) takes the form

$$\left(\frac{\partial \eta}{\partial p_o} \right) = \left(\frac{\partial \eta}{\partial p_f} \right) \Omega^{(4)} + \left(\frac{\partial \eta}{\partial x_f} \right) \Omega^{(2)} \qquad (2.5.39)$$

which is now easily evaluated.

The step by step calculation procedure is as follows:

(1) Guess p_{o_1} and integrate the system

$$\dot{x} = f(x, u, v, t) = 0$$

$$\dot{p} = -\frac{\partial H}{\partial x}$$

$$\frac{\partial H}{\partial u} = 0 \; ; \; v = v_{MAX} \quad \theta > 0$$

$$v = v_{MIN} \quad \theta < 0$$

(2) Integrate the system $\dot{\Lambda} = C(t)\Lambda\,;\ \Lambda(t_o) = I$

from t_o to t_f. (Note that since δx_o is zero only a portion of this integration need be performed.)

(3) Compute $\gamma(t_i^{(0)})$ using Eqs. (2.5.34) and (2.5.31).

(4) Evaluate $\Omega^{(2)}$ and $\Omega^{(4)}$ using Eq. (2.5.38).

(5) Calculate $(\partial\eta/\partial p_o)$ from Eq. (2.5.39).

(6) Set $\delta p_o = -(\partial\eta/\partial p_o)^{-1}\eta(x_{f_1}, p_{f_1})$

(7) Update p_o by δp_o

(8) Go to step (2).

The process is continued until the correct p_o vector is found.

2.5.3 Steepest Descent

The steepest descent technique as applied to the nonclassical problem of Eqs. (2.5.6) to (2.5.9) is much the same as that used in the classical Mayer problem of Section 2.4.2. The iteration starts with an admissible control program, u and ν, which satisfies the terminal and transversality conditions of Eqs. (2.5.8) and (2.5.14) but which is not optimal. This program is then corrected in the direction in which the minimizing function, $\phi(x_f)$, is decreasing most rapidly but with the requirement that the corrected program continue to satisfy the problem terminal conditions. Since changes in the control variable ν (denoted by $\delta\nu$) will not be small over the entire trajectory, a resort must again be made to the perturbation method described in Eqs. (2.5.19) to (2.5.23) in which the nominal and corrected solutions are compared at different values of time. By this device, the problem of successively adjusting the variable ν is transformed into that of successively adjusting the switching times at which ν jumps from one extreme value to another.

From Eq. (2.5.23)

$$\delta\dot{x} = \left[\frac{\partial f}{\partial x}\right]^T \delta x + \left[\frac{\partial f}{\partial u}\right]^T \delta u + f\,\delta\dot{t} \qquad (2.5.40)$$

where δt is a function of time and is required to satisfy the conditions

$$t_i^{(2)} = t_i^{(1)} + \delta t\,(t_i^{(1)});\quad i = 1, K$$

The quantities, $t_i^{(2)}$ and $t_i^{(1)}$ denote the i^{th} switch time along the second and first iterations respectively. But, the adjoint equations are

$$\dot{p} = -\frac{\partial H}{\partial x}$$

Thus, the identity

$$p_f^T \delta x_f - p_o^T \delta x_o = \int_{t_o}^{t_f} \left[\frac{\partial H}{\partial u}\right]^T \delta u \, dt + \int_{t_o}^{t_f} H \delta \dot{t} \, dt \qquad (2.5.41)$$

immediately follows. Since the state of the system is specified initially and since p_f is required to satisfy the transversality condition of Eq. (2.5.14); that is

$$p + \frac{\partial \phi}{\partial x} + \left[\frac{\partial \psi}{\partial x}\right]^T \nu = 0 \; ; \; t = t_f \qquad (2.5.42)$$

Eq. (2.5.41) takes the form

$$d\phi + (d\psi)^T \nu = -\int_{t_o}^{t_f} \frac{\partial H}{\partial u} \delta u \, dt - \int_{t_o}^{t_f} H \delta \dot{t} \, dt \qquad (2.5.43)$$

Now, in the neighboring extremal technique of the previous section, only the values which δt took at the K distinct switching times ($t_i^{(2)}$, $i = 1, K$) entered into the analysis. The form of δt on the intervals between switches was of no concern. In the steepest descent technique, however, the functional form of $\delta t(t)$ must be specified beforehand. For simplicity $\delta \dot{t}$ will be taken to be a constant on the intervals between switches. Specifically, on the interval between the i-1th and ith switch, $(t_{i-1}^{(2)}, t_i^{(2)})$, let

$$\delta t(t) = \delta t_{i-1} + a_{i-1}(t - t_{i-1}^{(1)}) \qquad (2.5.44)$$

with

$$a_{i-1} = \frac{\delta t_i - \delta t_{i-1}}{t_i^{(1)} - t_{i-1}^{(1)}} \; ; \quad a_K = \frac{-\delta t_K}{t_f - t_K^{(1)}} \; ; \; \delta t_f = 0 \qquad (2.5.45)$$

Substitution of Eq. (2.5.44) into (2.5.43) provides

$$d\phi + (d\psi)^T \nu = \int_{t_o}^{t_f} -\frac{\partial H}{\partial u} \delta u \, dt - \sum_{i-1}^{K} \delta t_i \, (K_{i-1} - K_i) \qquad (2.5.46)$$

where

$$K_o = \int_{t_o}^{t_1^{(1)}} \frac{H}{t_1^{(1)} - t_o}\, dt \qquad ; K_K = \int_{t_K}^{t_f} \frac{H}{t_f - t_K^{(1)}}\, dt$$

$$K_i = \int_{t_i^{(1)}}^{t_{i+1}^{(1)}} \frac{H}{t_{i+1}^{(1)} - t_i^{(1)}}\, dt \tag{2.5.47}$$

Now as in the classical treatment, the quantity $d\phi$ is to be minimized subject to $d\psi = 0$ and subject also to some additional constraint on the amount of correction to be made. In the classical case, it was required that

$$\int_{t_o}^{t_f} \delta u^2\, dt = k_1^2 \tag{2.5.48}$$

Here, however, a restriction must also be placed on the amount by which the switch times can change; i.e., the δt_i . There are several ways in which such a restriction could be included. For example, in addition to (2.5.48), it might also be required that

$$\sum_{i=1}^{K} \delta t_i^2 = k_2^2 \tag{2.5.49}$$

Alternately, the corrections in δt_i and δu might be made sequentially, rather than simultaneously, by first correcting δu under the conditions that $\delta t_i = 0$ and then correcting δt_i with $\delta u = 0$. However, in the treatment to follow, Eqs. (2.5.48) and (2.5.49) will be combined into the single condition

$$\int_{t_o}^{t_f} \delta u^2 dt + \sum_{i=1}^{K} \delta t_i^2 = k^2 \tag{2.5.50}$$

Introducing the additional multiplier λ , the optimizing condition of (2.5.46) becomes

$$d\phi + (d\psi)\mu^T = -\int_{t_o}^{t_f} \frac{\partial H}{\partial u}\, \delta u\, dt - \sum_{i=1}^{K} \delta t_i (K_{i-1} - K_i) + \lambda \left\{ \int_{t_o}^{t_f} \delta u^2 dt + \sum_{i=1}^{K} \delta t_i^2 - k^2 \right\} \tag{2.5.51}$$

Thus, equating the first variation to zero provides

$$\delta u = \frac{1}{2\lambda} \frac{\partial H}{\partial u}$$

$$\delta t_i = \frac{1}{2\lambda} (K_{i-1} - K_i) \qquad (2.5.52)$$

and it follows that

$$d\phi + (d\psi)_\mu^T = -\frac{1}{2\lambda} \left\{ \int_{t_o}^{t_f} \left(\frac{\partial H}{\partial u} \right)^2 dt + \sum_{i=1}^{K} (K_{i-1} - K_i)^2 \right\} \qquad (2.5.53)$$

The multiplier, λ, is determined by substituting Eqs. (2.5.52) in (2.5.50). The μ_i are determined by integrating the adjoint equations backwards $M+1$ times where M is the number of terminal constraints of the form $\psi_i = 0$. The terminal value of p_f for each integration is

$$\begin{aligned} p_f^{(1)} &= -\partial \psi_1 / \partial x \\ &\vdots \\ p_f^{(m)} &= -\partial \psi_m / \partial x \qquad \text{at } t = t_f \\ p_f^{(M+1)} &= -\partial \phi / \partial x \end{aligned}$$

From Eq. (2.5.42), it follows that the adjoint vector to be used in the calculations has the terminal value

$$p_f = p_f^{(M+1)} + \sum_{j=1}^{M} \mu_j p_f^{(j)}$$

and that

$$d\psi_i = -p_f^{(i)} \delta x_f = -\frac{1}{2\lambda} \left\{ \int_{t_o}^{t_f} \left(p^{(i)} \frac{\partial f}{\partial u} \right) \left(p \frac{\partial f}{\partial u} \right) dt + \sum_{j=1}^{K} \left(K_{j-1}^{(i)} - K_j^{(i)} \right) \left(K_{j-1} - K_j \right) \right\} = 0$$

This solution represents a system of M equations in M unknowns (the μ_i) and can be evaluated numerically. The μ_i, so determined, are then used to evaluate the adjoint vector, p, where

$$p = p^{(M+1)} + \sum \mu_j p^{(j)}$$

This vector then allows the control correction (δu) and δt_i to be computed.

The step by step computation procedure takes the following form:

(1) Select $u_1(t)$ and $\nu_1(t)$ for which the terminal conditions are satisfied. $\psi(x_f) = 0$

(2) Integrate the state system $\dot{x} = f(x, u, \nu, t)$ from t_o to t_f.

(3) Integrate the adjoint system $\dot{p} = -\dfrac{\partial H}{\partial x}$ backwards $M+1$ times with the $M+1$ terminal condition

$$p_f^{(1)} = -\partial \psi_1 / \partial x$$
$$p_f^{(M)} = -\partial \psi_M / \partial x$$
$$p_f^{(M+1)} = -\partial \phi / \partial x$$

(4) Set

$$p = p^{(M+1)} + \sum_{j=1}^{M} \mu_j p^{(j)}$$

where μ_j are determined from

$$\int_{t_o}^{t_f} \left(p^{(j)} \frac{\partial f}{\partial u}\right)^T \left(p \frac{\partial f}{\partial u}\right) dt + \sum_{i=1}^{K} \left(K_{i-1}^{(j)} - K_i^{(j)}\right)\left(K_{i-1} - K_i\right) = 0$$

and where the K_i are given by Eq. (2.5.47)

(5) Determine λ from

$$\frac{1}{4\lambda^2}\left\{\int_{t_o}^{t} \left(\frac{\partial H}{\partial u}\right)^2 dt + \sum_{i=1}^{K} \left(K_{i-1} - K_i\right)^2 = k^2\right.$$

(6) Determine δt_i from

$$\delta t_i = \frac{1}{2\lambda}\left(K_{i-1} - K_i\right)$$

(7) Compute a_{i-1} from

$$a_{i-1} = \frac{\delta t_i - \delta t_{i-1}}{t_i^{(1)} \quad t_{i-1}^{(1)}}$$

(8) Compute

$$\tau = t + \delta(t)$$

where

$$\delta t(t) = \delta t_{i-1} + a_{i-1}\left(t - t_{i-1}^{(1)}\right) \quad ; \quad t_{i-1}^{(1)} < t < t_i^{(1)}$$

(9) Compute the new switch times

$$t_i^{(2)} = t_i^{(1)} + \delta t_i$$

(10) Update the control

$$u_2(\tau) = u_1(t) + \frac{1}{2\lambda} \frac{\partial H}{\partial u}$$

(11) Go to step (2)

Note that in the process, the number of switch times may decrease, but may never increase. If the starting control program is selected so that $\nu(t)$ has K switches while the optimal program has $K-1$, then it is not unlikely that the process will converge to the correct solution. However, if the optimal solution contains $K+1$ switches, the iteration process contains no mechanism for increasing a starting guess of K switches to $K+1$. Hence, the iteration process will not converge on the optimal solution. Most likely, the process will converge on the solution which is optimal under the additional restrictions that $\nu(t)$ contain K or fewer switches.

2.5.4 Quasilinearization

As in the classical Mayer problem, the iteration starts with the functions $p_1(t)$, $x_1(t)$, $u_1(t)$ and $\nu_1(t)$ which satisfy the boundary conditions

$$x = x_o \quad AT \quad t = t_o$$
$$\eta(x_f, p_f) = 0 \quad AT \quad t = t_f \tag{2.5.54}$$

but do not satisfy the differential equations

$$\dot{x} = f(x, u, \nu, t)$$
$$\dot{p} = -\frac{\partial H}{\partial x} \tag{2.5.55}$$

or the optimizing conditions

$$\frac{\partial H}{\partial u} = 0$$
$$\nu = \begin{cases} \nu_{MAX} ; & \theta > 0 \\ \nu_{MIN} ; & \theta < 0 \end{cases}$$

Subsequent programs, $p_2(t)$, $u_2(t)$ and $x_2(t)$ are computed so as to minimize the second-order approximation to J where

$$J = \phi + \psi_M^T + \int_{t_o}^{t_T} p^T(\dot{x} - f)\,dt$$

Using the device of comparing the two solutions at different values of time, the perturbation equations take the form

$$\delta\dot{x} = f - \dot{x}_1 + \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial u}\delta u + f\delta\dot{t}$$

$$\delta\dot{p} = -\frac{\partial H}{\partial x} - \dot{p}_1 - \frac{\partial^2 H}{\partial x^2}\delta x - \frac{\partial^2 H}{\partial p \partial x}\delta p - \frac{\partial^2 H}{\partial u \partial x}\delta u - \frac{\partial H}{\partial x}\delta\dot{t} \quad (2.5.56)$$

where $\delta t(t)$ is to be determined so that the i^{th} switch along the second and first solutions are related by

$$t_i^{(2)} = t_i^{(1)} + \delta t\,(t_i^{(1)}) \quad (2.5.57)$$

This last condition provides

$$\delta w \equiv 0$$

The change in control δu is now determined from the equation

$$\frac{\partial H}{\partial u} + \frac{\partial^2 H}{\partial u^2}\delta u + \frac{\partial^2 H}{\partial u \partial x}\delta x + \frac{\partial^2 H}{\partial u \partial p}\delta p = 0 \quad (2.5.58)$$

Substituting (2.5.58) into (2.5.56) and noting the identities

$$\delta u = u(t+\delta t) - u_1(t)$$

$$\delta x = x(t+\delta t) - x_1(t)$$

$$\delta p = p(t+\delta t) - p_1(t) \quad (2.5.59)$$

provides

$$\begin{pmatrix} \dot{x}_2(\tau) \\ \dot{p}_2(\tau) \end{pmatrix} = C(t)\begin{pmatrix} x_2(\tau) \\ p_2(\tau) \end{pmatrix} + \begin{pmatrix} d^{(1)}(t) \\ d^{(2)}(t) \end{pmatrix} + \begin{pmatrix} f \\ -\frac{\partial H}{\partial x} \end{pmatrix}\delta\dot{t} \quad (2.5.60)$$

where $C^{(i)}$ and $d^{(i)}$ are given by Eqs. (2.4.64) and (2.4.89), respectively, and where τ and t are related by

$$\tau = t + \delta t(t) \tag{2.5.61}$$

Eq. (2.5.60) is to satisfy the boundary conditions

$$x = x_0 \quad AT \; t_0$$

$$\frac{\partial \eta}{\partial x_f}(x_{f_2} - x_{f_1}) + \frac{\partial \eta}{\partial p_f}(p_{f2} - p_{f_1}) = 0 \tag{2.5.62}$$

The time history of $\delta t(t)$ is arbitrary except for the corner condition of Eq. (2.5.57) which requires that the updated switching function, θ, vanish at the new switching times; that is

$$\theta\left(x_1(t_i), p_1(t_i), u_1(t_i)\right) + \frac{\partial \theta}{\partial x}\delta x + \frac{\partial \theta}{\partial p}\delta p + \frac{\partial \theta}{\partial u}\delta u = 0; \; i = 1, K \tag{2.5.63}$$

where K is the number of switches. This condition, which determines $\delta t(t_i^{\textcircled{0}})$, can be rewritten as

$$\left[\alpha(t_i^{\textcircled{1}}), \beta(t_i^{\textcircled{1}})\right]\begin{pmatrix} x_2 - x_1 \\ p_2 - p_1 \end{pmatrix} = -\theta(x_1, p_1, u) \; ; \; t = t_i^{\textcircled{1}}; \; i = 1, K \tag{2.5.64}$$

where α and β are given by Eq. (2.5.31).

Eqs. (2.5.60) to (2.5.64) can be solved for x_2 and p_2 as soon as the form of $\delta t(t)$ is specified on the intervals between switches. Since this function is arbitrary, it should be chosen to augment the iterative process in both the rate of convergence and computation required. Undoubtedly, the simplest choice in regard to the amount of computation is to take δt as a series of delta functions such that δt is zero on the interval between switches while taking on the appropriate value at the switching corners. In this case, however, the behavior of x_2, p_2 and u_2 on the intervals $(t_i^{\textcircled{0}}, t_i^{\textcircled{2}})$ would be lost since the time variable, τ in Eq. (2.5.61), would jump discontinuously from $t_i^{\textcircled{0}}$ to $t_i^{\textcircled{2}}$.

Alternately, $\delta t(t)$ could be equated to some smooth function closely approximating a sequence of step functions. By this smoothing process, the behavior of x_2 and p_2 on $(t_i^{\textcircled{0}}, t_i^{\textcircled{2}})$ would be retained. But, unless considerable care were taken, these values would contain errors which would influence the switching time adjustments on the next iteration as indicated in Eq. (2.5.63) (i.e., the change in switch times is a function of the state and adjoint values at the previously iterated switching points).

In the development here, $\delta\dot{t}$ will be taken to be constant. As in the steepest descent method, define $\delta t(t)$ by

$$\delta t(t) = \delta t_{i-1} + a_{i-1}(t - t_{i-1}^{①}) \; ; \; t_{i-1}^{①} \le t \le t_i^{①} \qquad (2.5.65)$$

with

$$a_{i-1} = \frac{\delta t_i^{①} - \delta t_{i-1}^{①}}{t_i^{①} - t_{i-1}^{①}} \quad ; \quad a_K = \frac{-\delta t_K}{t_f - t_K^{①}} \qquad (2.5.66)$$

With $\delta t(t)$ defined, the variables x_2 and p_2 can now be determined.

The solutions to Eq. (2.5.60) can be represented by

$$\begin{pmatrix} x_2(\tau) \\ p_2(\tau) \end{pmatrix} = \Lambda(t)\left\{ \begin{pmatrix} x_o \\ p_{o_2} \end{pmatrix} + \int_{t_o}^{t} \Lambda^{-1} \begin{pmatrix} d^{①} \\ d^{②} \end{pmatrix} dt + \int_{t_o}^{t} \Lambda^{-1} \begin{pmatrix} f \\ -\frac{\partial H}{\partial x} \end{pmatrix} \delta\dot{t}\, dt \right. \qquad (2.5.67)$$

This solution involves $n+K$ parameters, the initial p vector, p_{o_2}, and the K distinct but constant values of the derivative $\delta\dot{t}$ on intervals between switches (i.e., from Eq. (2.5.66), these constants take the values $a_o, a_1, \ldots a_{K-1}$). These $n+K$ parameters are to be determined to satisfy the n terminal conditions of Eq. (2.5.62) and the K corner conditions of Eq. (2.5.64). Since these equations are linear and since the parameters appear linearly in Eq. (2.5.67), a closed-form solution is possible. The step by step calculation procedure is as follows:

(1) Select an $x_1(t)$, $p_1(t)$, $u_1(t)$ and $\nu(t)$ which satisfy the boundary conditions

$$x = x_o \; ; \; t = t_o$$
$$\eta(x_f, p_f) = 0 \; ; \; t = t_f$$

(2) Develop the fundamental matrix solution

$$\dot{\Lambda} = C(t)\Lambda \; ; \; \Lambda(t_o) = I$$

where the C matrix has components given in Eq. (2.5.26)

(3) Compute the quantities $\alpha(t_i^{②})$ and $\beta(t_i^{②})$ using Eq. (2.5.31)

(4) Compute the value of the $2n \times 1$ vector

$$\delta_{i-1} = \int_{t_{i-1}^{①}}^{t_i^{①}} \Lambda^{-1} \begin{pmatrix} f \\ -\frac{\partial H}{\partial x} \end{pmatrix} dt \qquad (2.5.68)$$

(5) Using Eqs. (2.5.66) to (2.5.68), evaluate x_2 and p_2 at the corner and final times as

$$\begin{pmatrix} x_2(\tau_i) \\ p_2(\tau_i) \end{pmatrix} = \Lambda(t_i^{①}) \left\{ \begin{pmatrix} x_0 \\ p_{0_2} \end{pmatrix} + \int_{t_0}^{t_i^{①}} \Lambda^{-1} \begin{pmatrix} d^{①} \\ d^{②} \end{pmatrix} dt + \sum_{j=1}^{i} a_{j-1}\, \delta_{j-1} \right\} \quad (2.5.69)$$

$$i = 1, K$$

$$\begin{pmatrix} x_2(t_f) \\ p_2(t_f) \end{pmatrix} = \Lambda(t_f) \left\{ \begin{pmatrix} x_0 \\ p_{0_2} \end{pmatrix} + \int_{t_0}^{t_f} \Lambda^{-1} \begin{pmatrix} d^{①} \\ d^{②} \end{pmatrix} dt + \sum_{1}^{K+1} a_{j-1}\, \delta_{j-1} \right\} \quad (2.5.70)$$

(6) Use the n terminal conditions of Eq. (2.5.62) and the K corner conditions of (2.5.64) to evaluate the $n+K$ constants, p_{0_2} and a_i $(i=0, K-1)$ appearing in Eqs. (2.5.69) and (2.5.70). From Eq. (2.5.66) the constant a_K is given by

$$a_K = \frac{1}{t_K - t_f} \sum_{i=1}^{K} a_{i-1} \left(t_i^{①} - t_{i-1}^{①} \right)$$

(7) Calculate $\delta t(t)$ from Eq. (2.5.65) using the values of a_i computed above. Set

$$\tau = t + \delta t(t)$$

(8) With $\delta t(t)$ and p_{0_2} known determine the solution

$$x_2(t) = x_2(\tau(t)) = x_2(t + \delta t(t))$$

$$p_2(t) = p_2(\tau(t)) = p_2(t + \delta t(t))$$

where $x_2(\tau)$ and $p_2(\tau)$ are computer using Eq. (2.5.67).

(9) Go to step (2).

The process is repeated until additional iteratives provide essentially no change in the time histories of x, p and u. At this point, the solution has been achieved.

2.5.5 Final Time Open

In the development to this point, it has been assumed that the final time, t_f, is specified. In many cases, such as the vehicle problem of Section 2.1, the final time is open and is to be selected so that the total system performance is optimal. As discussed in Section 2.1, this selection is accomplished by requiring that the final time satisfy the additional transversality condition

$$H = \sum_{i=1}^{n} p_i f_i = 0 \; ; \; t = t_f \quad (2.5.71)$$

(see Eq. (2.1.18b)).

When the final time is not specified, the three iterative techniques must be slightly modified to allow for adjustments in the value of t_f . One such modification for the steepest-descent procedure is offered in Ref. (13) and another, which can be used in both second-order techniques, is offered in Ref. (23). In the technique to be presented here, adjustments in the final time are made in exactly the same manner as adjustments in the switching or corner times; that is, successive iterates are compared at different values of time.

Let

$$\begin{aligned} \delta x &= x(t+\delta t) - x_1(t) \\ \delta p &= p(t+\delta t) - p_1(t) \\ \delta u &= u(t+\delta t) - u_1(t) \end{aligned} \tag{2.5.72}$$

where δt is itself a function of time, that is

$$\delta t = \delta t(t)$$

At the final time $\delta t(t_f) = \delta t_f$ is selected so that

$$\delta t_f = t_f^{(2)} - t_f^{(1)} \tag{2.5.73}$$

The criterion for selection depends on the particular iterative process being used.

In the neighboring extremal or quasilinear techniques, the quantity dt_f would be selected so that the linearized version of the transversality condition in Eq. (2.5.71) holds. That is,

$$H(p_1, x_1, u_1) + \frac{\partial H}{\partial x}\,\delta x_f + \frac{\partial H}{\partial p}\,\delta p_f + \frac{\partial H}{\partial u}\,\delta u_f = 0 \tag{2.5.74}$$

Thus, the terminal time would be treated exactly the same as a switch time except that instead of a linearized switching condition being used to determine the correction δt_f , Eq. (2.5.74) would be used.

In the steepest descent technique the magnitude constraint of Eq. (2.5.50)

$$\int_{t_o}^{t_f} \delta u^2 \, dt + \sum_{i=1}^{K} \delta t_i^2 = k^2$$

would be replaced by

$$\int_{t_o}^{t_f} \delta u^2 \, dt + \sum_{i=1}^{K} \delta t_i^2 + \delta t_f^2 = k^2 \qquad (2.5.75)$$

The quantity δt_f would be chosen as are the other δt_i, so as to minimize the first-order approximation to the performance index but subject to the above magnitude constraints.

In both first and second-order techniques, the final time open situation is treated in exactly the same manner as a switching corner in the "bang-bang" problem. This holds for classical and nonclassical Mayer problems. Note that in the neighboring extremal technique, only the value of δt_f at $t_f^{(1)}$ enters into the analysis with the functional form of $\delta t(t)$ immaterial. This is due to the fact that in the calculations to be performed δt_f is treated as an infinitesimal. In both the steepest descent and quasilinear methods δt_f is a finite quantity and the behavior of the solution on the interval $(t_f^{(1)}, t_f^{(2)})$ is important. Hence, the form of $\delta t(t)$ does enter into the analysis and it is important to choose this form so that the best accuracy is achieved while keeping the computation required to a minimum.

3.0 RECOMMENDED PROCEDURES

It has been shown that the iterative numerical procedures used for solving nonlinear optimization problems consist in approximating the problem by a succession of linear problems. The three techniques presented, steepest descent, neighboring extremal and quasilinearization are essentially extrapolation into the variational domain of techniques that have been used for over two hundred years in ordinary maxima-minima theory. The step by step calculation procedure to be used in each of these techniques has been tabulorized in Sections 2.4.2, 2.4.3, and 2.4.4 for the classical Mayer problem and in Sections 2.5.2, 2.5.3, and 2.5.4 for the nonclassical problem. Also, the relative merits of each technique were discussed in Section 2.4.5. All other techniques represent only minor variations or combination of the three. Since there exists at the present no nonlinear theory, there is very little more that can be done in this area in regard to the development of new and improved procedures. However, there is much practical work which should be pursued.

It would be of considerable value to extensively experiment with the three iterative procedures in connection with a variety of problems currently of interest in flight mechanics and control theory. Information relating to the effectiveness of each of these techniques on typical problems in regard to range and rate of convergence, programming difficulty, computer time required, etc. would be highly useful and valuable. Work of this type is currently underway and as more and more numerical experimentation is conducted and disseminated in the literature, the numerical solution of optimization problems, will become more tractable.

4.0 REFERENCES

1. Pontryagin, L. S., et al. The Mathematical Theory of Optimal Processes. Translated from the Russian by K. N. Trirogoff and L. W. Neustadt, Intersicience Publishers (1962).

2. Bellman, R. and Drefus, S. Applied Dynamic Programming. Princeton University Press (1962).

3. Miele, A. "The Calculus of Variations in Applied Aerodynamics and Flight Mechanics," Chapter 4, Optimization Techniques. Academic Press (1962).

4. Mangad, M. and Schwartz, M. The Calculus of Variations and Modern Applications. NAA S&ID SID 65-1200-4 (November 1965).

5. MyIntyre, J. The Pontryagin Maximum Principle. NAA S&ID, SID 65-1200-7 (May 1966).

6. Abbott, A. and McIntyre, J. Dynamic Programming. NAA S&ID, SID 66-1678-2 (November 1966).

7. Bliss, G. A. Lectures on the Calculus of Variations. University of Chicago Press (1946).

8. Hildebrand, F. B. Introduction to Numerical Analysis. McGraw-Hill (1956).

9. Hochstrasser, U. "Numerical Methods for Finding Solutions of Nonlinear Equations," Chapter 7, Survey of Numerical Analysis. McGraw-Hill (1962) edited by John Todd.

10. Broyden, C. G. "A Class of Methods for Solving Nonlinear Simultaneous Equations," Mathematics of Computation. Vol. 19, pp. 577-593 (October 1965).

11. Dreyfus, S. E. Dynamic Programming and the Jacobi Condition of the Calculus of Variations. The Rand Corporation, Memorandum RM-3417-PR (August 1963).

12. Kelley, H. "Method of Gradients," Chapter 6, Optimization Techniques. edited by G. Leitmann, Academic Press, 1962

13. Bryson, A. E. and Denham, W. F. "A Steepest-Ascent Method for Solving Optimum Programming Problems," Journal of Applied Mechanics, Vol. 29, pp. 247-257 (1962).

14. Breakwell, J. V. "The Optimization of Trajectories," Journal for the Society of Industrial Applied Mathematics (June 1959).

15. Breakwell, J. V., Speyer, J. L. and Bryson, A. E. "Optimization and Control of Nonlinear Systems Using the Second Variation," Journal of the Society of Industrial Applied Mathematics Control, Vol. 1, No. 2, (1963).

16. Jazwinski, A. "Optimal Trajectories and Linear Control of Nonlinear Systems," AIAA Journal (August 1964).

17. Jurovics, S. A. and McIntyre, J. E. "Adjoint Method and Its Application to Optimum Trajectory Analysis," AIAA Journal.(September 1962).

18. McGill, R. and Kenneth, P. "Solution to Variational Problems by Means of a Generalized Newton-Raphson Operator," AIAA Journal (October 1964).

19. Bellman, R. E., Kagiwala, H. H. and Kalaba, R. E. Numerical Studies of a Two-Point Nonlinear Boundary Value Problem Using Dynamic Programming, Invariant Imbedding, and Quasilinearization. The Rand Corporation, Memorandum RM-4069-PR (March 1964).

20. McGill, R. and Kenneth, P. A Convergence Theorem on the Iterative Solution of Nonlinear Two-Point Boundary Value Systems. Paper presented at XIVth International Astronautical Federation Congress, Paris, France (September 1963).

21. McCue, G. A. Quasilinearization Determination of Optimum Finite-Thrust Orbital Transfers. NAA S&ID, SID 66-1278 (1966).

22. Lewallen, J. and Tapley, B. Analysis and Comparison of Several Numerical Optimization Methods. AIAA Paper No. 67-58, presented at 5th Aerospace Sciences Meeting, New York (January 1967).

23. Long. R. S. "Newton-Raphson Operator: Problems with Undetermined End Points," AIAA Journal, Vol. 3, No. 7 (July 1965).