4

The Calculus of Variations

A branch of mathematics that is extremely useful in solving optimization problems is the calculus of variations. Queen Dido of Carthage was apparently the first person to attack a problem that can readily be solved by using variational calculus.† Dido, having been promised all of the land she could enclose with a bull's hide, cleverly cut the hide into many lengths and tied the ends together. Having done this, her problem was to find the closed curve with a fixed perimeter that encloses the maximum area. We know that she should have chosen a circle. The calculus of variations enables us to prove this fact and, in addition, other results that are more useful, since real estate transactions are performed somewhat differently today.

Although the history of the calculus of variations dates back to the ancient Greeks, it was not until the seventeenth century in western Europe that substantial progress was made. Sir Isaac Newton used variational principles to determine the shape of a body moving in air that encounters the least resistance. Another problem of historical interest is the brachistochrone problem shown in Fig. 4-1, posed by Johann Bernoulli in 1696. Under the influence of gravity, the bead slides along a frictionless wire with fixed end points A and B. The problem is to find the shape of the wire that causes the bead to move from A to B in minimum time. The solution, a cycloid lying in the vertical plane, is credited to Johann and Jacob Bernoulli, Newton, and L'Hospital.

† See [M-2].

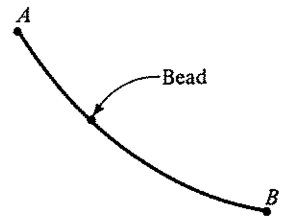


Figure 4-1 The brachistochrone problem

In Dido's problem, and in the brachistochrone problem, curves are sought which cause some criterion to assume extreme values. The connection with the optimal control problem, wherein we seek a control function that minimizes a performance measure, should be apparent.

4.1 FUNDAMENTAL CONCEPTS

In optimal control problems the objective is to determine a function that minimizes a specified *functional*—the performance measure. The analogous problem in calculus is to determine a point that yields the minimum value of a function. In this section we shall introduce some new concepts concerning functionals by appealing to some familiar results from the theory of functions.†

Functionals

To begin, let us review the definition of a function.

DEFINITION 4-1

A *function* f is a rule of correspondence that assigns to each element $\mathbf{q}$ in a certain set $\mathscr{D}$ a unique element in a set $\mathscr{R}$. $\mathscr{D}$ is called the *domain* of f and $\mathscr{R}$ is the *range*.

We shall be considering functions that assign a real number to each point (or vector) in n-dimensional Euclidean space.‡

Example 4.1-1. Suppose $q_1, q_2, \ldots, q_n$ are the coordinates of a point in n-dimensional Euclidean space and

† Appropriate references for functions of real variables are [B-4] and [O-2]. For additional reading on the calculus of variations see [G-1] and [E-1].

‡ It is assumed that the reader is familiar with the concept of a Euclidean space. See [O-2], pp. 293–301 for a detailed exposition.

$$f(\mathbf{q}) = \sqrt{q_1^2 + q_2^2 + \cdots + q_n^2}. \tag{4.1-1}$$

The real number assigned by f is the distance of the point $\mathbf{q}$ from the origin.

The definition of a functional parallels that of a function.

DEFINITION 4-2

A *functional* J is a rule of correspondence that assigns to each function $\mathbf{x}$ in a certain class Ω a unique real number. Ω is called the *domain* of the functional, and the set of real numbers associated with the functions in Ω is called the *range* of the functional.

Notice that the domain of a functional is a class of functions; intuitively, we might say that a functional is a "function of a function."

Example 4.1-2. Suppose that x is a continuous function of t defined in the interval $[t_0, t_f]$ and

$$J(x) = \int_{t_0}^{t_f} x(t)\, dt; \tag{4.1-2}$$

the real number assigned by the functional J is the area under the $x(t)$ curve.

Linearity of Functionals

Let us review the concept of linearity, which will be useful to us later, by considering a function f of $\mathbf{q}$, defined for $\mathbf{q} \in \mathscr{D}$.

DEFINITION 4-3

f is a *linear function* of $\mathbf{q}$ if and only if it satisfies the *principle of homogeneity*

$$f(\alpha\mathbf{q}) = \alpha f(\mathbf{q}) \tag{4.1-3}$$

for all $\mathbf{q} \in \mathscr{D}$ and for all real numbers α such that $\alpha\mathbf{q} \in \mathscr{D}$, *and* the *principle of additivity*

$$f(\mathbf{q}^{(1)} + \mathbf{q}^{(2)}) = f(\mathbf{q}^{(1)}) + f(\mathbf{q}^{(2)}) \tag{4.1-4}$$

for all $\mathbf{q}^{(1)}$, $\mathbf{q}^{(2)}$, and $\mathbf{q}^{(1)} + \mathbf{q}^{(2)}$ in $\mathscr{D}$.†

† In our applications we shall be concerned only with functions of real variables, so α and the components of $\mathbf{q}$ will be real numbers.

Example 4.1-3. If $f(t) = 5t$ for all t, then

$$f(\alpha t) = 5[\alpha t] \tag{4.1-5a}$$

and

$$\alpha f(t) = \alpha[5t]; \tag{4.1-5b}$$

therefore, since

$$5[\alpha t] = \alpha[5t] \tag{4.1-5c}$$

for all t, the principle of homogeneity is satisfied. Now, let us test to see if the property of additivity is satisfied.

$$f(t^{(1)} + t^{(2)}) = 5[t^{(1)} + t^{(2)}] \tag{4.1-6a}$$

and

$$f(t^{(1)}) + f(t^{(2)}) = 5t^{(1)} + 5t^{(2)}; \tag{4.1-6b}$$

thus, since

$$5[t^{(1)} + t^{(2)}] = 5t^{(1)} + 5t^{(2)} \tag{4.1-6c}$$

for all $t^{(1)}$, $t^{(2)}$, the principle of additivity is satisfied. Since the principle of homogeneity *and* the principle of additivity are both satisfied, f is a linear function.

Now consider the function g; with $g(t) = 2/t$ for all $t > 0$, then

$$g(\alpha t) = \frac{2}{\alpha t} \tag{4.1-7a}$$

and

$$\alpha g(t) = \alpha\left[\frac{2}{t}\right] \tag{4.1-7b}$$

Clearly,

$$\frac{2}{\alpha t} \neq \alpha\left[\frac{2}{t}\right] \tag{4.1-7c}$$

for all α; therefore, the principle of homogeneity is not satisfied, and g is a nonlinear function.

Next, we shall define a linear functional. Assume that $\mathbf{x}$ is a function which is a member of some class Ω, and J is a functional of $\mathbf{x}$; that is, to each $\mathbf{x}$ in Ω, J assigns a unique real number.

DEFINITION 4-4

J is a *linear functional* of $\mathbf{x}$ if and only if it satisfies the *principle of homogeneity*

$$J(\alpha\mathbf{x}) = \alpha J(\mathbf{x}) \tag{4.1-8a}$$

for all $\mathbf{x} \in \Omega$ and for all real numbers α such that $\alpha\mathbf{x} \in \Omega$, *and* the *principle of additivity*

$$J(\mathbf{x}^{(1)} + \mathbf{x}^{(2)}) = J(\mathbf{x}^{(1)}) + J(\mathbf{x}^{(2)}) \tag{4.1-8b}$$

for all $\mathbf{x}^{(1)}$, $\mathbf{x}^{(2)}$, and $\mathbf{x}^{(1)} + \mathbf{x}^{(2)}$ in Ω.

Example 4.1-4. Consider the functional

$$J(x) = \int_{t_0}^{t_f} x(t)\, dt, \tag{4.1-9}$$

where x is a continuous function of t. Let us see if this functional satisfies the principles of homogeneity and additivity.

Homogeneity:

$$\alpha J(x) = \alpha \int_{t_0}^{t_f} x(t)\, dt, \tag{4.1-10a}$$

$$J(\alpha x) = \int_{t_0}^{t_f} \alpha x(t)\, dt; \tag{4.1-10b}$$

therefore,

$$J(\alpha x) = \alpha J(x) \tag{4.1-10c}$$

for all real α and for all x and αx in Ω.

Additivity:

$$J(x^{(1)} + x^{(2)}) = \int_{t_0}^{t_f} [x^{(1)}(t) + x^{(2)}(t)]\, dt, \tag{4.1-11a}$$

$$J(x^{(1)}) = \int_{t_0}^{t_f} x^{(1)}(t)\, dt, \tag{4.1-11b}$$

$$J(x^{(2)}) = \int_{t_0}^{t_f} x^{(2)}(t)\, dt; \tag{4.1-11c}$$

therefore,

$$J(x^{(1)} + x^{(2)}) = J(x^{(1)}) + J(x^{(2)}) \tag{4.1-11d}$$

for all $x^{(1)}$, $x^{(2)}$, and $x^{(1)} + x^{(2)}$ in Ω.

Since additivity *and* homogeneity are both satisfied, the functional is linear.

Now consider the functional

$$J(x) = \int_{t_0}^{t_f} x^2(t)\,dt, \tag{4.1-12}$$

where x is a continuous function of t. Again let us ascertain whether homogeneity and additivity are satisfied.

Homogeneity:

$$\begin{aligned} J(\alpha x) &= \int_{t_0}^{t_f} [\alpha x(t)]^2\,dt \\ &= \alpha^2 \int_{t_0}^{t_f} x^2(t)\,dt, \end{aligned} \tag{4.1-13a}$$

$$\alpha J(x) = \alpha \int_{t_0}^{t_f} x^2(t)\,dt. \tag{4.1-13b}$$

Clearly,

$$J(\alpha x) \neq \alpha J(x) \tag{4.1-13c}$$

for all α, so the functional (4.1-12) is *nonlinear*.

Closeness of Functions

If two points are said to be close to one another, a geometric interpretation springs immediately to mind. But what do we mean when we say two *functions* are close to one another? To give a precise meaning to the term "close" we next introduce the concept of a norm.

DEFINITION 4-5

The *norm* in n-dimensional Euclidean space is a rule of correspondence that assigns to each point $\mathbf{q}$ a real number. The norm of $\mathbf{q}$, denoted by $\|\mathbf{q}\|$, satisfies the following properties:

1. $\|\mathbf{q}\| \geq 0$ and $\|\mathbf{q}\| = 0$ if and only if $\mathbf{q} = \mathbf{0}$. (4.1-14a)
2. $\|\alpha\mathbf{q}\| = |\alpha| \cdot \|\mathbf{q}\|$ for all real numbers α. (4.1-14b)
3. $\|\mathbf{q}^{(1)} + \mathbf{q}^{(2)}\| \leq \|\mathbf{q}^{(1)}\| + \|\mathbf{q}^{(2)}\|$. (4.1-14c)

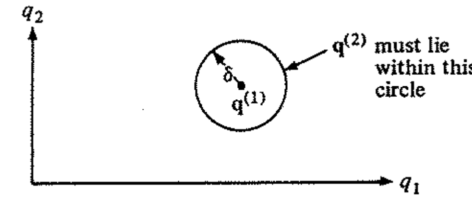
When we say that two points $\mathbf{q}^{(1)}$ and $\mathbf{q}^{(2)}$ are close together, we mean that

$$\|\mathbf{q}^{(1)} - \mathbf{q}^{(2)}\| \text{ is small.}$$

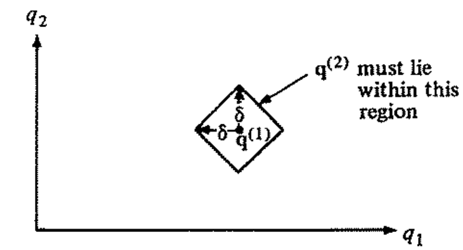
Example 4.1-5. What is a suitable norm for two-dimensional Euclidean space? It is easily verified that

$$\|\mathbf{q}\|_2 \triangleq \sqrt{q_1^2 + q_2^2}, \quad \text{or} \quad \|\mathbf{q}\|_1 \triangleq |q_1| + |q_2|$$

satisfies properties (4.1-14). Now suppose that a point $\mathbf{q}^{(1)}$ is specified and it is required that $\|\mathbf{q}^{(2)} - \mathbf{q}^{(1)}\| < \delta$. What are the acceptable locations for $\mathbf{q}^{(2)}$? If $\|\mathbf{q}\|_2$ is used as the norm, $\mathbf{q}^{(2)}$ must lie within the circle centered at $\mathbf{q}^{(1)}$ having radius δ as shown in Fig. 4-2(a). On the other hand, if $\|\mathbf{q}\|_1$ is used as the norm, the acceptable locations for $\mathbf{q}^{(2)}$ are as shown in Fig. 4-2(b).



(a)



(b)

Fig. 4-2 (a) The set of points that satisfy $\|\mathbf{q}^{(2)} - \mathbf{q}^{(1)}\|_2 < \delta$ (b) The set of points that satisfy $\|\mathbf{q}^{(2)} - \mathbf{q}^{(1)}\|_1 < \delta$

Next, let us define the norm of a function.

DEFINITION 4-6

The *norm of a function* is a rule of correspondence that assigns to each function $\mathbf{x} \in \Omega$, defined for $t \in [t_0, t_f]$, a real number. The norm of $\mathbf{x}$, denoted by $\|\mathbf{x}\|$, satisfies the following properties:

1. $\|\mathbf{x}\| \geq 0$ and $\|\mathbf{x}\| = 0$ if and only if $\mathbf{x}(t) = \mathbf{0}$ for all $t \in [t_0, t_f]$. (4.1-15a)
2. $\|\alpha\mathbf{x}\| = |\alpha| \cdot \|\mathbf{x}\|$ for all real numbers α. (4.1-15b)
3. $\|\mathbf{x}^{(1)} + \mathbf{x}^{(2)}\| \leq \|\mathbf{x}^{(1)}\| + \|\mathbf{x}^{(2)}\|$. (4.1-15c)

To compare the closeness of two functions $\mathbf{y}$ and $\mathbf{z}$ that are defined for $t \in [t_0, t_f]$, let $\mathbf{x}(t) = \mathbf{y}(t) - \mathbf{z}(t)$.

Intuitively speaking, the norm of the difference of two functions should be zero if the functions are identical, small if the functions are "close," and large if the functions are "far apart."

Example 4.1-6. x is a continuous scalar function of t defined in the interval $[t_0, t_f]$. Define an acceptable norm for x.

$$\|x\| = \max_{t_0 \le t \le t_f} \{|x(t)|\} \tag{4.1-16}$$

is a suitable norm because it satisfies the three properties given in (4.1-15).

The Increment of a Functional

In order to consider extreme values of a function, we now define the concept of an increment.

DEFINITION 4-7

If $\mathbf{q}$ and $\mathbf{q} + \Delta\mathbf{q}$ are elements for which the function f is defined, then the *increment* of f, denoted by Δf, is

$$\Delta f \triangleq f(\mathbf{q} + \Delta\mathbf{q}) - f(\mathbf{q}). \tag{4.1-17}$$

Notice that Δf depends on both $\mathbf{q}$ and $\Delta\mathbf{q}$, in general, so to be more explicit we would write $\Delta f(\mathbf{q}, \Delta\mathbf{q})$.

Example 4.1-7. Consider the function

$$f(\mathbf{q}) = q_1^2 + 2q_1q_2 \quad \text{for all real } q_1, q_2. \tag{4.1-18}$$

The increment of f is

$$\begin{aligned}\Delta f = f(\mathbf{q} + \Delta\mathbf{q}) - f(\mathbf{q}) &= [q_1 + \Delta q_1]^2 \\ &\quad + 2[q_1 + \Delta q_1][q_2 + \Delta q_2] - [q_1^2 + 2q_1q_2] \\ &= 2q_1\,\Delta q_1 + [\Delta q_1]^2 + 2\,\Delta q_1 q_2 + 2\,\Delta q_2 q_1 + 2\,\Delta q_1\,\Delta q_2\end{aligned} \tag{4.1-19}$$

In an analogous manner, we next define the increment of a functional.

DEFINITION 4-8

If $\mathbf{x}$ and $\mathbf{x} + \delta\mathbf{x}$ are functions for which the functional J is defined, then the *increment of* J, denoted by ΔJ, is

$$\Delta J \triangleq J(\mathbf{x} + \delta\mathbf{x}) - J(\mathbf{x}). \tag{4.1-20}$$

Again, to be more explicit, we would write $\Delta J(\mathbf{x}, \delta\mathbf{x})$ to emphasize that the increment depends on the functions $\mathbf{x}$ and $\delta\mathbf{x}$. $\delta\mathbf{x}$ is called the *variation* of the function $\mathbf{x}$.

Example 4.1-8. Find the increment of the functional

$$J(x) = \int_{t_0}^{t_f} x^2(t)\,dt, \tag{4.1-21}$$

where x is a continuous function of t.

The increment is

$$\begin{aligned}\Delta J &= J(x + \delta x) - J(x) \\ &= \int_{t_0}^{t_f} [x(t) + \delta x(t)]^2\,dt - \int_{t_0}^{t_f} x^2(t)\,dt \\ &= \int_{t_0}^{t_f} [2x(t)\delta x(t) + [\delta x(t)]^2]\,dt.\end{aligned} \tag{4.1-22}$$

The Variation of a Functional

The preceding definitions have laid the foundation for considering the variation of a functional. The variation plays the same role in determining extreme values of *functionals* as the differential does in finding maxima and minima of *functions*. As review, we next state the definition of the differential of a function.

DEFINITION 4-9

The increment of a function of n variables can be written as

$$\Delta f(\mathbf{q}, \Delta\mathbf{q}) = df(\mathbf{q}, \Delta\mathbf{q}) + g(\mathbf{q}, \Delta\mathbf{q}) \cdot \|\Delta\mathbf{q}\|, \tag{4.1-23}$$

where df is a linear function of $\Delta\mathbf{q}$. If

$$\lim_{\|\Delta\mathbf{q}\| \to 0} \{g(\mathbf{q}, \Delta\mathbf{q})\} = 0,$$

then f is said to be *differentiable* at $\mathbf{q}$, and df is *the differential of* f at the point $\mathbf{q}$.

If f is a differentiable function of *one* variable t, then the differential can be written

$$df(t, \Delta t) = f'(t)\,\Delta t; \tag{4.1-24}$$

$f'(t)$ is called the *derivative* of f at t. Figure 4-3 gives a geometric interpretation of the increment Δf, the differential df, and the derivative f': $f'(t_1)$ is the slope of the line that is tangent to f at the time t_1; $f'(t_1)\,\Delta t$ is a first-order (linear) approximation to Δf (the smaller Δt, the better the approximation).

Example 4.1-9. Find the differential of

$$f(\mathbf{q}) = q_1^2 + 2q_1q_2 \tag{4.1-25}$$

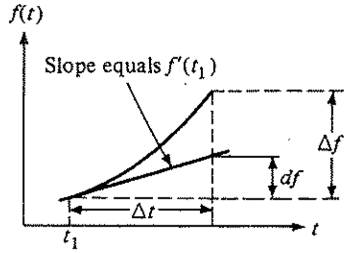


Figure 4-3 Geometric interpretation of $\Delta f, df, f'$

In Example 4.1-7 we found that the increment is

$$\begin{aligned}\Delta f(\mathbf{q}, \Delta\mathbf{q}) = [2q_1 + 2q_2]\,\Delta q_1 + [2q_1]\,\Delta q_2 \\ + [\Delta q_1]^2 + 2\,\Delta q_1\,\Delta q_2.\end{aligned} \tag{4.1-26}$$

The first two terms are linear in $\Delta\mathbf{q}$. Letting

$$\|\Delta\mathbf{q}\| \triangleq \sqrt{[\Delta q_1]^2 + [\Delta q_2]^2}, \tag{4.1-27}$$

we can write the last two terms as

$$\frac{[\Delta q_1]^2 + 2\,\Delta q_1\,\Delta q_2}{\sqrt{[\Delta q_1]^2 + [\Delta q_2]^2}} \cdot \sqrt{[\Delta q_1]^2 + [\Delta q_2]^2}, \tag{4.1-28}$$

which is of the form $g(\mathbf{q}, \Delta\mathbf{q})\cdot\|\Delta\mathbf{q}\|$. To show that f is differentiable we must verify that

$$\lim_{\|\Delta\mathbf{q}\|\to 0} \left\{\frac{[\Delta q_1]^2 + 2\,\Delta q_1\,\Delta q_2}{\sqrt{[\Delta q_1]^2 + [\Delta q_2]^2}}\right\} = 0. \tag{4.1-29}$$

It will be left as an exercise for the interested reader to verify that this limit exists and is zero; hence f is differentiable, and the differential is

$$df(\mathbf{q}, \Delta\mathbf{q}) = [2q_1 + 2q_2]\,\Delta q_1 + [2q_1]\,\Delta q_2. \tag{4.1-30}$$

Rather than go through all of these steps, we can use Definition 4-9 to develop a rule for finding the differential of a function. In particular, if f is a differentiable function of n variables, the differential df is given by

$$df = \frac{\partial f}{\partial q_1}\Delta q_1 + \frac{\partial f}{\partial q_2}\Delta q_2 + \cdots + \frac{\partial f}{\partial q_n}\Delta q_n. \tag{4.1-31}$$

We shall also find it convenient to develop a formal procedure for finding the variation of a functional rather than starting each time from the definition which follows.

DEFINITION 4-10

The increment of a functional can be written as

$$\Delta J(\mathbf{x}, \delta\mathbf{x}) = \delta J(\mathbf{x}, \delta\mathbf{x}) + g(\mathbf{x}, \delta\mathbf{x})\cdot\|\delta\mathbf{x}\|, \tag{4.1-32}$$

where δJ is linear in $\delta\mathbf{x}$. If

$$\lim_{\|\delta\mathbf{x}\|\to 0} \{g(\mathbf{x}, \delta\mathbf{x})\} = 0,$$

then J is said to be *differentiable* on $\mathbf{x}$ and δJ is the *variation of J* evaluated for the function $\mathbf{x}$.

Example 4.10. Let x be a continuous scalar function defined for $t \in [0, 1]$. Find the variation of the functional

$$J(x) = \int_0^1 [x^2(t) + 2x(t)]\,dt. \tag{4.1-33}$$

First, find the increment of J,

$$\begin{aligned}\Delta J(x, \delta x) &= J(x + \delta x) - J(x) \\ &= \int_0^1 \{[x(t) + \delta x(t)]^2 + 2[x(t) + \delta x(t)]\}\,dt \\ &\quad - \int_0^1 [x^2(t) + 2x(t)]\,dt.\end{aligned} \tag{4.1-34}$$

Expanding, and combining these integrals, we obtain

$$\Delta J(x, \delta x) = \int_0^1 \{[2x(t) + 2]\,\delta x(t) + [\delta x(t)]^2\}\,dt. \tag{4.1-35}$$

Separating the terms which are linear in δx, we have

$$\Delta J(x, \delta x) = \int_0^1 \{[2x(t) + 2]\,\delta x(t)\}\,dt + \int_0^1 [\delta x(t)]^2\,dt. \tag{4.1-36}$$

Now let us verify that the second integral can be written

$$\int_0^1 [\delta x(t)]^2\,dt = g(x, \delta x)\cdot\|\delta x\| \tag{4.1-37}$$

and that

$$\lim_{\|\delta x\|\to 0} \{g(x, \delta x)\} = 0. \tag{4.1-38}$$

Since x is a continuous function, let

$$\|\delta x\| \triangleq \max_{0\le t\le 1} \{|\delta x(t)|\}. \tag{4.1-39}$$

Multiplying the left side of (4.1-37) by $\|\delta x\|/\|\delta x\|$ gives

$$\frac{\|\delta x\|}{\|\delta x\|}\cdot\int_0^1[\delta x(t)]^2\,dt=\|\delta x\|\cdot\int_0^1\frac{[\delta x(t)]^2}{\|\delta x\|}\,dt; \tag{4.1-40}$$

the right side of Eq. (4.1-40) follows because $\|\delta x\|$ does not depend on t. Comparing (4.1-40) with (4.1-37), we observe that

$$g(x,\delta x)=\int_0^1\frac{[\delta x(t)]^2}{\|\delta x\|}\,dt. \tag{4.1-41}$$

Writing $[\delta x(t)]^2$ as $|\delta x(t)|\cdot|\delta x(t)|$ gives

$$\int_0^1\frac{|\delta x(t)|\cdot|\delta x(t)|}{\|\delta x\|}\,dt\le\int_0^1|\delta x(t)|\,dt, \tag{4.1-42}$$

because of the definition of the norm of δx, which implies that $\|\delta x\|\ge|\delta x(t)|$ for all $t\in[0,1]$. Clearly, if $\|\delta x\|\to 0$, $|\delta x(t)|\to 0$ for all $t\in[0,1]$, and thus

$$\lim_{\|\delta x\|\to 0}\left\{\int_0^1|\delta x(t)|\,dt\right\}=0. \tag{4.1-43}$$

We have succeeded in verifying that the increment can be written in the form of Eq. (4.1-32) and that $g(x,\delta x)\to 0$ as $\|\delta x\|\to 0$; therefore, the variation of J is

$$\delta J(x,\delta x)=\int_0^1\{[2x(t)+2]\,\delta x(t)\}\,dt. \tag{4.1-44}$$

This expression can also be obtained by formally expanding the integrand of ΔJ in a Taylor series about $x(t)$ and retaining only the terms of first order in $\delta x(t)$.

It is very important to keep in mind that δJ is the linear approximation to the difference in the functional J caused by two comparison curves. If the comparison curves are close ($\|\delta \mathbf{x}\|$ small), then the variation should be a good approximation to the increment; however, δJ may be a poor approximation to ΔJ if the comparison curves are far apart. The analogy in calculus is illustrated in Fig. 4-3, where it is seen that df is a good approximation to Δf for small Δt.

As with differentials, we would prefer to avoid using the definition each time the variation of a functional is to be determined; in Section 4.2 we shall develop a formal procedure for finding variations of functionals.

Maxima and Minima of Functionals

Let us now review the definition of an extreme value of a function.

DEFINITION 4-11

A function f with domain $\mathscr{D}$ has a *relative extremum* at the point $\mathbf{q}^*$ if there is an $\epsilon>0$ such that for all points $\mathbf{q}$ in $\mathscr{D}$ that satisfy $\|\mathbf{q}-\mathbf{q}^*\|<\epsilon$ the increment of f has the same sign. If

$$\Delta f=f(\mathbf{q})-f(\mathbf{q}^*)\ge 0, \tag{4.1-45}$$

$f(\mathbf{q}^*)$ is a *relative minimum;* if

$$\Delta f=f(\mathbf{q})-f(\mathbf{q}^*)\le 0, \tag{4.1-46}$$

$f(\mathbf{q}^*)$ is a *relative maximum.*

If (4.1-45) is satisfied for arbitrarily large ϵ, then $f(\mathbf{q}^*)$ is a *global,* or *absolute, minimum.* Similarly, if (4.1-46) holds for arbitrarily large ϵ, then $f(\mathbf{q}^*)$ is a *global,* or *absolute, maximum.*

Recall the procedure for locating extrema of functions. Generally, one attempts to find points where the differential vanishes—a necessary condition for an extremum at an interior point of $\mathscr{D}$. Assuming that there are such points and that they can be determined, then one can examine the behavior of the function in the vicinity of these points.

Example 4.1-11. Consider the function of one variable illustrated in Fig. 4-4. The function is defined for $t\in[t_0,t_f]$. Since the interval is bounded

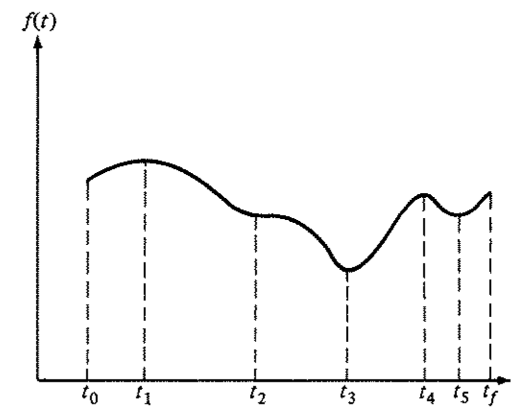


Figure 4-4 A function with several extrema

and closed, candidates for extrema are located at points where the differential vanishes and also at the end points. For this function, the differential vanishes at t_1, t_2, t_3, t_4, and t_5—these are called stationary points. t_2, however, is not an extreme point; it is a horizontal inflection point. t_1 and t_4 are relative maxima, and t_3 and t_5 are relative minima. Examining the function at the end points, we see that t_0 is a relative minimum and t_f is a relative maximum. It is easily shown for a function of one variable that at the left end point

$$\frac{df}{dt} > 0 \quad \text{implies that } t_0 \text{ is a relative minimum,}$$

and

$$\frac{df}{dt} < 0 \quad \text{implies that } t_0 \text{ is a relative maximum.}$$

For the right-hand end point the sense of the inequalities is reversed. Finally, observe that t_1 is the absolute or global maximum point and t_3 is the global minimum.

Next, consider a functional J which is defined for all functions $\mathbf{x}$ in a class Ω.

DEFINITION 4-12

A functional J with domain Ω has a relative extremum at $\mathbf{x}^*$ if there is an $\epsilon > 0$ such that for all functions $\mathbf{x}$ in Ω which satisfy $\|\mathbf{x} - \mathbf{x}^*\| < \epsilon$ the increment of J has the same sign. If

$$\Delta J = J(\mathbf{x}) - J(\mathbf{x}^*) \geq 0, \tag{4.1-47}$$

$J(\mathbf{x}^*)$ is a *relative minimum*; if

$$\Delta J = J(\mathbf{x}) - J(\mathbf{x}^*) \leq 0, \tag{4.1-48}$$

$J(\mathbf{x}^*)$ is a *relative maximum*.

If (4.1-47) is satisfied for arbitrarily large ϵ, then $J(\mathbf{x}^*)$ is a *global*, or *absolute*, *minimum*. Similarly, if (4.1-48) holds for arbitrarily large ϵ, then $J(\mathbf{x}^*)$ is a *global*, or *absolute*, *maximum*. $\mathbf{x}^*$ is called an *extremal*, and $J(\mathbf{x}^*)$ is referred to as an *extremum*.

The Fundamental Theorem of the Calculus of Variations

The fundamental theorem used in finding extreme values of functions is the necessary condition that the differential vanish at an extreme point (except extrema at the boundaries of closed regions). In variational problems, the analogous theorem is that the variation must be zero on an extremal curve, provided that there are no bounds imposed on the curves. We next state this theorem and give the proof.

Let $\mathbf{x}$ be a vector function of t in the class Ω, and $J(\mathbf{x})$ be a differentiable functional of $\mathbf{x}$. *Assume that the functions in Ω are not constrained by any boundaries.*

The fundamental theorem of the calculus of variations is

If $\mathbf{x}^*$ is an extremal, the variation of J must vanish on $\mathbf{x}^*$; that is,

$$\delta J(\mathbf{x}^*, \delta\mathbf{x}) = 0 \text{ for all admissible } \delta\mathbf{x}.† \tag{4.1-49}$$

Proof by contradiction: Assume that $\mathbf{x}^*$ is an extremal and that $\delta J(\mathbf{x}^*, \delta\mathbf{x}) \neq 0$. Let us show that these assumptions imply that the increment ΔJ can be made to change sign in an arbitrarily small neighborhood of $\mathbf{x}^*$.

The increment is

$$\begin{aligned}\Delta J(\mathbf{x}^*, \delta\mathbf{x}) &= J(\mathbf{x}^* + \delta\mathbf{x}) - J(\mathbf{x}^*) \\ &= \delta J(\mathbf{x}^*, \delta\mathbf{x}) + g(\mathbf{x}^*, \delta\mathbf{x}) \cdot \|\delta\mathbf{x}\|,\end{aligned} \tag{4.1-50}$$

where $g(\mathbf{x}^*, \delta\mathbf{x}) \to 0$ as $\|\delta\mathbf{x}\| \to 0$; thus, there is a neighborhood, $\|\delta\mathbf{x}\| < \epsilon$, where $g(\mathbf{x}^*, \delta\mathbf{x}) \cdot \|\delta\mathbf{x}\|$ is small enough so that δJ dominates the expression for ΔJ.

Now let us select the variation

$$\delta\mathbf{x} = \alpha\delta\mathbf{x}^{(1)} \tag{4.1-51}$$

shown in Fig. 4-5 (for a scalar function), where $\alpha > 0$ and $\|\alpha\delta\mathbf{x}^{(1)}\| < \epsilon$. Suppose that

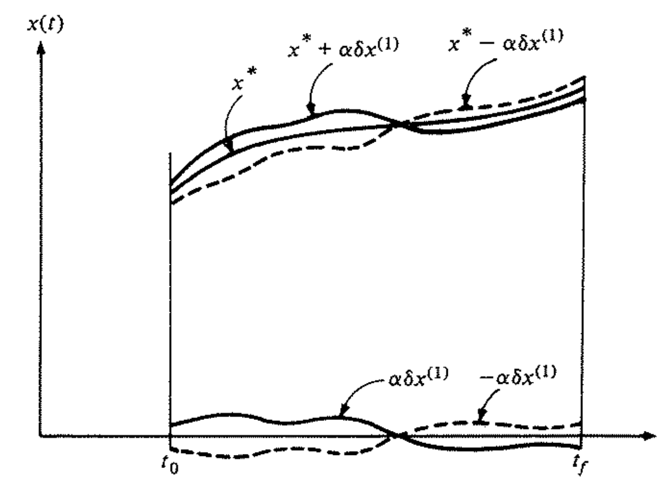


Figure 4-5 An extremal and two neighboring curves

† By admissible $\delta\mathbf{x}$ we mean that $\mathbf{x} + \delta\mathbf{x}$ must be a member of the class Ω; thus, if Ω is the class of continuous functions, $\mathbf{x}$ and $\delta\mathbf{x}$ are required to be continuous.

$$\delta J(\mathbf{x}^*, \alpha\delta\mathbf{x}^{(1)}) < 0. \tag{4.1-52}$$

Since δJ is a linear functional of $\delta\mathbf{x}$, the principle of homogeneity [see Eq. (4.1-8a)] gives

$$\delta J(\mathbf{x}^*, \alpha\delta\mathbf{x}^{(1)}) = \alpha\delta J(\mathbf{x}^*, \delta\mathbf{x}^{(1)}) < 0. \tag{4.1-53}$$

The signs of ΔJ and δJ are the same for $\|\delta\mathbf{x}\| < \epsilon$; thus,

$$\Delta J(\mathbf{x}^*, \alpha\delta\mathbf{x}^{(1)}) < 0. \tag{4.1-54}$$

Next, we consider the variation

$$\delta\mathbf{x} = -\alpha\delta\mathbf{x}^{(1)}$$

shown in Fig. 4-5. Clearly, $\|\alpha\delta\mathbf{x}^{(1)}\| < \epsilon$ implies that $\|-\alpha\delta\mathbf{x}^{(1)}\| < \epsilon$; therefore, the sign of $\Delta J(\mathbf{x}^*, -\alpha\delta\mathbf{x}^{(1)})$ is the same as the sign of $\delta J(\mathbf{x}^*, -\alpha\delta\mathbf{x}^{(1)})$. Again using the principle of homogeneity, we obtain

$$\delta J(\mathbf{x}^*, -\alpha\delta\mathbf{x}^{(1)}) = -\alpha\delta J(\mathbf{x}^*, \delta\mathbf{x}^{(1)}); \tag{4.1-55}$$

therefore, since $\delta J(\mathbf{x}^*, \alpha\delta\mathbf{x}^{(1)}) < 0$, $\delta J(\mathbf{x}^*, -\alpha\delta\mathbf{x}^{(1)}) > 0$, and this implies

$$\Delta J(\mathbf{x}^*, -\alpha\delta\mathbf{x}^{(1)}) > 0. \tag{4.1-56}$$

To recapitulate, we have shown that if $\delta J(\mathbf{x}^*, \delta\mathbf{x}) \neq 0$, then in an arbitrarily small neighborhood of $\mathbf{x}^*$

$$\Delta J(\mathbf{x}^*, \alpha\delta\mathbf{x}^{(1)}) < 0 \tag{4.1-57}$$

and

$$\Delta J(\mathbf{x}^*, -\alpha\delta\mathbf{x}^{(1)}) > 0, \tag{4.1-58}$$

thus contradicting the assumption that $\mathbf{x}^*$ is an extremal (see Definition 4-12). Therefore, if $\mathbf{x}^*$ is an extremal it is necessary that

$$\delta J(\mathbf{x}^*, \delta\mathbf{x}) = 0 \quad \text{for arbitrary } \delta\mathbf{x}. \tag{4.1-59}$$

The assumption that the functions in Ω are not bounded guarantees that $\alpha\delta\mathbf{x}^{(1)}$ and $-\alpha\delta\mathbf{x}^{(1)}$ are both admissible variations.

Summary

In this section important definitions have been given and the fundamental theorem of the calculus of variations has been proved. The analogy between

certain concepts of calculus and the calculus of variations has been exploited. It is helpful to think in terms of the analogies that exist; by doing so, we can appeal to familiar geometric ideas from the calculus. At the same time, we must be careful not to extrapolate results from calculus to the calculus of variations merely by using "intuitive continuation." In the next section we shall apply the fundamental theorem to problems that become progressively more general; eventually, we shall be able to attack the optimal control problem.

4.2 FUNCTIONALS OF A SINGLE FUNCTION

In this section we shall use the fundamental theorem to determine extrema of functionals depending on a single function. To relate our discussion to "the optimal control problem" posed in Chapter 1 we shall think in terms of finding state trajectories that minimize performance measures. In control problems state trajectories are determined by control histories (and initial conditions); however, to simplify the discussion it will be assumed initially that there are no such constraints and that the states can be directly and independently varied. Subsequently, this assumption will be removed.

The Simplest Variational Problem

Problem 1: Let x be a scalar function in the class of functions with continuous first derivatives. It is desired to find the function x^* for which the functional

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt \tag{4.2-1}$$

has a relative extremum. The notation $J(x)$ means that J is a functional of the function x; $g(x(t), \dot{x}(t), t)$, on the other hand, is a function—g assigns a real number to the point $(x(t), \dot{x}(t), t)$. It is assumed that the integrand g has continuous first and second partial derivatives with respect to all of its arguments; t_0 and t_f are fixed, and the end points of the curve are specified as x_0 and x_f.

Curves in the class Ω which also satisfy the end conditions are called admissible. Several admissible curves are shown in Fig. 4-6.

We wish to find the curves (if any exist) that extremize $J(x)$. The search begins by finding the curves that satisfy the fundamental theorem. Let x be any curve in Ω, and determine the variation $\delta J(x, \delta x)$ from the increment

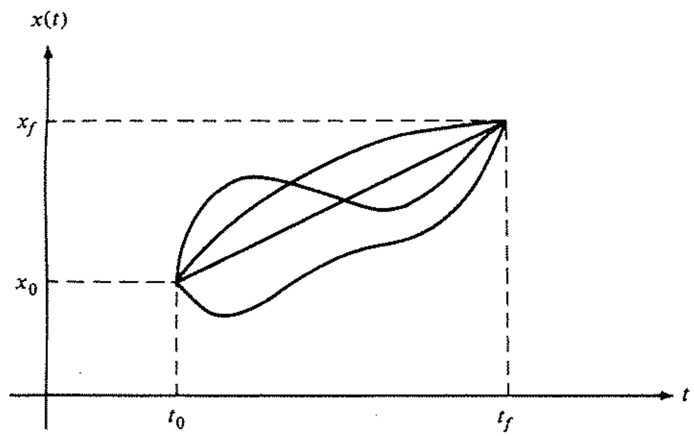


Figure 4-6 Admissible curves for *Problem 1*

$$\begin{aligned}\Delta J(x, \delta x) &= J(x + \delta x) - J(x)\\ &= \int_{t_0}^{t_f} g(x(t) + \delta x(t), \dot{x}(t) + \delta \dot{x}(t), t)\, dt \\ &\quad - \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt.\end{aligned} \tag{4.2-2}$$

Combining the integrals gives

$$\Delta J(x, \delta x) = \int_{t_0}^{t_f} [g(x(t) + \delta x(t), \dot{x}(t) + \delta \dot{x}(t), t) - g(x(t), \dot{x}(t), t)]\, dt. \tag{4.2-3}$$

Notice that the dependence on $\dot{x}$ and $\delta\dot{x}$ is not indicated in the argument of ΔJ, because x and $\dot{x}$, δx and $\delta\dot{x}$ are not independent;

$$\dot{x}(t) = \frac{d}{dt}[x(t)], \qquad \delta\dot{x}(t) = \frac{d}{dt}[\delta x(t)].$$

Eventually, ΔJ will be expressed entirely in terms of x, $\dot{x}$ and δx.

Expanding the integrand of (4.2-3) in a Taylor series about the point $x(t)$, $\dot{x}(t)$ gives

$$\begin{aligned}\Delta J = \int_{t_0}^{t_f} \Big\{ & g(x(t), \dot{x}(t), t) + \left[\frac{\partial g}{\partial x}(x(t), \dot{x}(t), t)\right] \delta x(t) \\ & + \left[\frac{\partial g}{\partial \dot{x}}(x(t), \dot{x}(t), t)\right] \delta \dot{x}(t) \\ & + \frac{1}{2}\bigg[\left[\frac{\partial^2 g}{\partial x^2}(x(t), \dot{x}(t), t)\right] [\delta x(t)]^2\end{aligned} \tag{4.2-4}$$

$$\begin{aligned} & + 2\left[\frac{\partial^2 g}{\partial x \partial \dot{x}}(x(t), \dot{x}(t), t)\right] \delta x(t) \delta \dot{x}(t) \\ & + \left[\frac{\partial^2 g}{\partial \dot{x}^2}(x(t), \dot{x}(t), t)\right] [\delta \dot{x}(t)]^2\bigg] \\ & + o([\delta x(t)]^2, [\delta \dot{x}(t)]^2) - g(x(t), \dot{x}(t), t)\Big\}\, dt.\end{aligned}$$

The notation $o([\delta x(t)]^2, [\delta \dot{x}(t)]^2)$ denotes terms in the expansion of order three and greater in $\delta x(t)$ and $\delta\dot{x}(t)$—these terms are smaller in magnitude than $[\delta x(t)]^2$ and $[\delta \dot{x}(t)]^2$ as $\delta x(t)$ and $\delta \dot{x}(t)$ approach zero. As indicated, the partial derivatives in Eq. (4.2-4) are evaluated on the trajectory x, $\dot{x}$.

Next, we extract the terms in ΔJ that are linear in $\delta x(t)$ and $\delta \dot{x}(t)$ to obtain the variation

$$\begin{aligned}\delta J(x, \delta x) = \int_{t_0}^{t_f} \bigg\{ & \left[\frac{\partial g}{\partial x}(x(t), \dot{x}(t), t)\right] \delta x(t) \\ & + \left[\frac{\partial g}{\partial \dot{x}}(x(t), \dot{x}(t), t)\right] \delta \dot{x}(t)\bigg\}\, dt.\end{aligned} \tag{4.2-5}$$

$\delta x(t)$ and $\delta \dot{x}(t)$ are related by

$$\delta x(t) = \int_{t_0}^{t} \delta \dot{x}(t)\, dt + \delta x(t_0); \tag{4.2-6}$$

thus, selecting δx uniquely determines $\delta \dot{x}$. We shall regard δx as being the function that is varied independently. To express (4.2-5) entirely in terms containing δx, we integrate by parts the term involving $\delta \dot{x}$ to obtain

$$\begin{aligned}\delta J(x, \delta x) = & \left[\frac{\partial g}{\partial \dot{x}}(x(t), \dot{x}(t), t)\right] \delta x(t) \Big|_{t_0}^{t_f} \\ & + \int_{t_0}^{t_f} \bigg\{\left[\frac{\partial g}{\partial x}(x(t), \dot{x}(t), t)\right] \\ & - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x(t), \dot{x}(t), t)\right]\bigg\} \delta x(t)\, dt.\end{aligned} \tag{4.2-7}$$

Since $x(t_0)$ and $x(t_f)$ are specified, all admissible curves must pass through these points; therefore, $\delta x(t_0) = 0$, $\delta x(t_f) = 0$, and the terms outside the integral vanish.

If we now consider an extremal curve, applying the fundamental theorem yields

$$\begin{aligned}\delta J(x^*, \delta x) = 0 = \int_{t_0}^{t_f} \bigg\{ & \frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) \\ & - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\bigg\} \delta x(t)\, dt.\end{aligned} \tag{4.2-8}$$

Thus, the integral must be zero; does this tell us anything about the integrand?

To answer this question, consider the function δx; it has continuous derivatives, and must be zero at t_0 and t_f, but aside from these requirements it is completely arbitrary. The assumptions made regarding the function g guarantee that the term which multiplies $\delta x(t)$ in Eq. (4.2-8) is continuous. It can be shown that if a function h is continuous and

$$\int_{t_0}^{t_f} h(t)\,\delta x(t)\,dt = 0 \qquad (4.2\text{-}9)$$

for every function δx that is continuous in the interval $[t_0, t_f]$, then h must be zero everywhere in the interval $[t_0, t_f]$.

This result, called the *fundamental lemma of the calculus of variations*, is proved in references [E-1] and [G-1]. The essence of the proof is as follows: Suppose that h is not zero everywhere in the interval; then, since h is continuous, there is a neighborhood in $[t_0, t_f]$ in which h has the same sign everywhere. Select δx, which is arbitrary, to be positive (or negative) throughout the neighborhood where h has the same sign, and zero elsewhere. By selecting δx in this manner the integral in Eq. (4.2-9) will be nonzero; thus, h must be identically zero for (4.2-9) to be satisfied.

Figure 4-7 shows a function h that is not identically zero in the interval

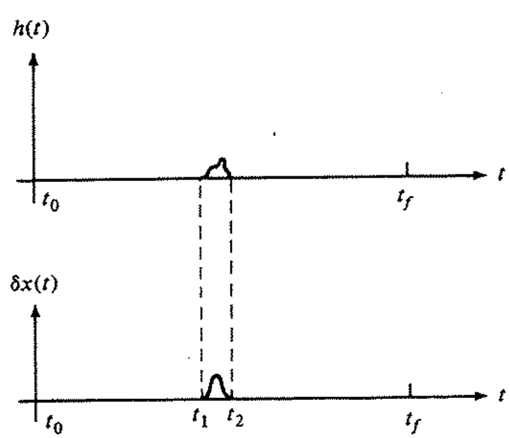


Figure 4-7 A nonzero h and an admissible δx

$[t_0, t_f]$. Selecting δx as shown makes the product $h(t)\,\delta x(t)$ greater than zero in the interval $[t_1, t_2]$, and zero elsewhere. By inspection, the integral of $h(t)\,\delta x(t)$ is certainly not zero. Notice that it does not matter what values h assumes outside of the interval $[t_1, t_2]$.

An intuitive way of looking at this lemma is the following: Given *any*

continuous function h that is not identically zero in the interval $[t_0, t_f]$, a function δx, with continuous derivatives, can be selected which makes the integral $\int_{t_0}^{t_f} h(t)\,\delta x(t)\,dt \neq 0$.

Applying the fundamental lemma to (4.2-8), we find that a necessary condition for x^* to be an extremal is

$$\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right] = 0 \qquad (4.2\text{-}10)$$

for all $t \in [t_0, t_f]$.

Let us now examine Eq. (4.2-10), called the *Euler equation*, in more detail. The presence of d/dt and/or $\dot{x}^*(t)$ means that this is a differential equation.

$$\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]$$

is, in general, a function of $x^*(t)$, $\dot{x}^*(t)$, and t; thus, when this function is differentiated with respect to t, $\ddot{x}^*(t)$ may be present. This means that the differential equation is generally of second order. There may also be terms involving products or powers of $\ddot{x}^*(t)$, $\dot{x}^*(t)$, and $x^*(t)$, in which case the differential equation is nonlinear, and the presence of t in the arguments indicates that the coefficients may be time-varying. Differential equations of this type are normally hard to solve analytically. There are, however, certain special cases (summarized in Appendix 3) in which the Euler equation can be reduced to a first-order differential equation, or solved by evaluating integrals.

In summary then, the Euler equation for *Problem 1* is generally a nonlinear, ordinary, time-varying, hard-to-solve, second-order differential equation.

Since the Euler equation usually cannot be solved analytically, one naturally thinks of using numerical integration. The characteristics of the Euler equation which make analytical solution difficult do not present serious difficulties numerically. Unfortunately, there is another factor that prevents us from simply solving the Euler equation by numerical integration—*the boundary conditions are split*. Instead of having $x(t_0)$ and $\dot{x}(t_0)$ specified [or $x(t_f)$, $\dot{x}(t_f)$], we know $x(t_0)$ and $x(t_f)$. To integrate numerically, we need values for all of the boundary conditions at one end. Thus, we see that to obtain the optimal trajectory x^*, a *nonlinear, two-point boundary-value problem* must be solved. The problem is difficult because of the combination of split boundary values and the nonlinearity of the differential equation. Separately, either of these difficulties can be surmounted without tremendous effort, but together they present a formidable challenge. For the moment we shall consider only problems that can be solved analytically. In Chapter 6 we shall consider some numerical techniques for solving nonlinear, two-point boundary-value problems.

It should be emphasized that since the Euler equation is a necessary condition, further investigation is required to ascertain whether a solution x^* is a minimizing curve, a maximizing curve, or neither.

Example 4.2-1. Find an extremal for the functional

$$J(x) = \int_0^{\pi/2} [\dot{x}^2(t) - x^2(t)]\, dt \tag{4.2-11}$$

which satisfies the boundary conditions $x(0) = 0$ and $x(\pi/2) = 1$.

The Euler equation is

$$\begin{aligned} 0 &= \frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right] \\ &= -2x^*(t) - \frac{d}{dt}[2\dot{x}^*(t)], \end{aligned} \tag{4.2-12}$$

or

$$\ddot{x}^*(t) + x^*(t) = 0. \tag{4.2-13}$$

Since Eq. (4.2-13) is linear and has constant coefficients, it can be readily solved by using classical differential equation theory. Assuming a solution of the form $x^*(t) = k\epsilon^{st}$ and substituting this in (4.2-13), we obtain

$$ks^2\epsilon^{st} + k\epsilon^{st} = 0. \tag{4.2-14}$$

Since (4.2-14) must be satisfied for all t,

$$s^2 + 1 = 0. \tag{4.2-15}$$

The roots of this characteristic equation are $s = \pm j1$,† so the solution has the form

$$x^*(t) = c_1\epsilon^{-jt} + c_2\epsilon^{jt}, \tag{4.2-16}$$

or

$$x^*(t) = c_3 \cos(t) + c_4 \sin(t), \tag{4.2-17}$$

where the c's are constants of integration.

To determine the constants that satisfy the boundary conditions $x(0) = 0$, $x(\pi/2) = 1$, we use the form of the solution in (4.2-17) to obtain

$$0 = c_3 \cos(0) + c_4 \sin(0) \Longrightarrow c_3 = 0† \tag{4.2-18}$$

and

$$1 = \overset{0}{c_3} \cos\left(\frac{\pi}{2}\right) + c_4 \sin\left(\frac{\pi}{2}\right) \Longrightarrow c_4 = 1. \tag{4.2-19}$$

Thus, the solution to the Euler equation is

$$x^*(t) = \sin(t). \tag{4.2-20}$$

The problem, as stated, has been solved, but let us investigate the increment for a neighboring curve to see if x^* is a minimum. As a comparison curve, consider the family

$$\begin{aligned} x(t) &= \sin(t) + \alpha \sin(2t) \\ &= x^*(t) + \delta x(t), \end{aligned} \tag{4.2-21}$$

with α as a real constant. Several curves for various values of α are shown in Fig. 4-8. Observe that each δx curve goes through zero at $t = 0$ and at $t = \pi/2$; thus $x^* + \delta x$ satisfies the required boundary conditions.

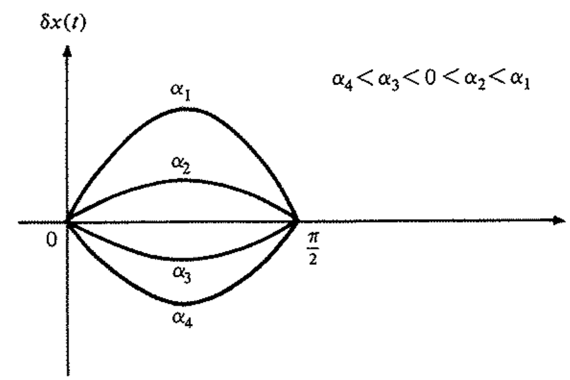


Figure 4-8 Several admissible δx curves

Substituting $x^*(t) = \sin(t)$ and $\dot{x}^*(t) = \cos(t)$ into the integrand of (4.2-11), we find that $J(x^*) = 0$. If $x(t) = \sin(t) + \alpha \sin(2t)$ and $\dot{x}(t) = \cos(t) + 2\alpha \cos(2t)$ are substituted into (4.2-11) and the integration performed, the result is

$$J(x^* + \delta x) = \left[\frac{3\pi}{4}\right]\alpha^2. \tag{4.2-22}$$

† $j \triangleq \sqrt{-1}$.

† $\Longrightarrow$ denotes "implies that."

Since $J(x^* + \delta x) > 0$ for all $\alpha \neq 0$, we conclude that

$$J(x^* + \delta x) > J(x^*) \quad \text{for } \alpha \neq 0. \tag{4.2-23}$$

What does this mean? It certainly indicates that x^* is not a maximizing curve, because we have just constructed a family of neighboring curves that gives larger values of J. Is x^* a minimizing curve? Our evidence is not conclusive, but it looks very much as if x^* does minimize J. We could try other neighboring curves to reinforce our suspicions, or else test x^* to see if it satisfies sufficient conditions for a minimum. Sufficient conditions for minima are beyond the scope of this book, so we shall content ourselves with investigating a few neighboring curves to ascertain whether a curve is maximal, minimal, or neither.

Now let us consider problems having end points that are not fixed. We shall consider only free end conditions at the final time; problems with unspecified boundary conditions at the initial time can be treated in a similar manner.

Final Time Specified, $x(t_f)$ Free

Problem 2: Find a necessary condition for a function to be an extremal for the functional

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt; \tag{4.2-24}$$

t_0, $x(t_0)$, and t_f are specified, and $x(t_f)$ is free. The admissible curves all begin at the same point and terminate on a vertical line, as, for example, is the case in Fig. 4-9. To use the fundamental theorem, we first find the variation as in *Problem 1*. After integrating by parts, we have [see Eq. (4.2-7)]

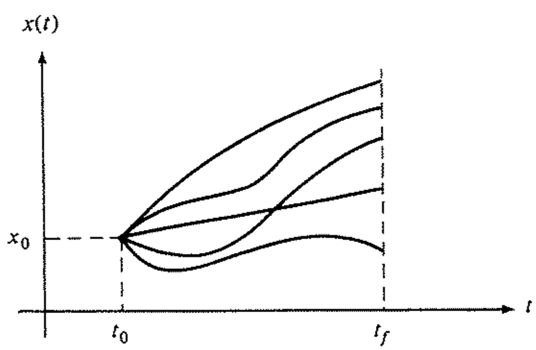


Figure 4-9 Several admissible curves for *Problem 2*

$$\begin{aligned}\delta J(x, \delta x) = \left[\frac{\partial g}{\partial \dot{x}}(x(t), \dot{x}(t), t)\right] \delta x(t)\Big|_{t_0}^{t_f} + \int_{t_0}^{t_f} \Big\{\left[\frac{\partial g}{\partial x}(x(t), \dot{x}(t), t)\right] \\ - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x(t), \dot{x}(t), t)\right]\Big\} \delta x(t)\, dt.\end{aligned} \tag{4.2-25}$$

Now $\delta x(t_0) = 0$ for all admissible curves, but $\delta x(t_f)$ is arbitrary.

For an extremal x^*, we know that $\delta J(x^*, \delta x)$ must be zero. Let us next show that the integral in (4.2-25) must be zero on an extremal. *Suppose that* the curve x^* shown in Fig. 4-10 is an extremal for the free end point problem.

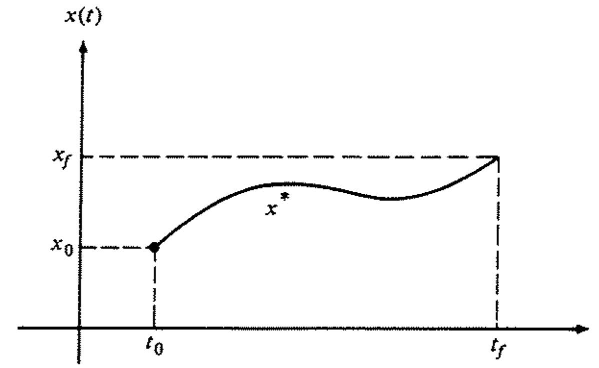


Figure 4-10 An extremal for a free end point problem

The value of $x^*(t_f)$ is x_f. Now consider a fixed end point problem with the same functional, the same initial and final times, and with *specified end points* $x(t_0) = x_0$ and $x(t_f) = x_f$ that are the same as for the extremal x^* in the free end point problem. The curve x^* in Fig. 4-10 must be an extremal for this fixed end point problem; therefore, x^* must be a solution of the Euler equation (4.2-10), and the integral term must be zero on an extremal. In other words, an extremal for a free end point problem is also an extremal for the fixed end point problem with the same end points, and the same functional; thus, *regardless of the boundary conditions, the Euler equation must be satisfied.*

Since

$$\delta J(x^*, \delta x) = 0, \quad \text{and} \quad \frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right] = 0$$

for all $t \in [t_0, t_f]$, from Eq. (4.2-25) we have

$$\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \delta x(t_f) = 0. \tag{4.2-26}$$

But since $x(t_f)$ is free, $\delta x(t_f)$ is arbitrary; therefore, it is necessary that

$$\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f) = 0. \tag{4.2-27}$$

The Euler equation is second order, and Eq. (4.2-27) provides the second required boundary condition [$x(t_0) = x_0$ is the other boundary condition]. We shall call Eq. (4.2-27) the *natural boundary condition;* notice that again we are confronted by a problem with split boundary values.

Example 4.2-2. Determine the smooth curve of smallest length connecting the point $x(0) = 1$ to the line $t = 5$.

It can be shown that the length of a curve lying in the $t - x(t)$ plane, with $t_0 = 0$ and $t_f = 5$, is

$$J(x) = \int_0^5 [1 + \dot{x}^2(t)]^{1/2}\, dt. \tag{4.2-28}$$

The Euler equation

$$-\frac{d}{dt}\left[\frac{\dot{x}^*(t)}{[1 + \dot{x}^{*2}(t)]^{1/2}}\right] = 0 \tag{4.2-29}$$

reduces to

$$\ddot{x}^*(t) = 0, \tag{4.2-30}$$

which has the solution

$$x^*(t) = c_1 t + c_2, \tag{4.2-31}$$

where c_1 and c_2 are constants of integration. $x^*(0) = 1$, so from (4.2-31) we have $c_2 = 1$. From Eq. (4.2-27),

$$\frac{\dot{x}^*(5)}{[1 + \dot{x}^{*2}(5)]^{1/2}} = 0, \tag{4.2-32}$$

which implies that $\dot{x}^*(5) = 0$. Substituting $\dot{x}^*(5) = 0$ into the equation

$$\dot{x}^*(t) = c_1, \tag{4.2-33}$$

obtained by differentiating (4.2-31), gives $c_1 = 0$. The solution then is

$$x^*(t) = 1, \tag{4.2-34}$$

a straight line parallel to the t axis.

Example 4.2-3. Determine an extremal for the functional

$$J(x) = \int_0^2 [\dot{x}^2(t) + 2x(t)\dot{x}(t) + 4x^2(t)]\, dt; \tag{4.2-35}$$

$x(0) = 1$, and $x(2)$ is free.

From (4.2-10) the Euler equation is

$$-\ddot{x}^*(t) + 4x^*(t) = 0. \tag{4.2-36}$$

The solution has the form

$$x^*(t) = c_1\epsilon^{-2t} + c_2\epsilon^{2t}. \tag{4.2-37}$$

To evaluate the constants of integration, use the boundary condition $x(0) = 1$, and the natural boundary condition

$$\frac{\partial g}{\partial \dot{x}}(x^*(2), \dot{x}^*(2)) = 0. \tag{4.2-38}$$

Equation (4.2-38) gives

$$\dot{x}^*(2) + x^*(2) = 0, \tag{4.2-39}$$

and from Eq. (4.2-37) we find that

$$\dot{x}^*(t) = -2c_1\epsilon^{-2t} + 2c_2\epsilon^{2t}. \tag{4.2-40}$$

Evaluating (4.2-37) and (4.2-40) with $t = 2$ and substituting in Eq. (4.2-39) we obtain

$$-c_1\epsilon^{-4} + 3c_2\epsilon^4 = 0. \tag{4.2-41}$$

The boundary value $x(0) = 1$ provides the equation

$$c_1 + c_2 = 1. \tag{4.2-42}$$

Solving these simultaneous algebraic equations for c_1 and c_2 yields

$$c_1 = \frac{3\epsilon^4}{\epsilon^{-4} + 3\epsilon^4}, \quad \text{and} \quad c_2 = \frac{\epsilon^{-4}}{\epsilon^{-4} + 3\epsilon^4}.$$

The final time was fixed in *Problems 1* and *2*; consequently, the variations of the functionals involved two integrals having the same limits of integration. If the final time is free, however, this is no longer the case; therefore, let us now generalize the results of our previous discussion. This is accomplished by separating the total variation of a functional into two partial variations: the variation resulting from the difference $\delta x(t)$ in the interval $[t_0, t_f]$ and the variation resulting from the difference in end points of two curves. The sum of these two variations is called the *general variation* of a functional. First, let us consider the case where $x(t_f)$ is specified.

Final Time Free, $x(t_f)$ Specified

In *Problem 2* we considered the situation where $x(t_f)$ was free, but the final time t_f was specified. Let us now investigate problems in which $x(t_f)$ is specified, but t_f is free.

Problem 3: Find a necessary condition that must be satisfied by an extremal of the functional

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt; \tag{4.2-43}$$

t_0, $x(t_0) = x_0$, and $x(t_f) = x_f$ are specified, and t_f is free.

The admissible curves, several of which are shown in Fig. 4-11, all begin at the point (x_0, t_0) and terminate on the horizontal line with ordinate x_f.

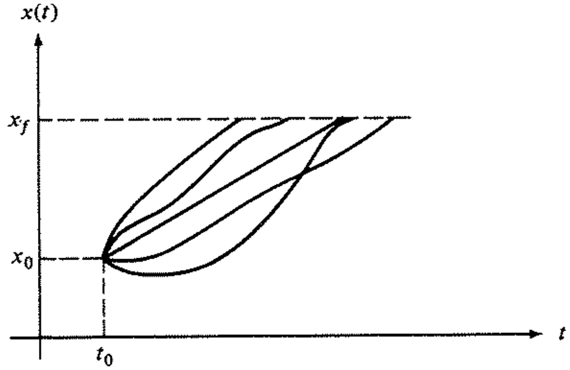


Figure 4-11 Several admissible curves for *Problem 3*

Because of the free final time, the development in *Problem 2* must be modified. In Fig. 4-12 an extremal curve x^*, terminating at the point (x_f, t_f), and a neighboring comparison curve x, terminating at the point $(x_f, t_f + \delta t_f)$, are shown.

From Fig. 4-12 it is apparent that $\delta x(t) = [x(t) - x^*(t)]$ has meaning only in the interval $[t_0, t_f]$, since x^* is not defined for $t \in (t_f, t_f + \delta t_f]$.†

† $t \in (t_f, t_f + \delta t_f]$ means $t_f < t \leq t_f + \delta t_f$.

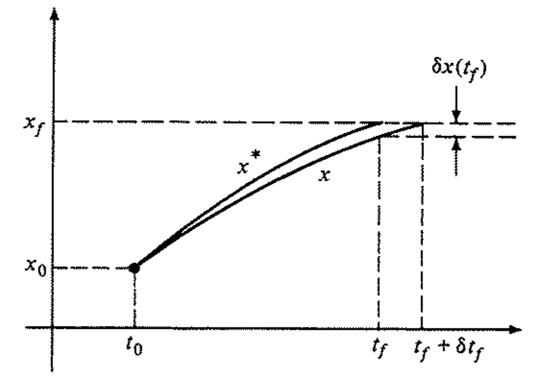


Figure 4-12 An extremal, x^*, and a neighboring comparison curve, x

First, we form the increment

$$\begin{aligned}\Delta J &= \int_{t_0}^{t_f+\delta t_f} g(x(t), \dot{x}(t), t)\, dt - \int_{t_0}^{t_f} g(x^*(t), \dot{x}^*(t), t)\, dt \\ &= \int_{t_0}^{t_f} \{g(x(t), \dot{x}(t), t) - g(x^*(t), \dot{x}^*(t), t)\}\, dt \\ &\quad + \int_{t_f}^{t_f+\delta t_f} g(x(t), \dot{x}(t), t)\, dt,\end{aligned} \tag{4.2-44}$$

or

$$\begin{aligned}\Delta J = \int_{t_0}^{t_f} &\{g(x^*(t) + \delta x(t), \dot{x}^*(t) + \delta \dot{x}(t), t) \\ &- g(x^*(t), \dot{x}^*(t), t)\}\, dt \\ &+ \int_{t_f}^{t_f+\delta t_f} g(x(t), \dot{x}(t), t)\, dt.\end{aligned} \tag{4.2-45}$$

The first integrand can be expanded about $x^*(t)$, $\dot{x}^*(t)$ in a Taylor series to give

$$\begin{aligned}\Delta J = \int_{t_0}^{t_f} &\left\{\left[\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t)\right] \delta x(t) \right. \\ &\left. + \left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right] \delta \dot{x}(t)\right\} dt \\ &+ o(\delta x(t), \delta \dot{x}(t)) + \int_{t_f}^{t_f+\delta t_f} g(x(t), \dot{x}(t), t)\, dt.\dagger\end{aligned} \tag{4.2-46}$$

† $o(\delta x(t), \delta \dot{x}(t))$ denotes terms of higher than first order in $\delta x(t)$ and $\delta \dot{x}(t)$; subsequently we will write simply $o(\cdot)$.

The second integral can be written

$$\int_{t_f}^{t_f+\delta t_f} g(x(t), \dot{x}(t), t)\, dt = [g(x(t_f), \dot{x}(t_f), t_f)]\, \delta t_f + o(\delta t_f). \quad (4.2\text{-}47)$$

Integrating by parts the term in Eq. (4.2-46) containing $\delta\dot{x}(t)$, and substituting (4.2-47), we obtain

$$\begin{aligned} \Delta J = &\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \delta x(t_f) + [g(x(t_f), \dot{x}(t_f), t_f)]\, \delta t_f \\ &+ \int_{t_0}^{t_f} \left\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) \right. \\ &\left. - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\right\} \delta x(t)\, dt + o(\cdot), \end{aligned} \quad (4.2\text{-}48)$$

where we have also used the fact that $\delta x(t_0) = 0$. Next, we shall express $g(x(t_f), \dot{x}(t_f), t_f)$ in terms of $g(x^*(t_f), \dot{x}^*(t_f), t_f)$ by the expansion

$$\begin{aligned} g(x(t_f), \dot{x}(t_f), t_f) = &\ g(x^*(t_f), \dot{x}^*(t_f), t_f) \\ &+ \left[\frac{\partial g}{\partial x}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \delta x(t_f) \\ &+ \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \delta \dot{x}(t_f) + o(\cdot). \end{aligned} \quad (4.2\text{-}49)$$

Substituting this expression in Eq. (4.2-48) yields

$$\begin{aligned} \Delta J = &\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \delta x(t_f) + [g(x^*(t_f), \dot{x}^*(t_f), t_f)]\, \delta t_f \\ &+ \int_{t_0}^{t_f} \left\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) \right. \\ &\left. - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\right\} \delta x(t)\, dt + o(\cdot). \end{aligned} \quad (4.2\text{-}50)$$

$\delta x(t_f)$, which is neither zero nor free, depends on δt_f. The variation of J, δJ, consists of the first-order terms in the increment ΔJ; therefore, the dependence of $\delta x(t_f)$ on δt_f must be linearly approximated. By inspection of Fig. 4-12 we have

$$\delta x(t_f) + \dot{x}^*(t_f)\, \delta t_f \doteq 0\dagger \quad (4.2\text{-}51)$$

or

$$\delta x(t_f) \doteq -\dot{x}^*(t_f)\, \delta t_f. \quad (4.2\text{-}52)$$

† $\doteq$ means "equal to first order."

Substituting (4.2-52) into Eq. (4.2-50), and retaining only first-order terms, we have the variation

$$\begin{aligned} \delta J(x^*, \delta x) = 0 = &\left\{\left[-\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \dot{x}^*(t_f) \right. \\ &\left. + g(x^*(t_f), \dot{x}^*(t_f), t_f)\right\} \delta t_f \\ &+ \int_{t_0}^{t_f} \left\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) \right. \\ &\left. - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\right\} \delta x(t)\, dt. \end{aligned} \quad (4.2\text{-}53)$$

Notice that the integral term represents the partial variation of J caused by $\delta x(t)$, $t \in [t_0, t_f]$, and the term involving δt_f is the partial variation of J caused by the difference in end points; together, these partial variations make up the general (or total) variation.

As in *Problem 2*, we argue that the extremal for this free end point problem is also an extremal for a particular fixed end point problem; therefore, the Euler equation

$$\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right] = 0 \quad (4.2\text{-}54)$$

must be satisfied, and the integral is zero. δt_f is arbitrary, so its coefficient must be zero, and the required boundary condition at t_f is

$$g(x^*(t_f), \dot{x}^*(t_f), t_f) - \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right] \dot{x}^*(t_f) = 0. \quad (4.2\text{-}55)$$

The following example illustrates the procedure for solving a problem with $x(t_f)$ specified and t_f free.

Example 4.2-4. Find an extremal for the functional

$$J(x) = \int_1^{t_f} [2x(t) + \tfrac{1}{2}\dot{x}^2(t)]\, dt; \quad (4.2\text{-}56)$$

the boundary conditions are $x(1) = 4$, $x(t_f) = 4$, and $t_f > 1$ is free.

The Euler equation

$$\ddot{x}^*(t) = 2 \quad (4.2\text{-}57)$$

has the solution

$$x^*(t) = t^2 + c_1 t + c_2. \quad (4.2\text{-}58)$$

t_f is unspecified, so the relationship

$$\begin{aligned} 0 &= g(x^*(t_f), \dot{x}^*(t_f), t_f) - \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right]\dot{x}^*(t_f) \\ &= 2x^*(t_f) - \tfrac{1}{2}\dot{x}^{*2}(t_f) \end{aligned} \tag{4.2-59}$$

must be satisfied. From (4.2-59) and the specified values of $x(1)$ and $x(t_f)$ we obtain

$$x^*(1) = 4 = 1 + c_1 + c_2, \text{ or } c_1 + c_2 = 3 \tag{4.2-60a}$$

$$x^*(t_f) = 4 = t_f^2 + c_1 t_f + c_2 \tag{4.2-60b}$$

$$2x^*(t_f) - \frac{1}{2}\dot{x}^{*2}(t_f) = 0 = 2c_2 - \frac{c_1^2}{2}. \tag{4.2-60c}$$

Solving Eqs. (4.2-60) for c_1, c_2, and t_f gives the extremal

$$x^*(t) = t^2 - 6t + 9, \quad \text{and} \quad t_f = 5. \tag{4.2-61}$$

Problems with Both the Final Time t_f and $x(t_f)$ Free

We are now ready to consider problems having *both* t_f and $x(t_f)$ unspecified. Not surprisingly, we shall find that the necessary conditions of *Problems 2* and *3* are included as special cases.

Problem 4: Find a necessary condition that must be satisfied by an extremal for a functional of the form

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt; \tag{4.2-62}$$

t_0 and $x(t_0) = x_0$ are specified, and t_f and $x(t_f)$ are free.

Figure 4-13 shows an extremal x^* and an admissible comparison curve x.

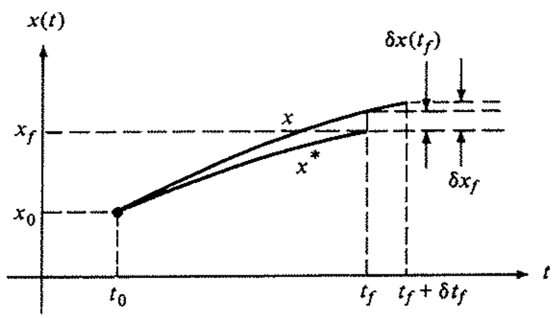


Figure 4-13 An extremal and a neighboring comparison curve for *Problem 4*

Notice that $\delta x(t_f)$ is the difference in ordinates at $t = t_f$ and δx_f is the difference in ordinates of the end points of the two curves. It is important to keep in mind that, in general, $\delta x(t_f) \neq \delta x_f$.

To use the fundamental theorem, we must first determine the variation by forming the increment. This is accomplished in exactly the same manner as in *Problem 3* as far as the equation

$$\begin{aligned} \Delta J = &\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right]\delta x(t_f) + [g(x^*(t_f), \dot{x}^*(t_f), t_f)]\,\delta t_f \\ &+ \int_{t_0}^{t_f}\left\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) \right. \\ &\left. - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\right\}\delta x(t)\, dt + o(\cdot). \end{aligned} \tag{4.2-50}$$

Next, we must relate $\delta x(t_f)$ to δt_f and δx_f. From Fig. 4-13 we have

$$\delta x_f \doteq \delta x(t_f) + \dot{x}^*(t_f)\,\delta t_f, \tag{4.2-63}$$

or

$$\delta x(t_f) = \delta x_f - \dot{x}^*(t_f)\,\delta t_f. \tag{4.2-64}$$

Substituting this in Eq. (4.2-50) and collecting terms, we obtain as the variation

$$\begin{aligned} \delta J(x^*, \delta x) = 0 = &\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right]\delta x_f \\ &+ \Big[g(x^*(t_f), \dot{x}^*(t_f), t_f) \\ &- \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right]\dot{x}^*(t_f)\Big]\delta t_f \\ &+ \int_{t_0}^{t_f}\left\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t) \right. \\ &\left. - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\right\}\delta x(t)\, dt. \end{aligned} \tag{4.2-65}$$

As before, we argue that the Euler equation must be satisfied; therefore, the integral is zero. There may be a variety of end point conditions in practice; however, for the moment we shall consider only two possibilities:

1. t_f and $x(t_f)$ unrelated. In this case δx_f and δt_f are independent of one another and arbitrary, so their coefficients must each be zero. From Eq. (4.2-65), then,

$$\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f) = 0, \tag{4.2-66}$$

and

$$g(x^*(t_f), \dot{x}^*(t_f), t_f) - \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right]\dot{x}^*(t_f) = 0, \quad (4.2\text{-}67)$$

which together imply that

$$g(x^*(t_f), \dot{x}^*(t_f), t_f) = 0. \quad (4.2\text{-}68)$$

2. t_f and $x(t_f)$ related. For example, the final value of x may be constrained to lie on a specified moving point, $\theta(t)$; that is,

$$x(t_f) = \theta(t_f). \quad (4.2\text{-}69)$$

In this case the difference in end points δx_f is related to δt_f by

$$\delta x_f \doteq \frac{d\theta}{dt}(t_f)\,\delta t_f. \quad (4.2\text{-}70)$$

The geometric interpretation of this relationship is shown in Fig. 4-14. The distance a is a linear approximation to δx_f; that is,

$$a = \left[\frac{d\theta}{dt}(t_f)\right]\delta t_f. \quad (4.2\text{-}71)$$

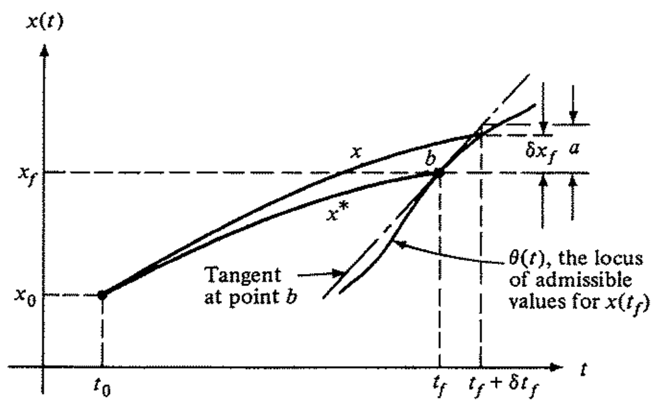


Figure 4-14 $x(t_f)$ and t_f free, but related

Substituting (4.2-70) into Eq. (4.2-65) and collecting terms gives

$$\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_f), \dot{x}^*(t_f), t_f)\right]\left[\frac{d\theta}{dt}(t_f) - \dot{x}^*(t_f)\right] + g(x^*(t_f), \dot{x}^*(t_f), t_f) = 0 \quad (4.2\text{-}72)$$

because δt_f is arbitrary. This equation is called the *transversality condition.*

In either of the two cases considered, integrating the Euler equation gives a solution $x^*(c_1, c_2, t)$, where c_1 and c_2 are constants of integration. c_1, c_2, and the unknown value of t_f can then be determined from $x^*(c_1, c_2, t_0) = x_0$ and Eqs. (4.2-66) and (4.2-68) if $x(t_f)$ and t_f are unrelated, or Eqs. (4.2-69) and (4.2-72) if $x(t_f)$ and t_f are related. Let us illustrate the use of these equations with the following examples.

Example 4.2-5. Find an extremal curve for the functional

$$J(x) = \int_{t_0}^{t_f} [1 + \dot{x}^2(t)]^{1/2}\,dt; \quad (4.2\text{-}73)$$

the boundary conditions $t_0 = 0$, $x(0) = 0$ are specified, t_f and $x(t_f)$ are free, but $x(t_f)$ is required to lie on the line

$$\theta(t) = -5t + 15. \quad (4.2\text{-}74)$$

The functional $J(x)$ is the length of the curve x; thus, the function that minimizes J is the shortest curve from the origin to the specified line. The Euler equation is

$$\frac{d}{dt}\left[\frac{\dot{x}^*(t)}{[1 + \dot{x}^{*2}(t)]^{1/2}}\right] = 0. \quad (4.2\text{-}75)$$

Performing the differentiation with respect to time and simplifying, we obtain

$$\ddot{x}^*(t) = 0, \quad (4.2\text{-}76)$$

which has the solution

$$x^*(t) = c_1 t + c_2. \quad (4.2\text{-}77)$$

We know that $x^*(0) = 0$, so $c_2 = 0$. To evaluate the other constant of integration, we use the transversality condition. From Eq. (4.2-72), since $x(t_f)$ and t_f are related,

$$\frac{\dot{x}^*(t_f)}{[1 + \dot{x}^{*2}(t_f)]^{1/2}} \cdot [-5 - \dot{x}^*(t_f)] + [1 + \dot{x}^{*2}(t_f)]^{1/2} = 0. \quad (4.2\text{-}78)$$

Simplifying, we have

$$-5\dot{x}^*(t_f) + 1 = 0, \quad (4.2\text{-}79)$$

from which, using Eq. (4.2-77), we obtain $c_1 = \frac{1}{5}$. The value of t_f, found from

$$x^*(t_f) = \theta(t_f)$$
$$\tfrac{1}{5}t_f = -5t_f + 15, \quad (4.2\text{-}80)$$

is

$$t_f = \frac{75}{26} = 2.88. \quad (4.2\text{-}81)$$

Thus, the solution is

$$x^*(t) = \tfrac{1}{5}t. \quad (4.2\text{-}82)$$

Figure 4-15 shows what we knew all along: the shortest path is along the perpendicular to the line that passes through the origin.

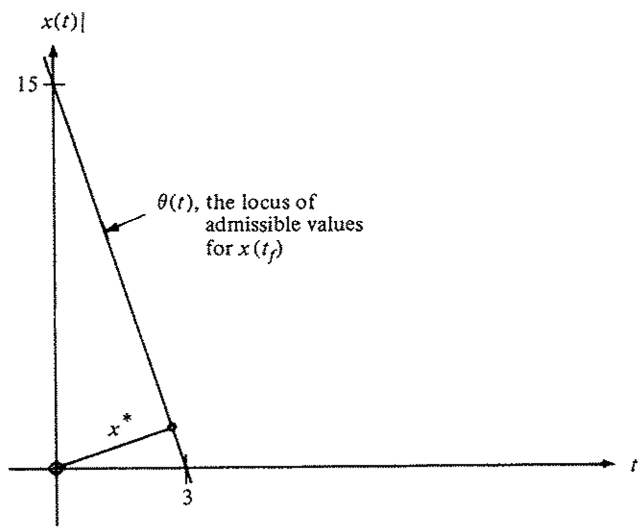


Figure 4-15 The extremal curve for Example 4.2-5

Example 4.2-6. Find an extremal for the functional in Eq. (4.2-73) which begins at the origin and terminates on the curve

$$\theta(t) = \tfrac{1}{2}[t - 5]^2 - \tfrac{1}{2}. \quad (4.2\text{-}83)$$

The Euler equation and its solution are the same as in the previous example, and since $x^*(0) = 0$ we again have $c_2 = 0$. From Eq. (4.2-72) the transversality condition is

$$\frac{\dot{x}^*(t_f)}{[1 + \dot{x}^{*2}(t_f)]^{1/2}} \cdot [t_f - 5 - \dot{x}^*(t_f)] + [1 + \dot{x}^{*2}(t_f)]^{1/2} = 0. \quad (4.2\text{-}84)$$

Simplifying, and substituting $\dot{x}^*(t_f) = c_1$, we obtain

$$c_1[t_f - 5] + 1 = 0. \quad (4.2\text{-}85)$$

Equating $x^*(t_f)$ and $\theta(t_f)$ yields

$$c_1 t_f = \tfrac{1}{2}[t_f - 5]^2 - \tfrac{1}{2}. \quad (4.2\text{-}86)$$

Solving the simultaneous equations (4.2-85) and (4.2-86), we find that $c_1 = \tfrac{1}{2}$ and $t_f = 3$, so the solution is

$$x^*(t) = \tfrac{1}{2}t. \quad (4.2\text{-}87)$$

Summary

We have now progressed from "the simplest variational problem" to problems having rather general boundary conditions. The key equation is (4.2-65), because from it we can deduce all of the results we have obtained so far. We have found that *regardless of the boundary conditions, the Euler equation must be satisfied;* thus, the integral term of (4.2-65) will be zero. If t_f and $x(t_f)$ are specified (*Problem 1*), then $\delta t_f = 0$ and $\delta x_f = \delta x(t_f) = 0$ in Eq. (4.2-65). To obtain the boundary condition equations for *Problem 2* [t_f specified, $x(t_f)$ free], simply let $\delta t_f = 0$ and $\delta x_f = \delta x(t_f)$ in (4.2-65). Similarly, to obtain the equations of *Problem 3*, substitute $\delta x_f = 0$ in Eq. (4.2-65).

Since the equations obtained for *Problems 1* through *3* can be obtained as special cases of Eq. (4.2-65), we suggest that the reader now consider the results of *Problem 4* as the starting point for solving problems of any of the foregoing types.

4.3 FUNCTIONALS INVOLVING SEVERAL INDEPENDENT FUNCTIONS

So far, the functionals considered have contained only a single function and its first derivative. We now wish to generalize our discussion to include functionals that may contain several *independent* functions and their first derivatives. We shall draw heavily on the results of Section 4.2—in fact, our terminal point will be the matrix version of Eq. (4.2-65).

Problems with Fixed End Points

Problem 1a: Consider the functional

$$J(x_1, x_2, \ldots, x_n) = \int_{t_0}^{t_f} g(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\, dt, \quad (4.3\text{-}1)$$

where $x_1, x_2, \ldots, x_n$ are independent functions with continuous first derivatives, and g has continuous first and second partial derivatives with respect to all of its arguments. t_0 and t_f are specified, and the boundary conditions are

$$\begin{aligned} x_1(t_0) &= x_{1_0}; & x_1(t_f) &= x_{1_f}; \\ &\vdots & &\vdots \\ x_n(t_0) &= x_{n_0}; & x_n(t_f) &= x_{n_f}. \end{aligned}$$

We wish to use the fundamental theorem to determine a necessary condition for the functions $x_1^*, x_2^*, \ldots, x_n^*$ to be extremal.

To begin, we find the variation of J by introducing variations in $x_1, \ldots, x_n$, determining the increment, and retaining only the first-order terms:

$$\begin{aligned} \Delta J = \int_{t_0}^{t_f} \{ & g(x_1(t) + \delta x_1(t), \ldots, x_n(t) + \delta x_n(t), \\ & \dot{x}_1(t) + \delta\dot{x}_1(t), \ldots, \dot{x}_n(t) + \delta\dot{x}_n(t), t) \\ & - g(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\}\, dt. \end{aligned} \tag{4.3-2}$$

Expanding in a Taylor series about $x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t)$ gives

$$\begin{aligned} \Delta J = \int_{t_0}^{t_f} \Big\{ & \sum_{i=1}^{n} \Big[\Big[\frac{\partial g}{\partial x_i}(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\Big] \delta x_i(t)\Big] \\ & + \sum_{i=1}^{n} \Big[\Big[\frac{\partial g}{\partial \dot{x}_i}(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\Big] \delta \dot{x}_i(t)\Big] \\ & + \sum_{i=1}^{n} [\text{terms of higher order in } \delta x_i(t), \delta\dot{x}_i(t)]\Big\}\, dt. \end{aligned} \tag{4.3-3}$$

The variation δJ is determined by retaining only the terms that are linear in δx_i and $\delta \dot{x}_i$. To eliminate the dependence of δJ on $\delta\dot{x}_i$, we integrate by parts the terms containing $\delta\dot{x}_i$ to obtain

$$\begin{aligned} \delta J = & \sum_{i=1}^{n} \Big[\Big[\frac{\partial g}{\partial \dot{x}_i}(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\Big] \delta x_i(t)\Big]\Big|_{t_0}^{t_f} \\ & + \int_{t_0}^{t_f} \Big\{ \sum_{i=1}^{n} \Big[\Big[\frac{\partial g}{\partial x_i}(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\Big] \\ & - \frac{d}{dt}\Big[\frac{\partial g}{\partial \dot{x}_i}(x_1(t), \ldots, x_n(t), \dot{x}_1(t), \ldots, \dot{x}_n(t), t)\Big] \delta x_i(t)\Big]\Big\}\, dt. \end{aligned} \tag{4.3-4}$$

Since the boundary conditions for all of the x_i's are fixed at t_0 and t_f, $\delta x_i(t_0) = 0$ and $\delta x_i(t_f) = 0$ $(i = 1, \ldots, n)$, and the terms outside the integral vanish. On an extremal [add*'s to the arguments in (4.3-4)], the variation

must be zero. The δx_i's are independent; let us select all of the δx_i's except δx_1 to be zero. Then

$$\begin{aligned} \delta J = 0 = \int_{t_0}^{t_f} \Big\{ & \frac{\partial g}{\partial x_1}(x_1^*(t), \ldots, x_n^*(t), \dot{x}_1^*(t), \ldots, \dot{x}_n^*(t), t) \\ & - \frac{d}{dt}\Big[\frac{\partial g}{\partial \dot{x}_1}(x_1^*(t), \ldots, x_n^*(t), \dot{x}_1^*(t), \ldots, \dot{x}_n^*(t), t)\Big]\Big\} \delta x_1(t)\, dt. \end{aligned} \tag{4.3-5}$$

But δx_1 can assume arbitrary values as long as it is zero at the end points t_0 and t_f; therefore, the fundamental lemma applies, and the coefficient of $\delta x_1(t)$ must be zero everywhere in the interval $[t_0, t_f]$. Repeating this argument for each of the δx_i's in turn gives

$$\begin{aligned} & \frac{\partial g}{\partial x_i}(x_1^*(t), \ldots, x_n^*(t), \dot{x}_1^*(t), \ldots, \dot{x}_n^*(t), t) \\ & \quad - \frac{d}{dt}\Big[\frac{\partial g}{\partial \dot{x}_i}(x_1^*(t), \ldots, x_n^*(t), \dot{x}_1^*(t), \ldots, \dot{x}_n^*(t), t)\Big] \\ & = 0 \quad \text{for all } t \in [t_0, t_f] \quad \text{and} \quad i = 1, \ldots, n. \end{aligned} \tag{4.3-6}$$

We now have n Euler equations. Notice that the same adjectives apply to these equations as in *Problem 1*; that is, each equation is, in general, a nonlinear, ordinary, hard-to-solve, second-order differential equation with split boundary values. The situation is further complicated by the fact that these differential equations are simultaneous—each differential equation generally contains terms involving all of the functions and their first and second derivatives.

Throughout the preceding development we have painfully (very!) written out each of the arguments. It is much more convenient and compact to use matrix notation; in the future we shall do so. To gain familiarity with the notation, let us re-derive the preceding equations using vector-matrix notation. Starting with the problem statement, we have

$$J(\mathbf{x}) = \int_{t_0}^{t_f} g(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\, dt \tag{4.3-1a}$$

and the boundary conditions $\mathbf{x}(t_0) = \mathbf{x}_0$, $\mathbf{x}(t_f) = \mathbf{x}_f$, where

$$\mathbf{x}(t) \triangleq \begin{bmatrix} x_1(t) \\ \cdot \\ \cdot \\ \cdot \\ x_n(t) \end{bmatrix} \quad \text{and} \quad \dot{\mathbf{x}}(t) \triangleq \begin{bmatrix} \frac{d}{dt}x_1(t) \\ \cdot \\ \cdot \\ \cdot \\ \frac{d}{dt}x_n(t) \end{bmatrix}.$$

The expression for the increment becomes

$$\Delta J = \int_{t_0}^{t_f} \{g(\mathbf{x}(t) + \delta\mathbf{x}(t), \dot{\mathbf{x}}(t) + \delta\dot{\mathbf{x}}(t), t) - g(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\}\, dt, \tag{4.3-2a}$$

which after expansion is

$$\Delta J = \int_{t_0}^{t_f} \left\{\left[\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\right]^T \delta\mathbf{x}(t) + \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\right]^T \delta\dot{\mathbf{x}}(t) + [\text{terms of higher order in } \delta\mathbf{x}(t), \delta\dot{\mathbf{x}}(t)]\right\} dt, \tag{4.3-3a}$$

where

$$\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) \triangleq \left[\frac{\partial g}{\partial x_1}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t), \ldots, \frac{\partial g}{\partial x_n}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\right]^T$$

(an $n \times 1$ column matrix), and similarly for $\partial g/\partial \dot{\mathbf{x}}$. Discarding terms that are nonlinear in $\delta\mathbf{x}(t)$ and $\delta\dot{\mathbf{x}}(t)$ and integrating by parts, we have

$$\begin{aligned}\delta J(\mathbf{x}, \delta\mathbf{x}) = &\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}(t_f), \dot{\mathbf{x}}(t_f), t_f)\right]^T \underbrace{\delta\mathbf{x}(t_f)}_{\mathbf{0}} \\ &- \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}(t_0), \dot{\mathbf{x}}(t_0), t_0)\right]^T \underbrace{\delta\mathbf{x}(t_0)}_{\mathbf{0}} \\ &+ \int_{t_0}^{t_f} \left\{\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\right]\right\}^T \delta\mathbf{x}(t)\, dt.\end{aligned} \tag{4.3-4a}$$

$\mathbf{0}$ is an $n \times 1$ matrix of zeros. Finally, the matrix representation of the Euler equations is

$$\boxed{\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right] = \mathbf{0};} \tag{4.3-6a}$$

Notice that Eq. (4.2-10), obtained previously, is the special case that results when $\mathbf{x}$ is a scalar.

Example 4.3-1. Find the Euler equations for the functional

$$J(\mathbf{x}) = \int_{t_0}^{t_f} [x_1^2(t)x_2(t) + t^2\dot{x}_1^2(t) - \dot{x}_2^2(t)\dot{x}_1(t)]\, dt; \tag{4.3-7}$$

the end points t_0, t_f, $x_1(t_0)$, $x_2(t_0)$, $x_1(t_f)$, and $x_2(t_f)$ are specified.

The Euler equations are given in Eq. (4.3-6a); writing out the indicated derivatives gives

$$\begin{aligned}0 &= \frac{\partial g}{\partial x_1}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}_1}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right] \\ &= 2x_1^*(t)x_2^*(t) - \frac{d}{dt}[2t^2\dot{x}_1^*(t) - \dot{x}_2^{*2}(t)] \\ &= 2x_1^*(t)x_2^*(t) - 4t\dot{x}_1^*(t) - 2t^2\ddot{x}_1^*(t) + 2\dot{x}_2^*(t)\ddot{x}_2^*(t),\end{aligned} \tag{4.3-8}$$

and

$$\begin{aligned}0 &= \frac{\partial g}{\partial x_2}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}_2}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right] \\ &= x_1^{*2}(t) - \frac{d}{dt}[-2\dot{x}_2^*(t)\dot{x}_1^*(t)] \\ &= x_1^{*2}(t) + 2\dot{x}_2^*(t)\ddot{x}_1^*(t) + 2\ddot{x}_2^*(t)\dot{x}_1^*(t).\end{aligned} \tag{4.3-9}$$

These differential equations are nonlinear and have time-varying coefficients.

Example 4.3-2. Find an extremal curve for the functional

$$J(\mathbf{x}) = \int_0^{\pi/4} [x_1^2(t) + 4x_2^2(t) + \dot{x}_1(t)\dot{x}_2(t)]\, dt \tag{4.3-10}$$

which satisfies the boundary conditions

$$\mathbf{x}(0) = \begin{bmatrix}0\\1\end{bmatrix}, \quad\text{and}\quad \mathbf{x}\left(\frac{\pi}{4}\right) = \begin{bmatrix}1\\0\end{bmatrix}.$$

The Euler equations, found from (4.3-6a),

$$2x_1^*(t) - \ddot{x}_2^*(t) = 0 \tag{4.3-11a}$$

$$8x_2^*(t) - \ddot{x}_1^*(t) = 0, \tag{4.3-11b}$$

are linear, time-invariant, and homogeneous. Solving these equations by classical methods (or Laplace transforms) gives

$$x_1^*(t) = c_1\epsilon^{2t} + c_2\epsilon^{-2t} + c_3\cos 2t + c_4\sin 2t, \tag{4.3-12}$$

where c_1, c_2, c_3, and c_4 are constants of integration. Differentiating $x_1^*(t)$ twice and substituting into Eq. (4.3-11b) gives

$$x_2^*(t) = \tfrac{1}{2}c_1\epsilon^{2t} + \tfrac{1}{2}c_2\epsilon^{-2t} - \tfrac{1}{2}c_3\cos 2t - \tfrac{1}{2}c_4\sin 2t. \tag{4.3-13}$$

Putting $t = 0$ and $t = \pi/4$ in (4.3-12) and (4.3-13), we obtain four equations and four unknowns; that is,

$$x_1^*(0) = 0; \quad x_2^*(0) = 1; \quad x_1^*\left(\frac{\pi}{4}\right) = 1; \quad x_2^*\left(\frac{\pi}{4}\right) = 0.$$

Solving these equations for the constants of integration yields

$$c_1 = \frac{-\frac{1}{2} + \epsilon^{-\pi/2}}{\epsilon^{-\pi/2} - \epsilon^{\pi/2}}; \quad c_2 = \frac{\frac{1}{2} - \epsilon^{\pi/2}}{\epsilon^{-\pi/2} - \epsilon^{\pi/2}}; \quad c_3 = -1; \quad c_4 = \frac{1}{2}.$$

Problems with Free End Points

Problem 4a: Consider the functional

$$J(\mathbf{x}) = \int_{t_0}^{t_f} g(\mathbf{x}(t), \dot{\mathbf{x}}(t), t)\, dt, \tag{4.3-14}$$

where $\mathbf{x}$ and g satisfy the continuity and differentiability requirements of *Problem 1a.* $\mathbf{x}(t_0)$ and t_0 are specified; $\mathbf{x}(t_f)$ and t_f are free. Find a necessary condition that must be satisfied by an extremal.

To obtain the generalized variation, we proceed in exactly the same manner as in *Problem 4* of Section 4.2. The only change is that now we are dealing with vector functions. Forming the increment, integrating by parts the term involving $\delta\dot{\mathbf{x}}(t)$, retaining terms of first order, and relating $\delta\mathbf{x}(t_f)$ to $\delta\mathbf{x}_f$ and δt_f [see Fig. 4-13 and Eq. (4.2-64)] by

$$\delta\mathbf{x}(t_f) = \delta\mathbf{x}_f - \dot{\mathbf{x}}^*(t_f)\, \delta t_f, \tag{4.3-15}$$

we obtain for the variation

$$\begin{aligned} \delta J(\mathbf{x}^*, \delta\mathbf{x}) = 0 = &\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \delta\mathbf{x}_f \\ &+ \left[g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right. \\ &- \left.\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \dot{\mathbf{x}}^*(t_f)\right]\delta t_f \\ &+ \int_{t_0}^{t_f} \left\{\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right. \\ &- \left.\frac{d}{dt}\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right]\right\}^T \delta\mathbf{x}(t)\, dt. \end{aligned} \tag{4.3-16}$$

As before, we argue that an extremal for this free end point problem must also be an extremal for a certain fixed end point problem; therefore, $\mathbf{x}^*$ must be a solution of the Euler equations

$$\boxed{\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right] = \mathbf{0}.} \tag{4.3-17}$$

The boundary conditions at the final time are then specified by the relationship

$$\boxed{\begin{aligned} \delta J(\mathbf{x}^*, \delta\mathbf{x}) = 0 = &\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \delta\mathbf{x}_f \\ &+ \left[g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right. \\ &- \left.\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \dot{\mathbf{x}}^*(t_f)\right]\delta t_f. \end{aligned}} \tag{4.3-18}$$

Equations (4.3-17) and (4.3-18) are the key equations, because they summarize necessary conditions that must be satisfied by an extremal curve. The boundary condition equations are obtained by making the appropriate substitutions in Eq. (4.3-18). The equations obtained by making these substitutions, which are contained in Table 4-1, are simply the vector analogs of the equations derived in Section 4.2. Notice that regardless of the problem specifications the boundary conditions are always split; thus, to find an optimal trajectory, in general, a nonlinear, two-point boundary-value problem must be solved.

Situations not included in Table 4-1 may arise; however, these can be handled by returning to Eq. (4.3-18). For example, suppose that t_f is fixed, $x_i(t_f)$, $i = 1, 2, \ldots, r$ are specified, and $x_j(t_f), j = r+1, \ldots, n$ are free. In this case, the appropriate substitutions are

$$\begin{aligned} &\delta t_f = 0; \\ &\delta x_i(t_f) = 0, \qquad i = 1, 2, \ldots, r; \\ &\delta x_j(t_f) \text{ arbitrary}, \qquad j = r+1, \ldots, n. \end{aligned}$$

Let us now consider several examples that illustrate the use of Table 4-1 and the key equations (4.3-17) and (4.3-18).

Example 4.3-3. Find an extremal for the functional

$$J(\mathbf{x}) = \int_0^{\pi/4} [x_1^2(t) + \dot{x}_1(t)\dot{x}_2(t) + \dot{x}_2^2(t)]\, dt. \tag{4.3-19}$$

The functions x_1 and x_2 are independent, and the boundary conditions are

$$\begin{aligned} &x_1(0) = 1; \qquad x_1\left(\frac{\pi}{4}\right) = 2; \\ &x_2(0) = \frac{3}{2}; \qquad x_2\left(\frac{\pi}{4}\right) \text{ free.} \end{aligned}$$

The Euler equations are, from Eq. (4.3-17),

$$2x_1^*(t) - \ddot{x}_2^*(t) = 0; \tag{4.3-20a}$$

$$-\ddot{x}_1^*(t) - 2\ddot{x}_2^*(t) = 0. \tag{4.3-20b}$$

Multiplying Eq. (4.3-20a) by 2 and subtracting (4.3-20b), we obtain

$$\ddot{x}_1^*(t) + 4x_1^*(t) = 0, \tag{4.3-21}$$

which has the solution

$$x_1^*(t) = c_1 \cos 2t + c_2 \sin 2t; \tag{4.3-22}$$

therefore,

$$\ddot{x}_2^*(t) = 2c_1 \cos 2t + 2c_2 \sin 2t. \tag{4.3-23}$$

Integrating twice yields

$$x_2^*(t) = -\frac{c_1}{2}\cos 2t - \frac{c_2}{2}\sin 2t + c_3 t + c_4. \tag{4.3-24}$$

Notice that the boundary conditions are such that this problem does not fit into any of the categories of Table 4-1, so we return to Eq. (4.3-18). $x_1(t_f)$ is specified, which means that $\delta x_{1_f} = \delta x_1(t_f) = 0$. $x_2(t_f)$, however, is free, so $\delta x_2(t_f)$ is arbitrary. We also have that $\delta t_f = 0$ because t_f is specified. Making these substitutions in Eq. (4.3-18) gives

$$\left[\frac{\partial g}{\partial \dot{x}_2}\left(\mathbf{x}^*\left(\frac{\pi}{4}\right), \dot{\mathbf{x}}^*\left(\frac{\pi}{4}\right)\right)\right]\delta x_2(t_f) = 0, \tag{4.3-25}$$

which implies [since $\delta x_2(t_f)$ is arbitrary] that

$$\frac{\partial g}{\partial \dot{x}_2}\left(\mathbf{x}^*\left(\frac{\pi}{4}\right), \dot{\mathbf{x}}^*\left(\frac{\pi}{4}\right)\right) = 0. \tag{4.3-26}$$

But

$$\frac{\partial g}{\partial \dot{x}_2}\left(\mathbf{x}^*\left(\frac{\pi}{4}\right), \dot{\mathbf{x}}^*\left(\frac{\pi}{4}\right)\right) = \dot{x}_1^*\left(\frac{\pi}{4}\right) + 2\dot{x}_2^*\left(\frac{\pi}{4}\right) = 2\cdot c_3,$$

so $c_3 = 0$. From the specified boundary conditions we have

$$x_1(0) = 1 = c_1\cdot 1 + c_2\cdot 0: \qquad c_1 = 1;$$

$$x_2(0) = \frac{3}{2} = -\frac{c_1}{2}\cdot 1 - \frac{c_2}{2}\cdot 0 + c_3\cdot 0 + c_4: \quad c_4 = 1.5 + \frac{c_1}{2} = 2;$$

$$x_1\left(\frac{\pi}{4}\right) = c_1\cdot 0 + c_2\cdot 1 = 2: \qquad c_2 = 2.$$

Table 4-1 DETERMINATION OF BOUNDARY-VALUE RELATIONSHIPS

Problem description	*Substitution*	*Boundary conditions*	*Remarks*
1. $\mathbf{x}(t_f)$, t_f both specified (*Problem 1*)	$\delta\mathbf{x}_f = \delta\mathbf{x}(t_f) = \mathbf{0}$ $\delta t_f = 0$	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\mathbf{x}^*(t_f) = \mathbf{x}_f$	$2n$ equations to determine $2n$ constants of integration
2. $\mathbf{x}(t_f)$ free; t_f specified (*Problem 2*)	$\delta\mathbf{x}_f = \delta\mathbf{x}(t_f)$ $\delta t_f = 0$	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) = \mathbf{0}$	$2n$ equations to determine $2n$ constants of integration
3. t_f free; $\mathbf{x}(t_f)$ specified (*Problem 3*)	$\delta\mathbf{x}_f = \mathbf{0}$	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\mathbf{x}^*(t_f) = \mathbf{x}_f$ $g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)$ $- \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \dot{\mathbf{x}}^*(t_f) = 0$	$(2n + 1)$ equations to determine $2n$ constants of integration and t_f
4. t_f, $\mathbf{x}(t_f)$ free and independent (*Problem 4*)	—	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) = \mathbf{0}$ $g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) = 0$	$(2n + 1)$ equations to determine $2n$ constants of integration and t_f
5. t_f, $\mathbf{x}(t_f)$ free but related by $\mathbf{x}(t_f) = \boldsymbol{\theta}(t_f)$ (*Problem 4*)	$\delta\mathbf{x}_f = \frac{d\boldsymbol{\theta}}{dt}(t_f)\,\delta t_f$†	$\mathbf{x}^*(t_0) = \mathbf{x}_0$ $\mathbf{x}^*(t_f) = \boldsymbol{\theta}(t_f)$ $g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)$ $+ \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \left[\frac{d\boldsymbol{\theta}}{dt}(t_f) - \dot{\mathbf{x}}^*(t_f)\right] = 0$†	$(2n + 1)$ equations to determine $2n$ constants of integration and t_f

† $\frac{d\boldsymbol{\theta}}{dt}$ denotes the $n \times 1$ column vector $\left[\frac{d\theta_1}{dt} \;\; \frac{d\theta_2}{dt} \;\; \cdots \;\; \frac{d\theta_n}{dt}\right]^T$.

The extremal curve is, then,

$$x_1^*(t) = \cos 2t + 2\sin 2t$$
$$x_2^*(t) = -\tfrac{1}{2}\cos 2t - \sin 2t + 2. \tag{4.3-27}$$

Example 4.3-4. Find the Euler equations for the functional

$$J(\mathbf{x}) = \int_0^{t_f} [x_1^3(t) + x_1^2(t) + 2\dot{x}_1(t)\,\dot{x}_2(t) + x_1(t)\,x_2^2(t)]\,dt, \tag{4.3-28}$$

and determine the relationships required to evaluate the constants of integration. The specified boundary conditions are

$$\mathbf{x}(0) = \begin{bmatrix} 2 \\ 1 \end{bmatrix}, \qquad \mathbf{x}(t_f) = \begin{bmatrix} -1 \\ 4 \end{bmatrix},$$

and t_f is free. The functions x_1 and x_2 are independent.

From Eq. (4.3-17) the Euler equations are

$$3x_1^{*2}(t) + 2x_1^*(t) + x_2^{*2}(t) - 2\ddot{x}_2^*(t) = 0$$
$$2x_1^*(t)x_2^*(t) - 2\ddot{x}_1^*(t) = 0. \tag{4.3-29}$$

The solution of these two nonlinear second-order differential equations, $\mathbf{x}^*(c_1, c_2, c_3, c_4, t)$, will contain the four constants of integration, c_1, c_2, c_3, c_4. From the specified boundary conditions we have

$$x_1^*(c_1, c_2, c_3, c_4, 0) = 2$$
$$x_2^*(c_1, c_2, c_3, c_4, 0) = 1$$
$$x_1^*(c_1, c_2, c_3, c_4, t_f) = -1$$
$$x_2^*(c_1, c_2, c_3, c_4, t_f) = 4, \tag{4.3-30}$$

but since t_f is unspecified, there are five unknowns. The other relationship that must be satisfied is obtained from Eq. (4.3-18) with $\delta\mathbf{x}_f = \mathbf{0}$:

$$x_1^{*3}(t_f) + x_1^{*2}(t_f) - 2\dot{x}_1^*(t_f)\dot{x}_2^*(t_f) + x_1^*(t_f)x_2^{*2}(t_f) = 0. \tag{4.3-31}$$

Thus, to determine c_1, c_2, c_3, c_4, and t_f the (nonlinear) algebraic equations (4.3-30) and (4.3-31) would have to be solved.

Example 4.3-5. Find the equation of the curve that is an extremal for the functional

$$J(x) = \int_0^{t_f} [t\dot{x}(t) + \dot{x}^2(t)]\,dt \qquad (t_f > 0) \tag{4.3-32}$$

for the boundary conditions specified below.

From (4.3-17) the Euler equation is

$$-\frac{d}{dt}[t + 2\dot{x}^*(t)] = 0, \tag{4.3-33}$$

or

$$1 + 2\ddot{x}^*(t) = 0. \tag{4.3-34}$$

The solution of this equation is

$$x^*(t) = -\tfrac{1}{4}t^2 + c_1 t + c_2. \tag{4.3-35}$$

(a) What is the extremal if the boundary conditions are $t_f = 1$, $x(0) = 1$, $x(1) = 2.75$?

$$x^*(0) = 1 = c_2$$
$$x^*(1) = 2.75 = -0.25 + c_1 + c_2, \quad \text{and} \quad c_1 = 2, \tag{4.3-36}$$

so

$$x^*(t) = -\tfrac{1}{4}t^2 + 2t + 1. \tag{4.3-37}$$

(b) Find the extremal curve if $x(0) = 1$, $t_f = 2$, and $x(2)$ is free.

Again we have

$$x^*(0) = 1, \quad \text{so} \quad c_2 = 1.$$

From entry 2 of Table 4-1,

$$t_f + 2\dot{x}^*(t_f) = 0$$
$$2 + 2[-\tfrac{1}{2}(2) + c_1] = 0; \tag{4.3-38}$$

therefore, $c_1 = 0$, so

$$x^*(t) = -\tfrac{1}{4}t^2 + 1. \tag{4.3-39}$$

(c) Find the extremal curve if $x(0) = 1$, $x(t_f) = 5$, and t_f is free.

As before, $x^*(0) = 1$ implies that $c_2 = 1$. From entry 3 of Table 4-1

$$t_f[\dot{x}^*(t_f)] + \dot{x}^{*2}(t_f) - [t_f + 2\dot{x}^*(t_f)]\dot{x}^*(t_f) = 0 \tag{4.3-40}$$

or

$$[t_f + \dot{x}^*(t_f) - t_f - 2\dot{x}^*(t_f)]\dot{x}^*(t_f) = 0, \tag{4.3-41}$$

which implies that $\dot{x}^*(t_f) = 0$, so

$$-\tfrac{1}{2}t_f + c_1 = 0, \tag{4.3-42}$$

and

$$5 = -\tfrac{1}{4}t_f^2 + c_1 t_f + 1, \tag{4.3-43}$$

since $c_2 = 1$. Solving these equations simultaneously gives $t_f = 4$ and $c_1 = 2$; therefore,

$$x^*(t) = -\tfrac{1}{4}t^2 + 2t + 1. \tag{4.3-44}$$

Summary

In Sections 4.2 and 4.3 we have progressed from the very restricted problem of a functional of one function with fixed end points to a rather general problem in which there can be several (independent) functions and free end points. Equations (4.3-17) and (4.3-18) are the important equations, because from them we can obtain the necessary conditions derived for more restricted problems.

To recapitulate, we have found that:

1. *Regardless of the boundary conditions*, the Euler equations

$$\frac{\partial g}{\partial \mathbf{x}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t), \dot{\mathbf{x}}^*(t), t)\right] = \mathbf{0} \tag{4.3-17}$$

must be satisfied.

2. The required boundary condition equations are found from the equation

$$\begin{aligned}&\left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \delta \mathbf{x}_f + \Big[g(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f) \\ &\quad - \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_f), \dot{\mathbf{x}}^*(t_f), t_f)\right]^T \dot{\mathbf{x}}^*(t_f)\Big]\delta t_f = 0\end{aligned} \tag{4.3-18}$$

by making the appropriate substitutions for $\delta \mathbf{x}_f$ and δt_f.

4.4 PIECEWISE-SMOOTH EXTREMALS

In the preceding sections we have derived necessary conditions that must be satisfied by extremal curves. The admissible curves were assumed to be continuous and to have continuous first derivatives; that is, the admissible curves were *smooth*. This is a very restrictive requirement for many practical problems. For example, if a control signal is the output of a relay, we know that this signal will contain discontinuities and that when such a control discontinuity occurs, one or more of the components of $\dot{\mathbf{x}}(t)$ will be discontinuous. Thus, we wish to enlarge the class of admissible curves to include

functions that have only *piecewise-continuous* first derivatives; that is, $\dot{\mathbf{x}}$ will be continuous except at a finite number of times in the interval (t_0, t_f).† At a time when $\dot{\mathbf{x}}$ is discontinuous, $\mathbf{x}$ is said to have a *corner*. Let us begin by considering functionals involving only a single function.

The problem is to find a necessary condition that must be satisfied by extrema of the functional

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\, dt. \tag{4.4-1}$$

It is assumed that g has continuous first and second partial derivatives with respect to all of its arguments, and that t_0, t_f, $x(t_0)$, and $x(t_f)$ are specified. $\dot{x}$ is a piecewise-continuous function (or we say that x is a *piecewise-smooth* curve). Assume that $\dot{x}$ has a discontinuity at some point $t_1 \in (t_0, t_f)$; t_1 is not fixed, nor is it usually known in advance.

Let us first express the functional J as

$$\begin{aligned}J(x) &= \int_{t_0}^{t_1} g(x(t), \dot{x}(t), t)\, dt + \int_{t_1}^{t_f} g(x(t), \dot{x}(t), t)\, dt \\ &\triangleq J_1(x) + J_2(x).\end{aligned} \tag{4.4-2}$$

We assert that if x^* is a minimizing extremal for J, then $x^*(t)$, $t \in [t_0, t_1]$, is an extremal for J_1 and $x^*(t)$, $t \in [t_1, t_f]$, is an extremal for J_2. To show this, assume that the final segment of an extremal for J is known; that is, we know $x^*(t)$, $t \in [t_1, t_f]$. Then to minimize J, we seek a curve defined in the interval $[t_0, t_1]$ which minimizes J_1; this curve is, by definition, an extremal of J_1. Similarly, if $x^*(t)$, $t \in [t_0, t_1]$, is known, to minimize J we seek a curve that minimizes J_2—an extremal for J_2.

Figure 4-16 shows an extremal curve x^* and a neighboring comparison

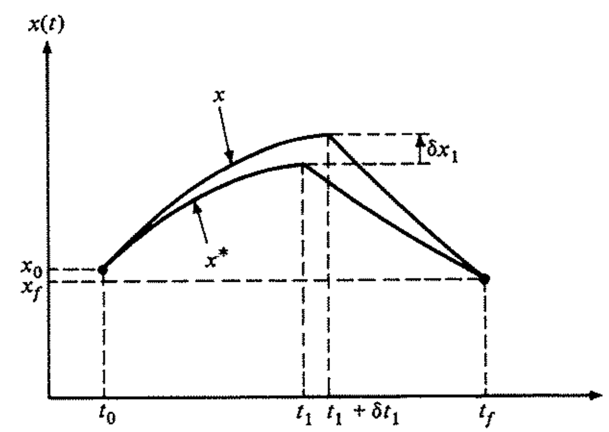


Figure 4-16 A piecewise-smooth extremal and a neighboring comparison curve

† The notation $t \in (t_0, t_f)$ means $t_0 < t < t_f$.

curve x. δt_1 and δx_1 are free, and from the fundamental theorem we know it is necessary that $\delta J(x^*, \delta x) = 0$. Since the coordinates of the corner point are free, we can use the results of *Problem 4* in Section 4.2 [see Eq. (4.2-65)] to obtain

$$\begin{aligned}\delta J(x^*, \delta x) = 0 = &\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1^-), \dot{x}^*(t_1^-), t_1^-)\right]\delta x_1 + \Big\{g(x^*(t_1^-), \dot{x}^*(t_1^-), t_1^-)\\ &- \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1^-), \dot{x}^*(t_1^-), t_1^-)\right]\dot{x}^*(t_1^-)\Big\}\delta t_1\\ &+ \int_{t_0}^{t_1}\Big\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t)\\ &- \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\Big\}\delta x(t)\,dt\\ &- \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1^+), \dot{x}^*(t_1^+), t_1^+)\right]\delta x_1\\ &- \Big\{g(x^*(t_1^+), \dot{x}^*(t_1^+), t_1^+)\\ &- \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1^+), \dot{x}^*(t_1^+), t_1^+)\right]\dot{x}^*(t_1^+)\Big\}\delta t_1\\ &+ \int_{t_1}^{t_f}\Big\{\frac{\partial g}{\partial x}(x^*(t), \dot{x}^*(t), t)\\ &- \frac{d}{dt}\left[\frac{\partial g}{\partial \dot{x}}(x^*(t), \dot{x}^*(t), t)\right]\Big\}\delta x(t)\,dt.\end{aligned} \tag{4.4-3}$$

δx_1 is the difference $x(t_1 + \delta t_1) - x^*(t_1)$, and t_1^- and t_1^+ denote the times just before and just after the discontinuity of $\dot{x}^*$. The terms that multiply δt_1 and δx_1 are due to the presence of t_1 as the upper limit of the first integral and as the lower limit of the second integral. We have shown that x^* is an extremal in both of the intervals $[t_0, t_1]$, and $[t_1, t_f]$; thus the Euler equation must be satisfied, and the integral terms are zero. In order that $\delta J(x^*, \delta x)$ be zero, it is then necessary that

$$\begin{aligned}&\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^-), t_1) - \frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^+), t_1)\right]\delta x_1\\ &+ \Big\{g(x^*(t_1), \dot{x}^*(t_1^-), t_1) - \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^-), t_1)\right]\dot{x}^*(t_1^-)\\ &- g(x^*(t_1), \dot{x}^*(t_1^+), t_1) + \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^+), t_1)\right]\dot{x}^*(t_1^+)\Big\}\delta t_1 = 0.\dagger\end{aligned} \tag{4.4-4}$$

If t_1 and $x(t_1)$ are unrelated, δx_1 and δt_1 are independently arbitrary, so their coefficients must each be zero and we have

† Notice that we have retained the t_1^+, t_1^- notation only where the distinction needs to be made.

$$\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^-), t_1) = \frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^+), t_1), \tag{4.4-5a}$$

and

$$\begin{aligned}&g(x^*(t_1), \dot{x}^*(t_1^-), t_1) - \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^-), t_1)\right]\dot{x}^*(t_1^-)\\ &= g(x^*(t_1), \dot{x}^*(t_1^+), t_1) - \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^+), t_1)\right]\dot{x}^*(t_1^+).\end{aligned} \tag{4.4-5b}$$

These two equations, called the *Weierstrass-Erdmann corner conditions*, are necessary conditions for an extremal. If there are several times $t_1, t_2, \ldots, t_r$ when corners exist, then at each such time these corner conditions must be satisfied.

It may be that $x(t_1)$ and t_1 are related by $x(t_1) = \theta(t_1)$. If so, δx_1 and δt_1 in Eq. (4.4-4) are not independently arbitrary; they are related by

$$\delta x_1 \doteq \frac{d\theta}{dt}(t_1)\,\delta t_1.\dagger \tag{4.4-6}$$

Substituting (4.4-6) into (4.4-4) and equating the coefficient of δt_1 equal to zero (since δt_1 is arbitrary), we obtain

$$\begin{aligned}&\left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^-), t_1)\right]\left[\frac{d\theta}{dt}(t_1) - \dot{x}^*(t_1^-)\right] + g(x^*(t_1), \dot{x}^*(t_1^-), t_1)\\ &= \left[\frac{\partial g}{\partial \dot{x}}(x^*(t_1), \dot{x}^*(t_1^+), t_1)\right]\left[\frac{d\theta}{dt}(t_1) - \dot{x}^*(t_1^+)\right] + g(x^*(t_1), \dot{x}^*(t_1^+), t_1).\end{aligned} \tag{4.4-7}$$

The extension of the Weierstrass-Erdmann corner conditions to the case where J involves several functions is straightforward. The reader can show that

$$\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_1), \dot{\mathbf{x}}^*(t_1^-), t_1) = \frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_1), \dot{\mathbf{x}}^*(t_1^+), t_1), \tag{4.4-8a}$$

and

$$\begin{aligned}&g(\mathbf{x}^*(t_1), \dot{\mathbf{x}}^*(t_1^-), t_1) - \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_1), \dot{\mathbf{x}}^*(t_1^-), t_1)\right]^T\dot{\mathbf{x}}^*(t_1^-)\\ &= g(\mathbf{x}^*(t_1), \dot{\mathbf{x}}^*(t_1^+), t_1) - \left[\frac{\partial g}{\partial \dot{\mathbf{x}}}(\mathbf{x}^*(t_1), \dot{\mathbf{x}}^*(t_1^+), t_1)\right]^T\dot{\mathbf{x}}^*(t_1^+)\end{aligned} \tag{4.4-8b}$$

are the appropriate equations when $\mathbf{x}$ represents n independent functions and $\mathbf{x}(t_1)$ and t_1 are not constrained by any relationship.

† For a geometric interpretation of this relationship, refer to *Problem 4*, Fig. 4-14.

To illustrate the role of the corner conditions, let us consider the following examples.

Example 4.4-1. Find a piecewise-smooth curve that begins at the point $x(0) = 0$, ends at the point $x(2) = 1$, and minimizes the functional

$$J(x) = \int_0^2 \dot{x}^2(t)[1 - \dot{x}(t)]^2\, dt. \tag{4.4-9}$$

The integrand g depends only on $\dot{x}(t)$; therefore, the solution of the Euler equation is (see Appendix 3, Case 1)

$$\dot{x}^*(t) = c_1 t + c_2. \tag{4.4-10}$$

The Weierstrass-Erdmann corner conditions are

$$\begin{aligned} &2\dot{x}^*(t_1^-)[1 - 2\dot{x}^*(t_1^-)][1 - \dot{x}^*(t_1^-)] \\ &= 2\dot{x}^*(t_1^+)[1 - 2\dot{x}^*(t_1^+)][1 - \dot{x}^*(t_1^+)] \end{aligned} \tag{4.4-11a}$$

and

$$\begin{aligned} &\dot{x}^{*2}(t_1^-)[1 - \dot{x}^*(t_1^-)][3\dot{x}^*(t_1^-) - 1] \\ &= \dot{x}^{*2}(t_1^+)[1 - \dot{x}^*(t_1^+)][3\dot{x}^*(t_1^+) - 1]. \end{aligned} \tag{4.4-11b}$$

Equation (4.4-11a) is satisfied by $\dot{x}^*(t_1^-) = 0, \frac{1}{2}, 1$ and $\dot{x}^*(t_1^+) = 0, \frac{1}{2}, 1$ in any combinations. Equation (4.4-11b) is satisfied by $\dot{x}^*(t_1^-) = 0, 1, \frac{1}{3}$ and $\dot{x}^*(t_1^+) = 0, 1, \frac{1}{3}$ in any combinations. Together these requirements give

$$\dot{x}^*(t_1^-) = 0 \quad \text{and} \quad \dot{x}^*(t_1^+) = 1,$$

or

$$\dot{x}^*(t_1^-) = 1 \quad \text{and} \quad \dot{x}^*(t_1^+) = 0$$

as the only nontrivial possibilities.

The curves labeled a, b, c in Fig. 4-17 are all extremals for this example. By inspection of the functional we see that each of these curves makes $J = 0$. Notice that if the admissible curves had been required to have continuous derivatives, the extremal would have been the straight line joining the points $x(0) = 0$ and $x(2) = 1$ (curve d in Fig. 4-17). The reader can verify that this curve makes $J = 0.125$.

Example 4.4-2. Find an extremal for the functional

$$J(x) = \int_0^{\pi/2} [\dot{x}^2(t) - x^2(t)]\, dt \tag{4.4-12}$$

with $x(0) = 0$ and $x(\pi/2) = 1$. Assume that $\dot{x}$ may have corners.

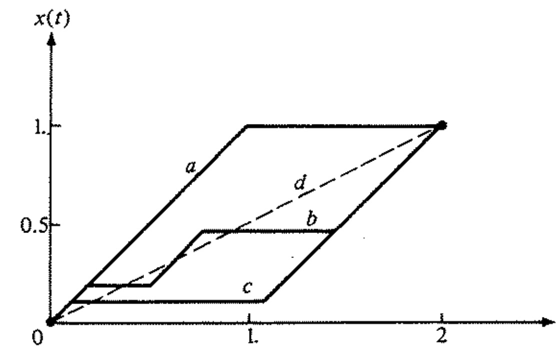


Figure 4-17 Extremal curves for Example 4.4-1

This problem was previously solved (see Example 4.2-1) under the assumption that x was required to be a smooth curve. The Euler equation

$$\ddot{x}^*(t) + x^*(t) = 0 \tag{4.4-13}$$

has a solution of the form

$$x^*(t) = c_3 \cos t + c_4 \sin t. \tag{4.4-14}$$

The Weierstrass-Erdmann corner conditions are

$$2\dot{x}^*(t_1^-) = 2\dot{x}^*(t_1^+), \tag{4.4-15a}$$

and

$$\begin{aligned} &\dot{x}^{*2}(t_1^-) - x^{*2}(t_1) - [2\dot{x}^*(t_1^-)]\dot{x}^*(t_1^-) \\ &= \dot{x}^{*2}(t_1^+) - x^{*2}(t_1) - [2\dot{x}^*(t_1^+)]\dot{x}^*(t_1^+). \end{aligned} \tag{4.4-15b}$$

From Eq. (4.4-15) we see that there can be no corners, because $\dot{x}^*(t_1^-)$ must equal $\dot{x}^*(t_1^+)$. So the extremal is, as in Example 4.2-1,

$$x^*(t) = \sin t. \tag{4.4-16}$$

Let us now consider an example in which the coordinates of the corner are constrained.

Example 4.4-3. Find the shortest piecewise-smooth curve joining the points $x(0) = 1.5$ and $x(1.5) = 0$ which intersects the line $x(t) = -t + 2$ at one point.

The functional to be minimized is (see Example 4.2-2)

$$J(x) = \int_0^{1.5} [1 + \dot{x}^2(t)]^{1/2}\, dt. \tag{4.4-17}$$

The solutions of the Euler equation are of the form

$$x^*(t) = c_1 t + c_2. \tag{4.4-18}$$

In this case the corner condition of Eq. (4.4-7) becomes

$$\begin{aligned}&\frac{\dot{x}^*(t_1^-)}{[1 + \dot{x}^{*2}(t_1^-)]^{1/2}}[-1 - \dot{x}^*(t_1^-)] + [1 + \dot{x}^{*2}(t_1^-)]^{1/2}\\ &= \frac{\dot{x}^*(t_1^+)}{[1 + \dot{x}^{*2}(t_1^+)]^{1/2}}[-1 - \dot{x}^*(t_1^+)] + [1 + \dot{x}^{*2}(t_1^+)]^{1/2}.\end{aligned} \tag{4.4-19}$$

Putting both sides over common denominators and reducing, we obtain

$$\frac{1 - \dot{x}^*(t_1^-)}{[1 + \dot{x}^{*2}(t_1^-)]^{1/2}} = \frac{1 - \dot{x}^*(t_1^+)}{[1 + \dot{x}^{*2}(t_1^+)]^{1/2}}. \tag{4.4-20}$$

The extremal subarcs have the form given by Eq. (4.4-18), but the constants of integration will generally be different on the two sides of the corner, so let

$$x^*(t) = c_1 t + c_2 \quad \text{for } t \in [0, t_1] \tag{4.4-21a}$$

$$x^*(t) = c_3 t + c_4 \quad \text{for } t \in [t_1, 1.5]. \tag{4.4-21b}$$

Substituting the derivatives of Eqs. (4.4-21) into (4.4-20) yields

$$\frac{1 - c_1}{[1 + c_1^2]^{1/2}} = \frac{1 - c_3}{[1 + c_3^2]^{1/2}} \tag{4.4-22}$$

The extremals must also satisfy the boundary conditions $x(0) = 1.5$ and $x(1.5) = 0$, so

$$c_1 \cdot 0 + c_2 = 1.5 \Longrightarrow c_2 = 1.5 \tag{4.4-23}$$

$$1.5c_3 + c_4 = 0. \tag{4.4-24}$$

At a corner, it must also be true that $x(t_1) = -t_1 + 2$; therefore, we have the additional equations

$$c_1 t_1 + c_2 = -t_1 + 2 \tag{4.4-25}$$

$$c_3 t_1 + c_4 = -t_1 + 2. \tag{4.4-26}$$

Equations (4.4-22) through (4.4-26) are a set of five nonlinear algebraic equations in the five unknowns c_1, c_2, c_3, c_4, and t_1. These equations can be solved by using (4.4-23) through (4.4-26) to express c_1 and c_3 solely in terms of t_1, substituting these expressions in Eq. (4.4-22), and solving for t_1. Doing this gives

$$\begin{aligned}x^*(t) &= -0.5t + 1.5, && t \in [0, 1.0]\\ x^*(t) &= -2t + 3, && t \in [1.0, 1.5]\end{aligned} \tag{4.4-27}$$

and $t_1 = 1.0$. This solution is shown in Fig. 4-18. The reader can show that we have found the shortest path to be the one whose angle of incidence θ_1 equals its angle of reflection θ_2. For further generalizations see reference [E-1], Chapter 2.

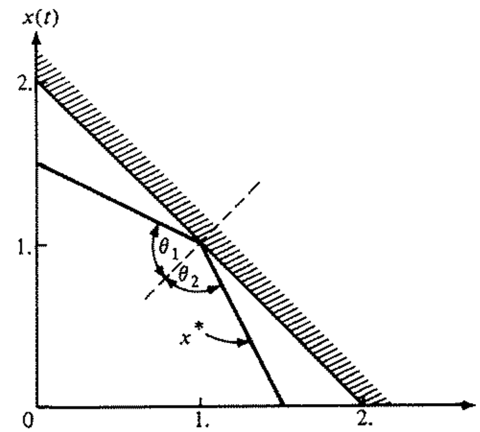


Figure 4-18 An extremal with a reflection

4.5 CONSTRAINED EXTREMA

So far, we have discussed functionals involving $\mathbf{x}$ and $\dot{\mathbf{x}}$, and we have derived necessary conditions for extremals *assuming that the components of* $\mathbf{x}$ *are independent*. In control problems the situation is more complicated, because the state trajectory is determined by the control $\mathbf{u}$; thus, we wish to consider functionals of $n + m$ functions, $\mathbf{x}$ and $\mathbf{u}$, but only m of the functions are independent—the controls. Let us now extend the necessary conditions we have derived to include problems with constraints.

To begin, we shall review the analogous problem from the calculus, and introduce some new variables—the Lagrange multipliers—that will be required for our subsequent discussion.

Constrained Minimization of Functions

Example 4.5-1. Find the point on the line $y_1 + y_2 = 5$ that is nearest the origin.

To solve this problem we need only apply elementary plane geometry to Fig. 4-19 to obtain the result that the minimum distance is $5/\sqrt{2}$, and the extreme point is $y_1^* = 2.5$, $y_2^* = 2.5$.

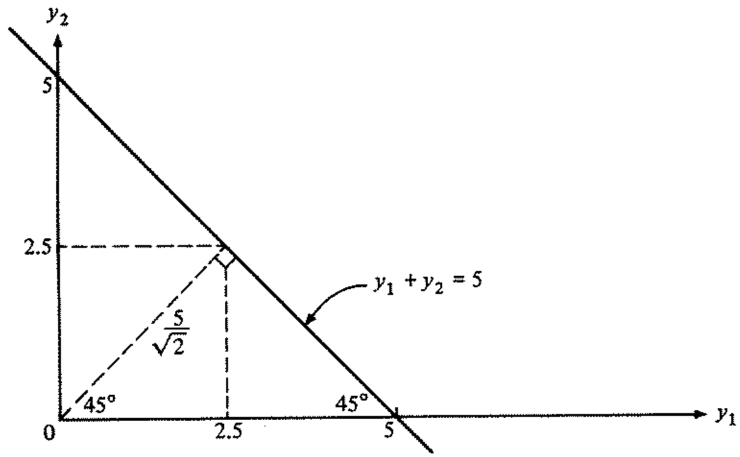


Figure 4-19 Geometrical interpretation of Example 4.5-1

Most problems cannot be solved by inspection, so let us consider alternative methods of solving this simple example.

The Elimination Method. If $\mathbf{y}^*$ is an extreme point of a function, it is necessary that the differential of the function, evaluated at $\mathbf{y}^*$, be zero.† In our example, the function

$$f(y_1, y_2) = y_1^2 + y_2^2 \quad \text{(the square of the distance)} \tag{4.5-1}$$

is to be minimized subject to the constraint

$$y_1 + y_2 = 5. \tag{4.5-2}$$

The differential is

$$df(y_1, y_2) = \left[\frac{\partial f}{\partial y_1}(y_1, y_2)\right]\Delta y_1 + \left[\frac{\partial f}{\partial y_2}(y_1, y_2)\right]\Delta y_2, \tag{4.5-3}$$

and if (y_1^*, y_2^*) is an extreme point,

$$df(y_1^*, y_2^*) = \left[\frac{\partial f}{\partial y_1}(y_1^*, y_2^*)\right]\Delta y_1 + \left[\frac{\partial f}{\partial y_2}(y_1^*, y_2^*)\right]\Delta y_2 = 0. \tag{4.5-4}$$

If y_1 and y_2 were independent, then Δy_1 and Δy_2 could be selected arbitrarily and Eq. (4.5-4) would imply that the partial derivatives must both be zero. In this example, however, y_1 and y_2 are constrained to lie on the specified line, so Δy_1 and Δy_2 are not independent. Solving Eq. (4.5-2) for y_1 and substituting into (4.5-1), we obtain

† Only interior points of bounded regions are considered.

$$\begin{aligned} f(y_2) &= [5 - y_2]^2 + y_2^2 \\ &= 25 - 10y_2 + 2y_2^2 \end{aligned} \tag{4.5-5}$$

The differential of f at the point y_2^* is then

$$df(y_2^*) = [-10 + 4y_2^*]\,\Delta y_2 = 0, \tag{4.5-6}$$

so $y_2^* = 2.5$. From (4.5-2) we then find that $y_1^* = 2.5$. The minimum value of the function is $\frac{25}{2}$, and the minimum distance is $5/\sqrt{2}$.†

The Lagrange Multiplier Method. Consider the *augmented* function

$$f_a(y_1, y_2, p) \triangleq y_1^2 + y_2^2 + p[y_1 + y_2 - 5], \tag{4.5-7}$$

with p a variable (the Lagrange multiplier) whose value is yet to be determined. For values of y_1 and y_2 that satisfy the constraining relation (4.5-2) (these are the only values of interest), the augmented function f_a equals f regardless of the value of p—we have simply added zero to f to obtain f_a. By satisfying the constraint and minimizing f_a, the constrained extreme point of f can be found. To find an extreme point of f_a, we use the necessary condition

$$\begin{aligned} df_a(y_1^*, y_2^*, p) = 0 = [2y_1^* + p]\,\Delta y_1 + [2y_2^* + p]\,\Delta y_2 \\ + [y_1^* + y_2^* - 5]\,\Delta p. \end{aligned} \tag{4.5-8}$$

Since only points that satisfy the constraining relation are acceptable,

$$y_1^* + y_2^* - 5 = 0, \tag{4.5-9}$$

but this is the coefficient of Δp. The remaining two terms must add to zero, but Δy_1 and Δy_2 are not independent—if Δy_1 is selected Δy_2 is determined, and vice versa; however, p comes to the rescue. Since the constraint must be satisfied, p can be any value, so we make a convenient choice—we select p so that the coefficient of Δy_2 (or Δy_1) is zero, and we denote this value of p by p^*. Then we have

$$2y_2^* + p^* = 0. \tag{4.5-10}$$

Δy_1 can assume arbitrary values; for each value of Δy_1 there is an associated dependent value of Δy_2, but this does not matter, because p was selected to make the coefficient of Δy_2 equal to zero. Since df_a must be zero and Δy_1 is arbitrary, the coefficient of Δy_1 must be zero; therefore,

$$2y_1^* + p^* = 0. \tag{4.5-11}$$

† Alternatively, we could reach the same final result by substituting $y_1 = 5 - y_2$ and $\Delta y_1 = \Delta y_2$ into Eq. (4.5-4), setting the coefficient of Δy_2 to zero, and solving for y_2^*.

Solving (4.5-9), (4.5-10), and (4.5-11) simultaneously gives

$$y_1^* = 2.5, \qquad y_2^* = 2.5, \qquad p^* = -5. \tag{4.5-12}$$

The reasoning that led to Eqs. (4.5-9), (4.5-10), and (4.5-11) is very important; we shall use it again shortly. Notice, however, that the same equations are obtained by forming $f_a(y_1, y_2, p)$ and then treating the three variables *as if* they were independent.

Let us now consider the "elimination method" and the method of Lagrange multipliers as they are applied in a general problem.

The problem is to find the extreme values for a function of $(n + m)$ variables, $y_1, \ldots, y_{n+m}$. The function that is to be extremized is given by $f(y_1, y_2, \ldots, y_{n+m})$. There are n constraints among the variables of the form

$$\begin{aligned} a_1(y_1, \ldots, y_{n+m}) &= 0 \\ &\vdots \\ a_n(y_1, \ldots, y_{n+m}) &= 0; \end{aligned} \tag{4.5-13}$$

thus, only $(n + m) - n = m$ of the variables are independent. Using the elimination method, we solve Eq. (4.5-13) for n of the variables in terms of the remaining m variables. For example, solving for the first n variables gives

$$\begin{aligned} y_1 &= e_1(y_{n+1}, \ldots, y_{n+m}) \\ &\vdots \\ y_n &= e_n(y_{n+1}, \ldots, y_{n+m}). \end{aligned} \tag{4.5-14}$$

Substituting these relations into f, we obtain a function of m *independent* variables, $f(y_{n+1}, \ldots, y_{n+m})$. To find the minimum value of this function, we solve the equations

$$\begin{aligned} \frac{\partial f}{\partial y_{n+1}}(y_{n+1}^*, \ldots, y_{n+m}^*) &= 0 \\ &\vdots \\ \frac{\partial f}{\partial y_{n+m}}(y_{n+1}^*, \ldots, y_{n+m}^*) &= 0 \end{aligned} \tag{4.5-15}$$

for $y_{n+1}^*, \ldots, y_{n+m}^*$, and substitute these values in (4.5-14) to obtain $y_1^*, \ldots, y_n^*$. The extreme value of f can then also be obtained. This procedure is conceptually straightforward; the principal difficulty is in obtaining the relations (4.5-14). The solution of (4.5-15) may also be difficult, but this problem is also present in the method of Lagrange multipliers.

Now let us consider the method of Lagrange multipliers. First, we form the augmented function

$$\begin{aligned} f_a(y_1, \ldots, y_{n+m}, p_1, \ldots, p_n) &\triangleq f(y_1, \ldots, y_{n+m}) \\ &+ p_1[a_1(y_1, \ldots, y_{n+m})] + \cdots + p_n[a_n(y_1, \ldots, y_{n+m})]. \end{aligned} \tag{4.5-16}$$

Then

$$\begin{aligned} df_a &= \frac{\partial f_a}{\partial y_1}\Delta y_1 + \cdots + \frac{\partial f_a}{\partial y_{n+m}}\Delta y_{n+m} + \frac{\partial f_a}{\partial p_1}\Delta p_1 + \cdots + \frac{\partial f_a}{\partial p_n}\Delta p_n \\ &= \frac{\partial f_a}{\partial y_1}\Delta y_1 + \cdots + \frac{\partial f_a}{\partial y_{n+m}}\Delta y_{n+m} + a_1\,\Delta p_1 + \cdots + a_n\,\Delta p_n. \end{aligned} \tag{4.5-17}$$

If the constraints are satisfied, the coefficients of $\Delta p_1, \ldots, \Delta p_n$ are zero. We then select the n p_i's so that the coefficients of Δy_i $(i = 1, \ldots, n)$ are zero. The remaining m Δy_i's are independent, and for df_a to equal zero their coefficients must vanish. The result is that the extreme point $y_1^*, \ldots, y_{n+m}^*$ is found by solving the equations

$$\left.\begin{aligned} &a_i(y_1^*, \ldots, y_{n+m}^*) = 0, && i = 1, 2, \ldots, n \\ &\frac{\partial f_a}{\partial y_j}(y_1^*, \ldots, y_{n+m}^*, p_1^*, \ldots, p_n^*) = 0, && j = 1, 2, \ldots, n+m \end{aligned}\right\} \begin{matrix} 2n+m \\ \text{equations} \end{matrix} \tag{4.5-18}$$

We shall now conclude our consideration of the calculus problem with another illustrative example.

Example 4.5-2 [H-1]. Find the point in three-dimensional Euclidean space that is nearest the origin and lies on the intersection of the surfaces

$$\begin{aligned} y_3 &= y_1 y_2 + 5 \\ y_1 + y_2 + y_3 &= 1. \end{aligned} \tag{4.5-19}$$

The function to be minimized is

$$f(y_1, y_2, y_3) = y_1^2 + y_2^2 + y_3^2. \tag{4.5-20}$$

The elimination method is left as an exercise for the reader. To use the method involving Lagrange multipliers, first form the augmented function

$$\begin{aligned} f_a(y_1, y_2, y_3, p_1, p_2) = y_1^2 + y_2^2 + y_3^2 &+ p_1[y_1 y_2 + 5 - y_3] \\ &+ p_2[y_1 + y_2 + y_3 - 1]. \end{aligned} \tag{4.5-21}$$

Using the same reasoning as before, we find that the equations corresponding to (4.5-18) are

$$\begin{aligned} y_1^* + y_2^* + y_3^* - 1 &= 0 \\ y_1^* y_2^* + 5 - y_3^* &= 0 \\ 2y_1^* + p_1^* y_2^* + p_2^* &= 0 \\ 2y_2^* + p_1^* y_1^* + p_2^* &= 0 \\ 2y_3^* - p_1^* + p_2^* &= 0. \end{aligned} \tag{4.5-22}$$

Solving these five equations gives

$$(y_1^*, y_2^*, y_3^*) = \begin{cases} (2, -2, 1) \\ \text{or} \\ (-2, 2, 1) \end{cases} \tag{4.5-23}$$

and $f_{a_{\min}} = 9$, so the distance is 3.

Constrained Minimization of Functionals

We are now ready to consider the presence of constraints in variational problems. To simplify the variational equations, it will be assumed that the admissible curves are smooth.

Point Constraints. Let us determine a set of necessary conditions for a function $\mathbf{w}^*$ to be an extremal for a functional of the form

$$J(\mathbf{w}) = \int_{t_0}^{t_f} g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\, dt; \tag{4.5-24}$$

$\mathbf{w}$ is an $(n + m) \times 1$ vector of functions $(n, m \geq 1)$ that is required to satisfy n relationships of the form

$$f_i(\mathbf{w}(t), t) = 0, \qquad i = 1, 2, \ldots, n, \tag{4.5-25}$$

which are called *point constraints.* Constraints of this type would be present if, for example, the admissible trajectories were required to lie on a specified surface in the $n + m + 1$-dimensional $\mathbf{w}(t) - t$ space. The presence of these n constraining relations means that only m of the $n + m$ components of $\mathbf{w}$ are independent.

We have previously found that the Euler equations must be satisfied regardless of the boundary conditions, so we will ignore, temporarily, terms that enter only into the determination of boundary conditions.

One way to attack this problem might be to solve Eqs. (4.5-25) for n

of the components of $\mathbf{w}(t)$ in terms of the remaining m components—which can then be regarded as m independent functions—and use these equations to eliminate the n dependent components of $\mathbf{w}(t)$ and $\dot{\mathbf{w}}(t)$ from J. If this can be done, then the equations of Sections 4.2 and 4.3 apply. Unfortunately, the constraining equations (4.5-25) are generally nonlinear algebraic equations, which may be quite difficult to solve.

As an alternative approach we can use Lagrange multipliers. The first step is to form the *augmented functional* by adjoining the constraining relations to J, which yields

$$\begin{aligned} J_a(\mathbf{w}, \mathbf{p}) &= \int_{t_0}^{t_f} \{g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + p_1(t)[f_1(\mathbf{w}(t), t)] \\ &\qquad + p_2(t)[f_2(\mathbf{w}(t), t)] + \cdots + p_n(t)[f_n(\mathbf{w}(t), t)]\}\, dt \\ &= \int_{t_0}^{t_f} \{g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + \mathbf{p}^T(t)[\mathbf{f}(\mathbf{w}(t), t)]\}\, dt. \end{aligned} \tag{4.5-26}$$

Since the constraints must be satisfied for all $t \in [t_0, t_f]$, the Lagrange multipliers $p_1, \ldots, p_n$ are assumed to be functions of time. This allows us the flexibility of multiplying the constraining relations by a *different* real number for each value of t; the reason for desiring this flexibility will become clear as we proceed.

Notice that if the constraints are satisfied, $J_a = J$ for any function $\mathbf{p}$. The variation of the functional J_a,

$$\begin{aligned} \delta J_a(\mathbf{w}, \delta\mathbf{w}, \mathbf{p}, \delta\mathbf{p}) = \int_{t_0}^{t_f} \Big\{ \Big[\frac{\partial g}{\partial \mathbf{w}}^T(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + \mathbf{p}^T(t) \Big[\frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\mathbf{w}(t), t) \Big] \Big] \delta\mathbf{w}(t) \\ + \Big[\frac{\partial g}{\partial \dot{\mathbf{w}}}^T(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \Big] \delta\dot{\mathbf{w}}(t) + [\mathbf{f}^T(\mathbf{w}(t), t)]\, \delta\mathbf{p}(t) \Big\}\, dt, \end{aligned} \tag{4.5-27}$$

is found in the usual manner by introducing variations in the functions $\mathbf{w}$, $\dot{\mathbf{w}}$, and $\mathbf{p}$. $\partial\mathbf{f}/\partial\mathbf{w}$ denotes the $n \times (n + m)$ matrix

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial w_1} & \cdots & \dfrac{\partial f_1}{\partial w_{n+m}} \\ \vdots & & \vdots \\ \dfrac{\partial f_n}{\partial w_1} & \cdots & \dfrac{\partial f_n}{\partial w_{n+m}} \end{bmatrix}.$$

Integrating by parts the term containing $\delta\dot{\mathbf{w}}$ and retaining only the terms inside the integral, we obtain

$$\delta J_a(\mathbf{w}, \delta\mathbf{w}, \mathbf{p}, \delta\mathbf{p}) = \int_{t_0}^{t_f} \Big\{ \Big[\frac{\partial g^T}{\partial \mathbf{w}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + \mathbf{p}^T(t) \Big[\frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\mathbf{w}(t), t) \Big] - \frac{d}{dt} \Big[\frac{\partial g^T}{\partial \dot{\mathbf{w}}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \Big] \Big] \delta\mathbf{w}(t) + [\mathbf{f}^T(\mathbf{w}(t), t)] \, \delta\mathbf{p}(t) \Big\} \, dt. \tag{4.5-28}$$

On an extremal, the variation must be zero; that is, $\delta J_a(\mathbf{w}^*, \mathbf{p}) = 0$. In addition, the point constraints must also be satisfied by an extremal; therefore,

$$\mathbf{f}(\mathbf{w}^*(t), t) = \mathbf{0}, \qquad t \in [t_0, t_f], \tag{4.5-29}$$

and the coefficient of $\delta\mathbf{p}(t)$ in Eq. (4.5-28) is zero. Since the constraints are satisfied, we can select the n Lagrange multipliers arbitrarily—let us choose the p's so that the coefficients of n of the components of $\delta\mathbf{w}(t)$ are zero throughout the interval $[t_0, t_f]$. The remaining $(n + m) - n = m$ components of $\delta\mathbf{w}$ are then independent; hence, the coefficients of these components of $\delta\mathbf{w}(t)$ must be zero. The final result is that, in addition to Eq. (4.5-29), the equations

$$\frac{\partial g}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) + \Big[\frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\mathbf{w}^*(t), t) \Big]^T \mathbf{p}^*(t) - \frac{d}{dt} \Big[\frac{\partial g}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) \Big] = \mathbf{0} \tag{4.5-30}$$

must be satisfied.

If we define the *augmented integrand function* as

$$g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), t) \triangleq g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + \mathbf{p}^T(t)[\mathbf{f}(\mathbf{w}(t), t)], \tag{4.5-31}$$

then Eq. (4.5-30) can be written

$$\boxed{\frac{\partial g_a}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), t) - \frac{d}{dt} \Big[\frac{\partial g_a}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), t) \Big] = \mathbf{0}} \tag{4.5-30a}$$

Equations (4.5-30a) are a set of $n + m$ second-order differential equations, and the constraining relations (4.5-29) are a set of n algebraic equations. Together, these $2n + m$ equations constitute a set of necessary conditions for $\mathbf{w}^*$ to be an extremal.

The reader may have already noticed that Eqs. (4.5-29) and (4.5-30a) are

the same as if the results from *Problem 1a* had been applied to the functional

$$J_a(\mathbf{w}, \mathbf{p}) = \int_{t_0}^{t_f} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), t) \, dt \tag{4.5-32}$$

with the assumption that the functions $\mathbf{w}$ and $\mathbf{p}$ are independent. It should be emphasized that, although the results are the same, the reasoning used is quite different.

Example 4.5-3. Find necessary conditions that must be satisfied by the curve of smallest length which lies on the sphere $w_1^2(t) + w_2^2(t) + t^2 = R^2$, for $t \in [t_0, t_f]$, and joins the specified points $\mathbf{w}_0, t_0$, and $\mathbf{w}_f, t_f$.

The functional to be minimized is

$$J(\mathbf{w}) = \int_{t_0}^{t_f} [1 + \dot{w}_1^2(t) + \dot{w}_2^2(t)]^{1/2} \, dt, \tag{4.5-33}$$

so the augmented integrand function is

$$g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), p(t), t) = [1 + \dot{w}_1^2(t) + \dot{w}_2^2(t)]^{1/2} + p(t)[w_1^2(t) + w_2^2(t) + t^2 - R^2]. \tag{4.5-34}$$

Performing the operations indicated by Eq. (4.5-30a) gives

$$2w_1^*(t)p^*(t) - \frac{d}{dt}\Big\{ \frac{\dot{w}_1^*(t)}{[1 + \dot{w}_1^{*2}(t) + \dot{w}_2^{*2}(t)]^{1/2}} \Big\} = 0 \tag{4.5-35a}$$

$$2w_2^*(t)p^*(t) - \frac{d}{dt}\Big\{ \frac{\dot{w}_2^*(t)}{[1 + \dot{w}_1^{*2}(t) + \dot{w}_2^{*2}(t)]^{1/2}} \Big\} = 0. \tag{4.5-35b}$$

In addition, of course, it is necessary that the constraining relation

$$w_1^{*2}(t) + w_2^{*2}(t) + t^2 = R^2 \tag{4.5-35c}$$

be satisfied.

Differential Equation Constraints. Let us now find necessary conditions for a function $\mathbf{w}^*$ to be an extremal for a functional

$$J(\mathbf{w}) = \int_{t_0}^{t_f} g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \, dt. \tag{4.5-36}$$

$\mathbf{w}$ is an $(n + m) \times 1$ vector of functions $(n, m \geq 1)$ which must satisfy the n differential equations

$$f_i(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) = 0, \qquad i = 1, 2, \ldots, n. \tag{4.5-37}$$

Because of the n differential equation constraints, only m of the $n + m$

components of $\mathbf{w}$ are independent. Constraints of this type may represent the state equation constraints in optimal control problems where $\mathbf{w}$ corresponds to the $n + m$ vector $[\mathbf{x} \mid \mathbf{u}]^T$.

As with point constraints, it is generally not feasible to eliminate n dependent functions and their derivatives from the functional J, so we shall again use the method of Lagrange multipliers. The derivation proceeds along the same lines as for problems with point constraints; that is, we first form the augmented functional

$$\begin{aligned} J_a(\mathbf{w}, \mathbf{p}) &= \int_{t_0}^{t_f} \{g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + p_1(t)[f_1(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)] \\ &\quad + p_2(t)[f_2(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)] + \cdots \\ &\quad + p_n(t)[f_n(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)]\}\, dt \\ &= \int_{t_0}^{t_f} \{g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + \mathbf{p}^T(t)[\mathbf{f}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)]\}\, dt. \end{aligned} \tag{4.5-38}$$

Again notice that if the constraints are satisfied, $J_a = J$ for any $\mathbf{p}(t)$. The variation of the functional J_a,

$$\begin{aligned} \delta J_a(\mathbf{w}, \delta\mathbf{w}, \mathbf{p}, \delta\mathbf{p}) = \int_{t_0}^{t_f} \Bigg\{ &\left[\frac{\partial g^T}{\partial \mathbf{w}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \right.\\ &\left. + \mathbf{p}^T(t)\left[\frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\right]\right]\delta\mathbf{w}(t) \\ &+ \left[\frac{\partial g^T}{\partial \dot{\mathbf{w}}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\right. \\ &\left. + \mathbf{p}^T(t)\left[\frac{\partial \mathbf{f}}{\partial \dot{\mathbf{w}}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\right]\right]\delta\dot{\mathbf{w}}(t) \\ &+ [\mathbf{f}^T(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)]\,\delta\mathbf{p}(t)\Bigg\}\, dt, \end{aligned} \tag{4.5-39}$$

is found in the usual manner by introducing variations in the functions $\mathbf{w}$, $\dot{\mathbf{w}}$, and $\mathbf{p}$. The notation $\partial\mathbf{f}/\partial\dot{\mathbf{w}}$ means

$$\begin{bmatrix} \dfrac{\partial f_1}{\partial \dot{w}_1} & \cdots & \dfrac{\partial f_1}{\partial \dot{w}_{n+m}} \\ \vdots & & \vdots \\ \dfrac{\partial f_n}{\partial \dot{w}_1} & \cdots & \dfrac{\partial f_n}{\partial \dot{w}_{n+m}} \end{bmatrix}.$$

Integrating by parts the terms containing $\delta\dot{\mathbf{w}}$ and retaining only the terms inside the integral, we obtain

$$\begin{aligned} \delta J_a(\mathbf{w}, \delta\mathbf{w}, \mathbf{p}, \delta\mathbf{p}) = \int_{t_0}^{t_f} \Bigg\{ &\left[\frac{\partial g^T}{\partial \mathbf{w}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \right.\\ &+ \mathbf{p}^T(t)\left[\frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\right] \\ &- \frac{d}{dt}\left[\frac{\partial g^T}{\partial \dot{\mathbf{w}}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\right. \\ &\left.\left. + \mathbf{p}^T(t)\left[\frac{\partial \mathbf{f}}{\partial \dot{\mathbf{w}}}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\right]\right]\right]\delta\mathbf{w}(t) \\ &+ [\mathbf{f}^T(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)]\,\delta\mathbf{p}(t)\Bigg\}\, dt. \end{aligned} \tag{4.5-40}$$

On an extremal, the variation must be zero, that is, $\delta J_a(\mathbf{w}^*, \mathbf{p}) = 0$, and the differential equation constraints must also be satisfied; therefore,

$$\mathbf{f}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) = \mathbf{0}, \tag{4.5-41}$$

and the coefficient of $\delta\mathbf{p}(t)$ in Eq. (4.5-40) is zero. Since the constraints are satisfied, we can choose the n Lagrange multipliers arbitrarily—let us select the p's so that the coefficients of n of the components of $\delta\mathbf{w}(t)$ are zero throughout the interval $[t_0, t_f]$. The remaining $(n + m) - n = m$ components of $\delta\mathbf{w}$ are then independent; hence, the coefficients of these components of $\delta\mathbf{w}(t)$ must be zero. The final result is that, in addition to Eq. (4.5-41), the equations

$$\begin{aligned} &\frac{\partial g}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) + \left[\frac{\partial \mathbf{f}}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t)\right]^T \mathbf{p}^*(t) \\ &- \frac{d}{dt}\left\{\frac{\partial g}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) + \left[\frac{\partial \mathbf{f}}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t)\right]^T \mathbf{p}^*(t)\right\} = \mathbf{0} \end{aligned} \tag{4.5-42}$$

must be satisfied.

If we define the augmented integrand function as

$$\begin{aligned} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), t) \\ = g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) + \mathbf{p}^T(t)[\mathbf{f}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)] \end{aligned} \tag{4.5-43}$$

then Eq. (4.5-42) can be written

$$\boxed{\frac{\partial g_a}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g_a}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), t)\right] = \mathbf{0}.} \tag{4.5-42a}$$

Equations (4.5-41) and (4.5-42a) comprise a set of $(2n + m)$ second-order differential equations. We shall see in Chapter 5 that in optimal control problems m of these equations are algebraic, and the remaining $2n$ differential equations are first order.

Equations (4.5-41) and (4.5-42a) are the same as if the results of *Problem 1a* had been applied to the functional

$$J_a(\mathbf{w}, \mathbf{p}) = \int_{t_0}^{t_f} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), t)\, dt \tag{4.5-44}$$

with the assumption that the functions $\mathbf{w}$ and $\mathbf{p}$ are independent. Again we emphasize that although the results are the same, the reasoning is quite different!

Example 4.5-4. Find the equations that must be satisfied by an extremal for the functional

$$J(\mathbf{w}) = \int_{t_0}^{t_f} \tfrac{1}{2}[w_1^2(t) + w_2^2(t)]\, dt, \tag{4.5-45}$$

where the functions w_1 and w_2 are related by

$$\dot{w}_1(t) = w_2(t). \tag{4.5-46}$$

There is one constraint, so the function $\mathbf{f}$ in Eq. (4.5-41) is

$$f(\mathbf{w}(t), \dot{\mathbf{w}}(t)) = w_2(t) - \dot{w}_1(t), \tag{4.5-47}$$

and one Lagrange multiplier $p(t)$ is required. The function g_a in Eq. (4.5-43) is

$$g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), p(t)) = \tfrac{1}{2}w_1^2(t) + \tfrac{1}{2}w_2^2(t) + p(t)w_2(t) - p(t)\dot{w}_1(t). \tag{4.5-48}$$

From Eq. (4.5-42a) we have

$$\begin{aligned} w_1^*(t) + \dot{p}^*(t) &= 0 \\ w_2^*(t) + p^*(t) &= 0, \end{aligned} \tag{4.5-49}$$

and satisfaction of (4.5-46) requires that

$$\dot{w}_1^*(t) = w_2^*(t). \tag{4.5-46a}$$

Equations (4.5-49) and (4.5-46a) are necessary conditions for $\mathbf{w}^*$ to be an extremal.

Example 4.5-5. Suppose that the system

$$\begin{aligned} \dot{x}_1(t) &= x_2(t) - x_1(t) \\ \dot{x}_2(t) &= -2x_1(t) - 3x_2(t) + u(t) \end{aligned} \tag{4.5-50}$$

is to be controlled to minimize the performance measure

$$J(\mathbf{x}, u) = \int_{t_0}^{t_f} \tfrac{1}{2}[x_1^2(t) + x_2^2(t) + u^2(t)]\, dt. \tag{4.5-51}$$

Find a set of necessary conditions for optimal control.

If we define $x_1 \triangleq w_1$, $x_2 \triangleq w_2$, and $u \triangleq w_3$, the problem statement and solution, using the notation of this section, are the following.

Find the equations that must be satisfied for a function $\mathbf{w}^*$ to be an extremal for the functional

$$J(\mathbf{w}) = \int_{t_0}^{t_f} \tfrac{1}{2}[w_1^2(t) + w_2^2(t) + w_3^2(t)]\, dt, \tag{4.5-52}$$

where the function $\mathbf{w}$ must satisfy the differential equation constraints

$$\begin{aligned} \dot{w}_1(t) &= w_2(t) - w_1(t) \\ \dot{w}_2(t) &= -2w_1(t) - 3w_2(t) + w_3(t). \end{aligned} \tag{4.5-53}$$

The function $\mathbf{f}$ is

$$\begin{aligned} f_1(\mathbf{w}(t), \dot{\mathbf{w}}(t)) &= w_2(t) - w_1(t) - \dot{w}_1(t) = 0 \\ f_2(\mathbf{w}(t), \dot{\mathbf{w}}(t)) &= -2w_1(t) - 3w_2(t) + w_3(t) - \dot{w}_2(t) = 0, \end{aligned} \tag{4.5-54}$$

and g_a is given by

$$\begin{aligned} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t)) = {} & \tfrac{1}{2}w_1^2(t) + \tfrac{1}{2}w_2^2(t) + \tfrac{1}{2}w_3^2(t) \\ & + p_1(t)[w_2(t) - w_1(t) - \dot{w}_1(t)] \\ & + p_2(t)[-2w_1(t) - 3w_2(t) + w_3(t) - \dot{w}_2(t)]. \end{aligned} \tag{4.5-55}$$

From Eq. (4.5-42a), we obtain the differential equations

$$\begin{aligned} \dot{p}_1^*(t) &= -w_1^*(t) + p_1^*(t) + 2p_2^*(t) \\ \dot{p}_2^*(t) &= -w_2^*(t) - p_1^*(t) + 3p_2^*(t), \end{aligned} \tag{4.5-56}$$

and the algebraic equation (since $\dot{w}_3$ does not appear in g_a),

$$w_3^*(t) + p_2^*(t) = 0. \tag{4.5-57}$$

The two additional equations that must be satisfied by an extremal are the constraints

$$\begin{aligned} \dot{w}_1^*(t) &= w_2^*(t) - w_1^*(t) \\ \dot{w}_2^*(t) &= -2w_1^*(t) - 3w_2^*(t) + w_3^*(t). \end{aligned} \tag{4.5-58}$$

Isoperimetric Constraints. Queen Dido's land transaction was perhaps the original problem with an *isoperimetric constraint*—she attempted to find the curve having a fixed length which enclosed the maximum area. Today, we say that any constraints of the form

$$\int_{t_0}^{t_f} e_i(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\,dt = c_i \qquad (i = 1, 2, \ldots, r) \tag{4.5-59}$$

are isoperimetric constraints. The c_i's are specified constants. In control problems such constraints often enter in the form of total fuel or energy available to perform a required task.

Suppose that it is desired to find necessary conditions for $\mathbf{w}^*$ to be an extremal for

$$J(\mathbf{w}) = \int_{t_0}^{t_f} g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\,dt \tag{4.5-60}$$

subject to the isoperimetric constraints given in Eq. (4.5-59).

These constraints can be put into the form of differential equation constraints by defining the new variables

$$z_i(t) \triangleq \int_{t_0}^{t} e_i(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)\,dt, \qquad i = 1, 2, \ldots, r.\dagger \tag{4.5-61}$$

The required boundary conditions for these additional variables are $z_i(t_0) = 0$ and $z_i(t_f) = c_i$. Differentiating Eq. (4.5-61) with respect to time gives

$$\dot{z}_i(t) = e_i(\mathbf{w}(t), \dot{\mathbf{w}}(t), t), \qquad i = 1, 2, \ldots, r, \tag{4.5-62}$$

or, in vector notation,

$$\dot{\mathbf{z}}(t) = \mathbf{e}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t). \tag{4.5-62a}$$

Equation (4.5-62a) is a set of r differential equation constraints which we treat, as before, by forming the augmented function

$$\begin{aligned} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), \dot{\mathbf{z}}(t), t) &\triangleq g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \\ &+ \mathbf{p}^T(t)[\mathbf{e}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) - \dot{\mathbf{z}}(t)]. \end{aligned} \tag{4.5-63}$$

Corresponding to Eq. (4.5-42a), we now have the set of $n + m$ equations

$$\frac{\partial g_a}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), \dot{\mathbf{z}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g_a}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), \dot{\mathbf{z}}^*(t), t)\right] = \mathbf{0}, \tag{4.5-64}$$

and the set of r equations

$$\frac{\partial g_a}{\partial \mathbf{z}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), \dot{\mathbf{z}}^*(t), t) - \frac{d}{dt}\left[\frac{\partial g_a}{\partial \dot{\mathbf{z}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), \dot{\mathbf{z}}^*(t), t)\right] = \mathbf{0}, \tag{4.5-65}$$

† Notice that the upper limit on the integral is t, *not* t_f.

a total of $(n + m + r)$ equations involving $(n + m + r + r)$ functions $(\mathbf{w}^*, \mathbf{p}^*, \mathbf{z}^*)$. The additional r equations required are

$$\dot{\mathbf{z}}^*(t) = \mathbf{e}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) \tag{4.5-66}$$

whose solution must satisfy the boundary conditions $z_i^*(t_f) = c_i$, $i = 1, \ldots, r$.

Notice that g_a does not contain $\mathbf{z}(t)$, so $\partial g_a/\partial \mathbf{z} \equiv \mathbf{0}$. In addition, $\partial g_a/\partial \dot{\mathbf{z}} = -\mathbf{p}^*(t)$; therefore, Eq. (4.5-65) always gives

$$\dot{\mathbf{p}}^*(t) = \mathbf{0}, \tag{4.5-67}$$

which implies that the Lagrange multipliers are constants.

To summarize, for problems with isoperimetric constraints, the necessary conditions for an extremal are given by Eqs. (4.5-64), (4.5-66), and (4.5-67). The following examples illustrate the use of these equations.

Example 4.5-6. Find necessary conditions for $\mathbf{w}^*$ to be an extremal of the functional

$$J(\mathbf{w}) = \int_{t_0}^{t_f} \tfrac{1}{2}[w_1^2(t) + w_2^2(t) + 2\dot{w}_1(t)\dot{w}_2(t)]\,dt \tag{4.5-68}$$

subject to the constraint

$$\int_{t_0}^{t_f} w_2^2(t)\,dt = c; \tag{4.5-69}$$

c is a specified constant.

Let $\dot{z}(t) \triangleq w_2^2(t)$; then

$$\begin{aligned} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), p(t), \dot{z}(t)) &= \tfrac{1}{2}w_1^2(t) + \tfrac{1}{2}w_2^2(t) + \dot{w}_1(t)\dot{w}_2(t) \\ &+ p(t)[w_2^2(t) - \dot{z}(t)]. \end{aligned} \tag{4.5-70}$$

From (4.5-64),

$$\begin{aligned} w_1^*(t) - \ddot{w}_2^*(t) &= 0 \\ w_2^*(t) + 2w_2^*(t)p^*(t) - \ddot{w}_1^*(t) &= 0, \end{aligned} \tag{4.5-71}$$

and Eq. (4.5-65) gives

$$\dot{p}^*(t) = 0. \tag{4.5-72}$$

In addition, the solution of the differential equation

$$\dot{z}^*(t) = w_2^{*2}(t), \qquad z^*(t_0) = 0 \tag{4.5-73}$$

must satisfy the boundary condition

$$z^*(t_f) = c. \tag{4.5-74}$$

In control problems, there are always state differential equation constraints, in addition to any isoperimetric constraints. Let us now consider an example having both types of constraints.

Example 4.5-7. The system with state equations

$$\dot{x}_1(t) = -x_1(t) + x_2(t) + u(t)$$
$$\dot{x}_2(t) = -2x_1(t) - 3x_2(t) + u(t) \tag{4.5-75}$$

is to be controlled to minimize the performance measure

$$J(\mathbf{x}, u) = \int_{t_0}^{t_f} \tfrac{1}{2}[x_1^2(t) + x_2^2(t)]\, dt. \tag{4.5-76}$$

The total control energy to be expended is

$$\int_{t_0}^{t_f} u^2(t)\, dt = c, \tag{4.5-77}$$

where c is a specified constant. Find a set of necessary conditions for optimal control.

If we define $x_1 \triangleq w_1$, $x_2 \triangleq w_2$, and $u \triangleq w_3$, then the problem stated in the notation of this section is as follows.

Find necessary conditions that must be satisfied by an extremal for the functional

$$J(\mathbf{w}) = \int_{t_0}^{t_f} \tfrac{1}{2}[w_1^2(t) + w_2^2(t)]\, dt. \tag{4.5-78}$$

The constraining relations are

$$\dot{w}_1(t) = -w_1(t) + w_2(t) + w_3(t)$$
$$\dot{w}_2(t) = -2w_1(t) - 3w_2(t) + w_3(t) \tag{4.5-79}$$

and

$$\int_{t_0}^{t_f} w_3^2(t)\, dt = c. \tag{4.5-80}$$

First, we form the function

$$\begin{aligned} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), \dot{z}(t)) = {} & \tfrac{1}{2}w_1^2(t) + \tfrac{1}{2}w_2^2(t) \\ & + p_1(t)[-w_1(t) + w_2(t) + w_3(t) - \dot{w}_1(t)] \\ & + p_2(t)[-2w_1(t) - 3w_2(t) + w_3(t) - \dot{w}_2(t)] \\ & + p_3(t)[w_3^2(t) - \dot{z}(t)]. \end{aligned}$$

The required equations are

$$\begin{aligned} \dot{p}_1^*(t) &= p_1^*(t) + 2p_2^*(t) - w_1^*(t) \\ \dot{p}_2^*(t) &= -p_1^*(t) + 3p_2^*(t) - w_2^*(t) \\ p_1^*(t) &+ p_2^*(t) + 2w_3^*(t)p_3^*(t) = 0 \\ \dot{p}_3^*(t) &= 0 \\ \dot{w}_1^*(t) &= -w_1^*(t) + w_2^*(t) + w_3^*(t) \\ \dot{w}_2^*(t) &= -2w_1^*(t) - 3w_2^*(t) + w_3^*(t) \\ \dot{z}^*(t) &= w_3^{*2}(t), \qquad z^*(t_0) = 0. \end{aligned} \tag{4.5-81}$$

The boundary condition $z^*(t_f) = c$ must also be satisfied.

To recapitulate, the important result of this section is that a necessary condition for problems with differential equation constraints, or point constraints, is

$$\boxed{\begin{aligned} &\frac{\partial g_a}{\partial \mathbf{w}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), t) \\ &\quad - \frac{d}{dt}\left[\frac{\partial g_a}{\partial \dot{\mathbf{w}}}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), \mathbf{p}^*(t), t)\right] = \mathbf{0}, \end{aligned}} \tag{4.5-42a}$$

where

$$\begin{aligned} g_a(\mathbf{w}(t), \dot{\mathbf{w}}(t), \mathbf{p}(t), t) &\triangleq g(\mathbf{w}(t), \dot{\mathbf{w}}(t), t) \\ &\quad + \mathbf{p}^T(t)[\mathbf{f}(\mathbf{w}(t), \dot{\mathbf{w}}(t), t)]. \end{aligned} \tag{4.5-43}$$

This means that to determine the necessary conditions for an extremal we simply form the function g_a and write the Euler equations *as if* there were no constraints among the functions $\mathbf{w}$. Naturally, the constraining relations

$$\mathbf{f}(\mathbf{w}^*(t), \dot{\mathbf{w}}^*(t), t) = \mathbf{0} \tag*{†(4.5-41)}$$

must also be satisfied.

4.6 SUMMARY

In this chapter, some basic ideas of the calculus of variations have been introduced. The analogy between familiar results of the calculus and corresponding results in the calculus of variations has been established and

† If $\dot{\mathbf{w}}(t)$ does not appear explicitly in $\mathbf{f}$, then we have point constraints.

exploited. First, some basic definitions were stated, and used to prove the fundamental theorem of the calculus of variations. The fundamental theorem was then applied to determine necessary conditions to be satisfied by an extremal. Initially, the problems considered were assumed to have trajectories with fixed end points; subsequently, problems with free end points were considered. We found that regardless of the boundary conditions, the fundamental theorem yields a set of differential equations (the Euler equations) that are the same for a specified functional. Furthermore, we observed that the Euler equations are generally *nonlinear* differential equations with *split boundary values*; these two characteristics combine to make the solution of optimal control problems a challenging task.

In control problems the system trajectory is determined by the applied control—we say that the optimization problem is *constrained* by the dynamics of the process. In the concluding section of this chapter we considered constrained problems and introduced the method of Lagrange multipliers.

With this background material, we are at last ready to tackle "the optimal control problem."

REFERENCES

B-4 Bartle, R. G., *The Elements of Real Analysis.* New York: John Wiley & Sons, Inc., 1964.

E-1 Elsgolc, L. E., *Calculus of Variations.* Reading, Mass.: Addison-Wesley Publishing Company, Inc., 1962.

G-1 Gelfand, I. M., and S. V. Fomin, *Calculus of Variations.* Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1963.

H-1 Hildebrand, F. B., *Methods of Applied Mathematics,* 1st ed. Englewood Cliffs, N. J.: Prentice-Hall, Inc., 1952.

M-2 Menger, K., "What Is Calculus of Variations and What Are Its Applications?", *The World of Mathematics,* Vol. 2. New York: Simon and Schuster, Inc., 1956.

O-2 Olmstead, J. M. H., *Real Variables.* New York: Appleton-Century-Crofts, Inc., 1959.

PROBLEMS

4-1. f is a differentiable function of n variables defined on the domain $\mathscr{D}$. If $\mathbf{q}^*$ is an interior point of $\mathscr{D}$ and $f(\mathbf{q}^*)$ is a relative extremum, prove that the differential of f must be zero at the point $\mathbf{q}^*$.

4-2. Prove the fundamental lemma; that is, show that if $h(t)$ is continuous for $t \in [t_0, t_f]$, and if

$$\int_{t_0}^{t_f} h(t)\,\delta x(t)\,dt = 0$$

for every function $\delta x(t)$ that is continuous in the interval $[t_0, t_f]$ with $\delta x(t_0) = \delta x(t_f) = 0$, then $h(t)$ must be identically zero in the interval $[t_0, t_f]$.

4-3. Using the definition, find the differentials of the following functions:
(a) $f(t) = 4t^3 + 5/t,\ t > 0.$
(b) $f(q_1, q_2) = 5q_1^2, + 6q_1q_2 + 2q_2^2.$
(c) $f(\mathbf{q}) = q_1^2 + q_2^2 + 5q_1q_2q_3 + 2q_1q_2 + 3q_3.$
Compare your answers with the results obtained by using formal procedures for determining the differential.

4-4. Determine the variations of the functionals:
(a) $J(x) = \int_{t_0}^{t_f} [x^3(t) - x^2(t)\dot{x}(t)]\,dt.$
(b) $J(\mathbf{x}) = \int_{t_0}^{t_f} [x_1^2(t) + x_1(t)x_2(t) + x_2^2(t) + 2\dot{x}_1(t)\dot{x}_2(t)]\,dt.$
(c) $J(x) = \int_{t_0}^{t_f} \epsilon^{x(t)}\,dt.$
Assume that the end points are specified.

4-5. Consider *Problem 1* of Section 4.2 and let η be a specified continuously differentiable function that is arbitrary in the interval $[t_0, t_f]$ except at the end points, where $\eta(t_0) = \eta(t_f) = 0$. If ϵ is an arbitrary real parameter, then $x^* + \epsilon\eta$ represents a family of curves. Evaluating the functional

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\,dt$$

on the family $x^* + \epsilon\eta$ makes J a *function* of ϵ, and if x^* is an extremal this function must have a relative extremum at the point $\epsilon = 0$.

Show that the Euler equation (4.2-10) is obtained from the necessary condition

$$\left.\frac{dJ(x^* + \epsilon\eta)}{d\epsilon}\right|_{\epsilon=0} = 0.$$

4-6. Euler derived necessary conditions to be satisfied by an extremal using finite differences. The first step in the finite-difference approach to *Problem 1* of Section 4.2 is to approximate the functional

$$J(x) = \int_{t_0}^{t_f} g(x(t), \dot{x}(t), t)\,dt$$

by the summation

$$J_d \approx \sum_{k=0}^{N-1} g(x(k), \dot{x}(k), k)\,\Delta t,$$