CHAPTER 1

VARIATIONAL PROBLEMS WITH FIXED BOUNDARIES

1.1 The Concept of Variation and Its Properties

As already pointed out in the introduction, a variable quantity $I[y(x)]$ is a functional dependent on a function $y(x)$ if to each function $f(x)$ belonging to a certain class of functions $\mathbb{C}$, there is a definite value of I. Thus there is a correspondence between a given function $y(x)$ and a number I.

By the increment or variation δy of the argument $y(x)$ of a functional I, we mean the difference $\delta y (= y(x) - y_1(x))$ between two functions belonging to a certain class. A functional $I[y(x)]$ is said to be continuous if a small change in $y(x)$ results in a small change in $I[y(x)]$. This definition is, however, somewhat imprecise since we have not specified what we exactly mean by the phrase 'a small change in $y(x)$'. In other words, under what conditions should we consider the curves $y = y(x)$ and $y = y_1(x)$ close?

One way of specifying the closeness of $y(x)$ and $y_1(x)$ is to say that the absolute value of their difference given by $|y(x) - y_1(x)|$ is small for all x for which $y(x)$ and $y_1(x)$ are defined. When this happens, we say $y(x)$ is close to $y_1(x)$ in the sense of zero-order proximity. But with this definition, the functional

$$I[y(x)] = \int_a^b F(x, y(x), y'(x))\, dx, \tag{1.1}$$

which occurs in many applications, is seldom continuous due to the presence of the argument $y'(x)$. This necessitates the extension of the notion of closeness of the curves $y = y(x)$ and $y = y_1(x)$ such that both $|y(x) - y_1(x)|$ and $|y'(x) - y_1'(x)|$ are small for all values of x for which these functions are prescribed. We then say that these two curves are close in the sense of first-order proximity. In general, the curves $y = y(x)$ and $y = y_1(x)$ are said to be close in the sense of nth order proximity if $|y(x) - y_1(x)|$, $|y'(x) - y_1'(x)|, \ldots, |y^{(n)}(x) - y_1^{(n)}(x)|$ are small for values of x for which these functions are defined.

Figure 1.1 shows two curves which are close in the sense of zero-order proximity but not in the sense of first-order proximity. Figure 1.2 shows two curves which are close in the sense of first-order proximity. It is clear from the above definitions that if two curves are close in the sense of nth order proximity, then they are certainly, close in the sense of any lower order (say, $(n-1)$th) proximity.

We are now in a position to refine the concept of the continuity of a functional. The functional $I[y(x)]$ is said to be continuous at $y = y_0(x)$, in the sense of nth

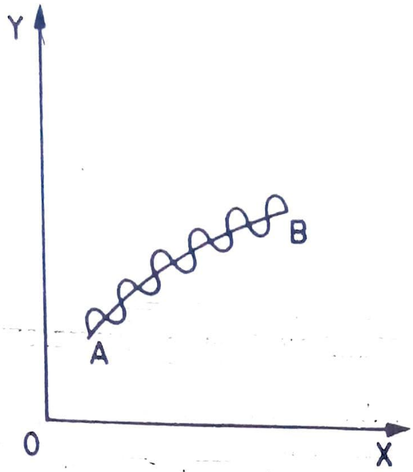

Fig. 1.1 Curves close in the sense of zero-order proximity.

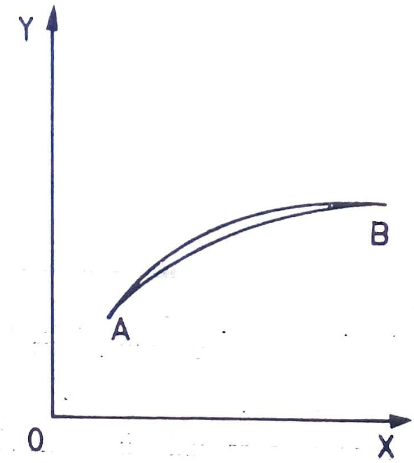

Fig. 1.2 Curves close in the sense of first-order proximity.

order proximity, if given any positive number ε, there exists a $\delta > 0$ such that

$$|I[y(x)] - I[y_0(x)]| < \varepsilon \text{ for } |y(x) - y_0(x)| < \delta$$

$$|y'(x) - y_0'(x)| < \delta, \ldots, |y^{(n)}(x) - y_0^{(n)}(x)| < \delta$$

Example 1. Show that the functional

$$I[y(x)] = \int_0^1 x^3[1 + y^2(x)]^{1/2}\,dx$$

defined on the set of functions $y(x) \in C[0, 1]$, (where $C[0, 1]$ is the set of all continuous functions on the closed interval $0 \le x \le 1$) is continuous on the function $y_0(x) = x^2$ in the sense of zero-order proximity.

Solution. Put $y(x) = x^2 + \alpha\eta(x)$, where $\eta(x) \in C[0, 1]$ and α is arbitrarily small. Then,

$$I[y(x)] = I[x^2 + \alpha\eta(x)] = \int_0^1 x^3\,[1 + (x^2 + \alpha\eta(x))^2]^{1/2}\,dx.$$

Passing to the limit $\alpha \to 0$, we find that,

$$\lim_{\alpha \to 0} I[y(x)] = \int_0^1 x^3(1 + x^4)^{1/2}\,dx = I[x^2]$$

and this establishes the continuity of the functional on $y_0(x) = x^2$.

It is, however, possible to define the notion of distance $\rho(y_1, y_2)$ between two curves $y = y_1(x)$ and $y = y_2(x)$ (with $x_0 \le x \le x_1$) as

$$\rho(y_1; y_2) = \max_{(x_0 \le x \le x_1)} |y_1(x) - y_2(x)| \tag{1.2}$$

Clearly, with this metric, we can introduce the concept of zero-order proximity. This notion can be extended to the case of nth order proximity of two curves $y = y_1(x)$ and $y = y_2(x)$ (admitting continuous derivatives upto order n inclusive) if one introduces the metric

$$\rho(y_1, y_2) = \sum_{p=1}^{n} \max_{(x_0 \le x \le x_1)} | y_1^{(p)}(x) - y_2^{(p)}(x) | \tag{1.3}$$

Let us now introduce the concept of a linear functional $I[y(x)]$ defined in the normed linear space M of the functions $y(x)$. This functional is said to be linear, if it satisfies

(i) $I[cy(x)] = cI[y(x)]$,

where c is an arbitrary constant,

(ii) $I[y_1(x) + y_2(x)] = I[y_1(x)] + I[y_2(x)]$,

where $y_1(x) \in M$ and $y_2(x) \in M$.

Take, for instance, the functional

$$I[y(x)] = \int_a^b [y'(x) + 2y(x)]\, dx \tag{1.4}$$

defined in the space $C^1[a, b]$, which consists of the set of all functions admitting continuous first order derivatives in $[a, b]$. Clearly, I in (1.4) is a linear functional.

It can, however, be shown that a functional $I[y(x)]$ is linear if (a) it is continuous and, (b) for any $y_1(x) \in M$ and $y_2(x) \in M$, satisfies the condition

$$I[y_1(x)] + I[y_2(x)] = I[y_1(x) + y_2(x)]$$

Let us now define the variation of a functional $I[y(x)]$. The increment ΔI is given by

$$\Delta I = I[y(x) + \delta y(x)] - I[y(x)]$$

which may be written in the form

$$\Delta I = L[y(x), \delta y] + \beta[y(x), \delta y] \max | \delta y | \tag{1.5}$$

Here, $L[y(x), \delta y]$ is a functional linear in δy and $\beta[y(x), \delta y] \to 0$ as the maximum value of δy (given by $\max |\delta y|$) $\to 0$. This sort of division of the increment ΔI is analogous to the differential, and the infinitesimal, in the case of a function of a single variable given by

$$\Delta f(x) = f(x + \Delta x) - f(x)$$
$$= A(x)\, \Delta x + \beta(x, \Delta x)\, \Delta x. \tag{1.6}$$

Here, $A(x)\, \Delta x$, known as the differential df, is the principal part of the increment and is linear in Δx. By the same token, the part $L[y(x), \delta y]$ is called the variation of the functional and is denoted by δI.

An alternative definition of the variation δI of a functional I can be given. Consider the functional $I[y(x) + \alpha \delta y]$ for fixed y and δy and different values of the parameter α.

Now using (1.5) the increment ΔI can be written as

$$\Delta I = I[y(x) + \alpha \delta y] - I[y(x)]$$
$$= L[y, \alpha \delta y] + \beta[y, \alpha \delta y]\, |\alpha| \max | \delta y |.$$

The derivative of $I[y(x) + \alpha\, \delta y]$ with respect to α at $\alpha = 0$ is

$$\lim_{\Delta\alpha \to 0} \frac{\Delta I}{\Delta\alpha} = \lim_{\alpha \to 0} \frac{\Delta I}{\alpha} = \lim_{\alpha \to 0} \frac{L[y, \alpha\delta y] + \beta[y, \alpha\delta y]\,|\alpha|\max|\delta y|}{\alpha}$$

$$= \lim_{\alpha \to 0} \frac{L[y, \alpha\delta y]}{\alpha} + \lim_{\alpha \to 0} \frac{\beta[y, \alpha\delta y]\,|\alpha|\max|\delta y|}{\alpha}$$

$$= L[y, \delta y] = \delta I,$$

since by linearity $L[y, \alpha\delta y] = \alpha L[y, \delta y]$ and $\beta \to 0$ as $\alpha \to 0$. Hence the variation of a functional $I[y(x)]$ is equal to

$$\frac{\partial}{\partial\alpha} I[y(x) + \alpha\delta y] \text{ at } \alpha = 0.$$

Definition. A functional $I[y(x)]$ attains a maximum on a curve $y = y_0(x)$, if the values of I on any curve close to $y = y_0(x)$ do not exceed $I[y_0(x)]$. This means that $\Delta I = I[y(x)] - I[y_0(x)] \le 0$. Further, if $\Delta I \le 0$ and $\Delta I = 0$ only on $y = y_0(x)$, we say that a strict maximum is attained on $y = y_0(x)$. In the case of a minimum of I on $y = y_0(x)$, $\Delta I \ge 0$ for all curves close to $y_0(x)$ and a strict minimum is defined in the same way.

Theorem. If a functional $I[y(x)]$ attains a maximum or minimum on $y = y_0(x)$, where the domain of definition belongs to certain class, then at $y = y_0(x)$,

$$\delta I = 0. \qquad (1.7)$$

Proof. For fixed $y_0(x)$ and δy, $I[y_0(x) + \alpha\delta y)] = \Psi(\alpha)$ is a function of α and this reaches a maximum or minimum for $\alpha = 0$. Thus $\Psi'(0) = 0$ leading to $\frac{\partial}{\partial\alpha} I[y_0(x) + \alpha\delta y]|_{\alpha=0} = 0$, i.e., $\delta I = 0$. This proves the theorem.

However, when we talk of maximum or minimum, we mean the largest or smallest value of the functional, relative to values of the functional on close-lying curves. But we have already seen that the closeness of curves may be understood in different ways depending on the order of proximity of the curves.

If a functional $I[y(x)]$ attains a maximum or minimum on the curve $y = y_0(x)$ with respect to all curves $y = y(x)$ such that $|y(x) - y_0(x)|$ is small, then the maximum or minimum is said to be strong.

If, on the other hand, $I[y(x)]$ attains a maximum or minimum on the curve $y = y_0(x)$ with respect to all curves $y = y(x)$ in the sense of first order proximity, i.e., $|y(x) - y_0(x)|$ and $|y'(x) - y_0'(x)|$ are both small, then the maximum or minimum is said to be weak. It is quite clear that if a strong maximum (or minimum) of a functional $I[y(x)]$ is attained on the curve $y = y_0(x)$, then a weak maximum (or minimum) is also attained on the same curve. This follows from the fact that if two curves are close in the sense of first-order proximity, then they are definitely close in the sense of zero-order proximity as well.

This theorem can be readily extended to functionals dependent on several unknown functions, or dependent on one or several functions of any number of variables, e.g.,

$$I[y_1(x),\ y_2(x),\ \ldots,\ y_n(x)] \quad \text{or} \quad I[z(x_1,\ x_2,\ \ldots,\ x_m)]$$

or

$$I[z_1(x_1,\ x_2,\ \ldots,\ x_m),\ z_2(x_1,\ x_2,\ \ldots,\ x_m),\ \ldots,\ z_p(x_1,\ x_2,\ \ldots,\ x_m)].$$

The necessary condition for extremum, in all these cases, is still, given by $\delta I = 0$, where the variation δ is defined in exactly the same way as that for a functional $I[y(x)]$.

1.2 Euler's Equation

Let us examine the extremum of the functional

$$I[y(x)] = \int_a^b F(x,\ y(x),\ y'(x))\,dx \tag{1.8}$$

subject to the boundary conditions $y(a) = y_1$ and $y(b) = y_2$, where y_1 and y_2 are prescribed at the fixed boundary points a and b. We assume that $F(x, y, y')$ is three times differentiable. We have already shown that the necessary condition for an extremum of a functional is that its variation must vanish. We shall now apply this condition to (1.8), and assume that the admissible curves on which an extremum is achieved, admits of continuous first-order derivatives. It can be proved, however, that the curve on which an extremum is achieved, admits of a continuous second-order derivative also (see Section 1.9).

Let $y = y(x)$ be the curve which extremizes the functional (1.8) such that $y(x)$ is twice differentiable and satisfies the above boundary conditions (see Fig. 1.3). Let $y = \bar{y}(x)$ be an admissible curve close to $y = y(x)$ such that both $y(x)$ and $\bar{y}(x)$ can be included in a one-parameter family of curves

$$y(x,\ \alpha) = y(x) + \alpha[\bar{y}(x) - y(x)] \tag{1.9}$$

For $\alpha = 0$, $y(x, \alpha) = y(x)$ and for $\alpha = 1$, $y(x, \alpha) = \bar{y}(x)$.

The difference $\bar{y}(x) - y(x)$ is the variation δy of the function y (see Fig. 1.4) and is similar to the role played by Δx, the increment in x while considering the extrema of a function $f(x)$. Now on the curves of the family (1.9), the functional (1.8) reduces to a function of α, say $\Psi(\alpha)$. Since the extremizing curve $y = y(x)$ corresponds to $\alpha = 0$, it follows that $\Psi(\alpha)$ is extremized for $\alpha = 0$. This implies that

$$\left(\frac{d\Psi}{d\alpha}\right)_{\alpha = 0} = 0, \tag{1.10}$$

where

$$\Psi(\alpha) = \int_a^b F(x,\ y(x,\ \alpha),\ y'(x,\ \alpha))\,dx \tag{1.11}$$

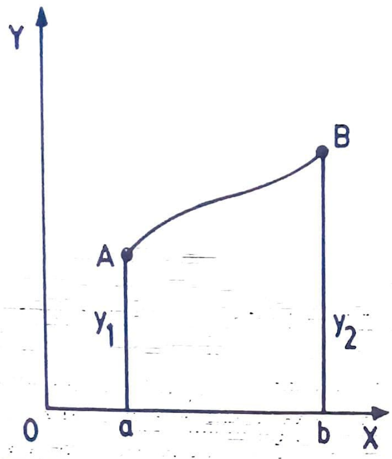

Fig. 1.3 Extremizing curve joining two fixed points.

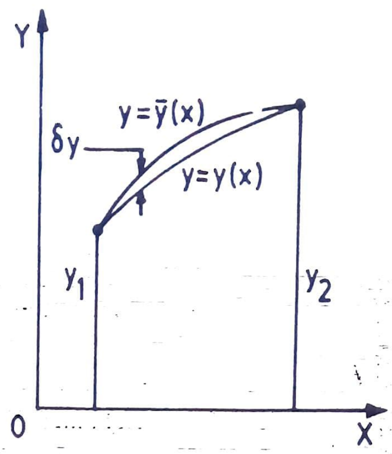

Fig. 1.4 Extremizing curve and an admissible curve between two fixed points.

Using (1.9) and (1.11), it follows that

$$\Psi'(\alpha) = \int_a^b [F_y(x, y(x, \alpha), y'(x, \alpha))\, \delta y + F_{y'}(x, y(x, \alpha), y'(x, \alpha))\delta y'] \, dx \tag{1.12}$$

where a subscript denotes partial derivative with respect to the indicated variable.

Further, the variation $\delta y (= \bar{y}(x) - y(x))$ is a function of x and can be differentiated once, or several times, such that $(\delta y)' = \bar{y}'(x) - y'(x) = \delta y'$. Finally, (1.10) gives from (1.12) the relation

$$\int_a^b [F_y(x, y(x), y'(x))\delta y + F_{y'}(x, y(x), y'(x))\delta y'] \, dx = 0 \tag{1.13}$$

Let us integrate the second term by parts subject to the boundary conditions $(\delta y)_a = 0$ and $(\delta y)_b = 0$ (as a consequence of y being fixed at $x = a$ and $x = b$). This gives from (1.13),

$$\int_a^b \left[F_y - \frac{d}{dx} F_{y'} \right] \delta y \, dx = 0 \tag{1.14}$$

In view of the assumptions made on $F(x, y(x), y'(x))$ and the extremizing curve $y(x)$, it follows that $F_y - \frac{d}{dx} F_{y'}$ on the curve $y(x)$ is a given continuous function, while δy is an arbitrary continuous function, subject to the vanishing of δy at $x = a$ and $x = b$.

Before proceeding further, we now prove the following lemma: If for every continuous function $\eta(x)$,

$$\int_a^b \Phi(x)\eta(x) \, dx = 0 \tag{1.15}$$

where $\Phi(x)$ is continuous in the closed interval $[a, b]$, then $\Phi(x) \equiv 0$ on $[a, b]$.

Proof. Assume that $\Phi(x) \neq 0$ (positive, say) at a point $x = \bar{x}$ in $a \leq x \leq b$. By virtue of the continuity of $\Phi(x)$, it follows that $\Phi(x) \neq 0$ and maintains positive sign in a small neighbourhood $x_0 \leq x \leq x_1$ of the point $\bar{x}$. Since $\eta(x)$ is an arbitrary continuous function, we might choose $\eta(x)$ such that $\eta(x)$ remains positive in $x_0 \leq x \leq x_1$ but vanishes outside this interval (see Fig. 1.5). It then follows from (1.15) that

$$\int_a^b \Phi(x)\eta(x)\,dx = \int_{x_0}^{x_1} \Phi(x)\eta(x)\,dx > 0 \tag{1.16}$$

since the product $\Phi(x)\eta(x)$ remains positive everywhere in $[x_0, x_1]$. The contradiction between (1.15) and (1.16) shows that our original assumption $\Phi(x) \neq 0$ at some point $\bar{x}$ must be wrong and hence $\Phi(x) \equiv 0$ on $[a, b]$.

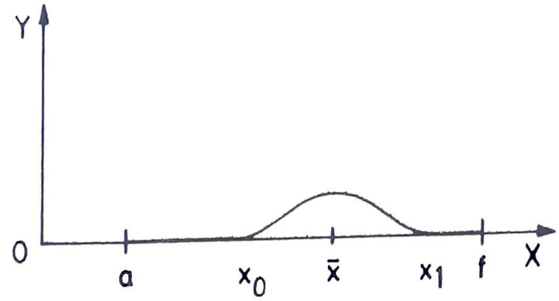

Fig. 1.5 A continuous function which is positive in an interval but vanishes outside.

Invoking this fundamental lemma, and from (1.14) we conclude, that

$$F_y - \frac{d}{dx} F_{y'} = 0 \tag{1.17}$$

on the extremizing curve $y = y(x)$. This equation is known as Euler's equation and the integral curves of this equation are known as extremals. It should be noted that the functional (1.8) can attain an extremum only on extremals. On expanding (1.17) we find that

$$F_y - F_{xy'} - F_{yy'}y' - F_{y'y'}y'' = 0 \tag{1.18}$$

which is, in general, a second-order differential equation in $y(x)$ (although sometimes it may reduce to a finite equation). The two arbitrary constants appearing in the solution $y(x)$ are determined from the boundary conditions $y(a) = y_1$ and $y(b) = y_2$.

It should be emphasized, however, that the existence of the solution of (1.17) satisfying the above boundary conditions cannot always be taken for granted, and even if a solution exists, it may not be unique. However, in many problems, the existence of a solution is evident from the geometrical or physical significance of the problem. Hence in such cases, if the existence of solution of Euler's equation is unique, then this solution will provide the solution of the variational problem.

Example 2. Test for an extremum the functional

$$I[y(x)] = \int_0^1 (xy + y^2 - 2y^2 y')\,dx, \qquad y(0) = 1,\ y(1) = 2.$$

Solution. Here Euler's equation is

$$x + 2y - 4yy' - \frac{d}{dx}(-2y^2) = 0$$

which reduces to $y = -x/2$. Clearly, this extremal cannot satisfy the boundary conditions $y(0) = 1$, $y(1) = 2$. Thus an extremum cannot be achieved in the class of continuous functions.

Example 3. Test for extremum the functional

$$I[y(x)] = \int_0^{\pi/2} (y'^2 - y^2)\,dx, \qquad y(0) = 0, \qquad y(\pi/2) = 1$$

Solution. In this case Euler's equation is

$$-y - \frac{d}{dx}(y') = 0$$

and its general solution is $y = C_1 \cos x + C_2 \sin x$. Using the boundary conditions, we find that $C_1 = 0$, $C_2 = 1$. Thus the extremum can be achieved only on the curve $y = \sin x$.

In the problems cited above, Euler's equation is readily integrable. But this is not always possible. In what follows that we consider some cases, where Euler's equation admits of integration.

(i) In this case F in (1.8) is a function of x and y only. Then the Euler equation reduces to $F_y(x, y) = 0$. This finite equation, when solved for y, does not involve any arbitrary constant. Thus, in general, it is not possible to find y satisfying the boundary conditions $y(a) = y_1$, and $y(b) = y_2$ and as such this variational problem does not, in general, admit of a solution. Example 2 cited above is an illustration of such a problem.

(ii) F in (1.8) depends only on x and y'. Here Euler's equation becomes

$$\frac{d}{dx} F_{y'}(x, y') = 0 \tag{1.19}$$

which has an integral $F_{y'}(x, y') = C_1$, a constant. Since this relation does not contain y, it can be solved for y' as a function of x. Another integration leads to a solution involving two arbitrary constants which can be found from the boundary conditions.

Example 4. Find the extremum of the function

$$I[y(x)] = \int_{x_0}^{x_1} \frac{(1 + y'^2)^{1/2}}{x}\,dx$$

Solution. Before we embark on the solution, it may be noticed that the functional t may be recognized as the time spent on translation along the curve $y = y(x)$ from one point to another, if the rate of motion $v = (ds/dt)$ is equal to x. This is due to the fact that $ds = (1 + y'^2)^{1/2}\, dx$.

Since the functional is independent of y, Euler's equation leads to

$$y' = C_1 x(1 + y'^2)^{1/2}. \tag{1.20}$$

This may be integrated by introducing $y' = \tan t$, t being a parameter. Then (1.20) gives $x = (1/C_1) \sin t = \bar{C}_1 \sin t$. Then

$$dy = \tan t\, dx = \bar{C}_1 \sin t\, dt, \tag{1.21}$$

which on integration leads to

$$y = -\bar{C}_1 \cos t + C_2$$

Elimination of t from the expressions for x and y then gives the extremals as

$$x^2 + (y - C_2)^2 = \bar{C}_1^2,$$

which is a family of circles.

(iii) F in (1.8) is dependent on y and y' only. In this case Euler's equation reduces to

$$F_y - F_{yy'}y' - F_{y'y'}y'' = 0 \tag{1.22}$$

But

$$\frac{d}{dx}(F - y'F_{y'}) = F_y y' + F_{y'}y'' - y''F_{y'} - F_{yy'}y'^2 - F_{y'y'}y''y'$$

$$= y'(F_y - F_{yy'}y' - F_{y'y'}y'')$$

Thus by virtue of (1.22), Euler's equation has the first integral

$$F - y'F_{y'} = C_1 \tag{1.23}$$

where C_1 is a constant. This equation may be integrated further after solving for y' and separation of variables.

Example 5. Find the curve joining given points A and B which is traversed by a particle moving under gravity from A to B in the shortest time (ignore friction along the curve and the resistance of the medium). This is known as the Brachistochrone problem to which we have alluded before.

Solution. Fix the origin at A with x-axis horizontal and y-axis vertically downward. The speed of the particle ds/dt is given by $(2gy)^{1/2}$, g being the acceleration due to gravity. Thus the time taken by the particle in moving from $A(0, 0)$ to $B(x_1, y_1)$ is

$$t[y(x)] = \frac{1}{\sqrt{2g}} \int_0^{x_1} \frac{\sqrt{1 + y'^2}}{\sqrt{y}}\, dx;$$

$$y(0) = 0, \quad y(x_1) = y_1 \tag{1.24}$$

Although the integral is improper, it is convergent. Since the integrand is independent of x, a first integral of Euler's equation is given by (1.23), which gives, on simplification, the relation

$$y(1 + y'^2) = C_1 \tag{1.25}$$

Put $y' = \cot t$, t being a parameter. Then (1.25) gives

$$y = \frac{C_1}{2}(1 - \cos 2t) \tag{1.26}$$

Now,

$$dx = \frac{dy}{y'} = \frac{2C_1 \sin t \cos t \, dt}{\cot t} = C_1(1 - \cos 2t)\, dt$$

which gives, on integration, the equation

$$x - C_2 = \frac{C_1}{2}(2t - \sin 2t) \tag{1.27}$$

Putting $2t = t_1$ and remembering that $y = 0$ at $x = 0$, we find that $C_2 = 0$. Thus (1.26) and (1.27) give the desired extremals in the parametric form

$$x = \frac{C_1}{2}(t_1 - \sin t_1), \quad y = \frac{C_1}{2}(1 - \cos t_1)$$

which is a family of cycloids with $C_1/2$ as the radius of the rolling circle. In fact, C_1 is determined by the fact that the cycloid passes through $B(x_1, y_1)$.

Example 6. Find the curvewith fixed boundary points such that its rotation about the axis of abscissae give rise to a surface of revolution of minimum surface area.

Solution. The area of the surface of revolution (Fig. 1.6) is

$$S[y(x)] = 2\pi \int_{x_1}^{x_2} y\sqrt{1 + y'^2}\, dx$$

where the end points A and B of the curve $y = y(x)$ have x-coordinates x_1 and x_2. Since the integrand is a function of y and y' only, a first integral of Euler's equation is

$$y\sqrt{1 + y'^2} - \frac{yy'^2}{\sqrt{1 + y'^2}} = C_1$$

which reduces to $y/\sqrt{1 + y'^2} = C_1$. To integrate this equation, we put $y' = \sinh t$. Then, clearly,

$$y = C_1 \cosh t, \quad dx = \frac{dy}{y'} = C_1\, dt \tag{1.28}$$

The second equation of (1.28) gives on integration the relation

$$x = C_1 t + C_2 \text{ with } y = C_1 \cosh t \tag{1.29}$$

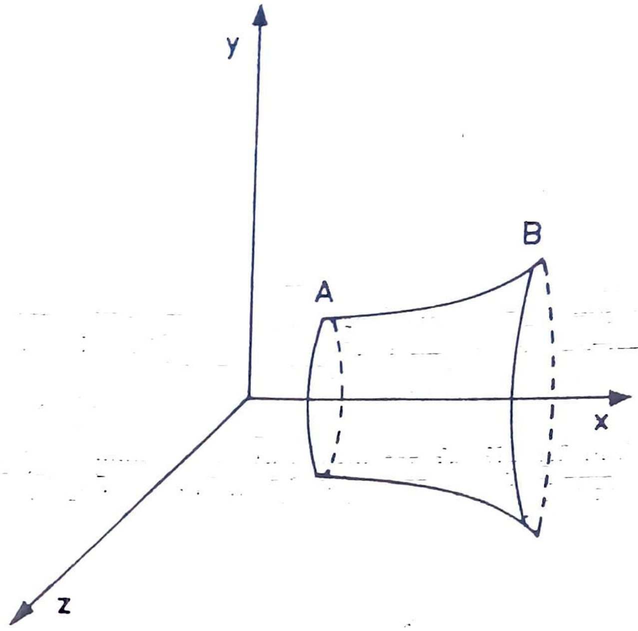

Fig. 1.6 Surface of revolution with minimum surface area.

The elimination of t from (1.29) gives as extremals

$$y = C_1 \cosh \frac{x - C_2}{C_1}$$

which constitutes a two-parameter family of catenaries. The constants C_1 and C_2 are determined from the conditions, that the given curve passes through the given points A and B.

As a last example of the extremum of a functional, we consider the following problem of gas dynamics.

Example 7. To determine the shape of a solid of revolution moving in a flow of gas with least resistance.

Solution. Referring to Fig. 1.7, assume that the gas density is sufficiently small such that the gas molecules are mirror reflected from the surface of the solid. The component of the gas pressure normal to the surface is

$$p = 2\rho v^2 \sin^2 \theta \tag{1.30}$$

where ρ, v and θ denote the density of the gas, the velocity of the gas relative to the solid, and the angle between the tangent at any point of the surface with the direction of flow.

The pressure given by (1.30) is normal to the surface and one can write down the force component along the x-axis acting on a ring PQ of width ds ($= \sqrt{1 + y'^2}\, dx$) and radius $y(x)$ in the form

$$dF = 2\rho v^2 \sin^2 \theta \cdot [2\pi y\sqrt{1 + y'^2}] \sin \theta \, dx \tag{1.31}$$

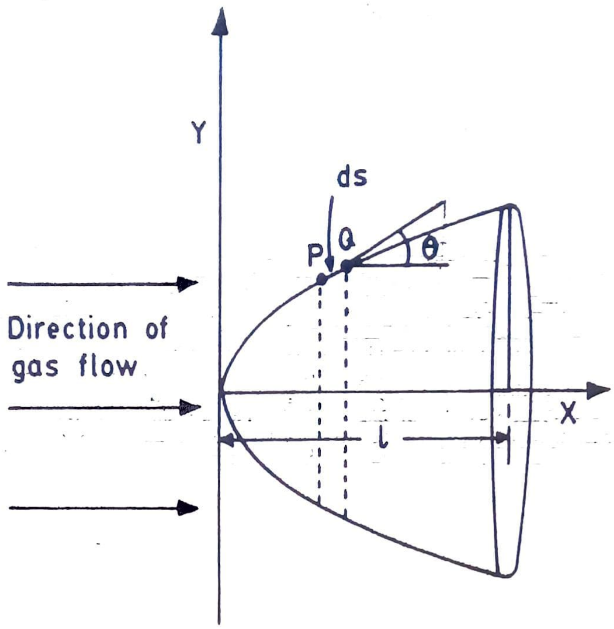

Fig. 1.7 Solid of revolution experiencing least resistance in a gas flow.

Hence the total force along the x-direction is

$$F = \int_0^l 4\pi\rho v^2 \sin^3\theta \cdot \sqrt{1 + y'^2}\, y\, dx \tag{1.32}$$

To make further progress, we assume

$$\sin\theta = \frac{y'}{(1 + y'^2)^{3/2}} \approx y'$$

where the slope y' is taken to be small. Thus from (1.32), the total resistance experienced by the body is

$$F = 4\pi\rho v^2 \int_0^l y'^3 y\, dx \tag{1.33}$$

The problem now is to find $y = y(x)$ for which F is minimum. Thus (1.33) constitutes a variational problem with the boundary conditions

$$y(0) = 0, \qquad y(l) = R \tag{1.34}$$

Since the integrand in (1.33) depends on y and y' only, a first integral of Euler's equation is

$$y'^3 - 3\frac{d}{dx}(yy'^2) = 0 \tag{1.35}$$

Multiplying (1.35) by y' and integrating, we get

$$y'^3 y = C_1^3,$$

C_1 being a constant. One more integration gives

$$y = (C_2 x + C_3)^{3/4} \tag{1.36}$$

Using the boundary conditions (1.34), we obtain

$$C_2 = \frac{R^{4/3}}{l}, \qquad C_3 = 0$$

Thus the required function $y(x)$ is given by

$$y(x) = R\left(\frac{x}{l}\right)^{3/4}$$

If F in (1.8) is linear in y' such that

$$F(x, y, y') = M(x, y) + N(x; y)y'$$

then Euler's equation reduces to

$$\frac{\partial M}{\partial y} - \frac{\partial N}{\partial x} = 0,$$

which is a finite equation, and not a differential equation. Thus the curve defined by the above equation does not, in general, satisfy the boundary conditions at $x = a$ and b. Clearly, in this case the variational problem (1.8) does not have (in general) a solution in the class of continuous functions. The reason for this lies in the fact, that, when the above equation holds in some domain of the xy-plane, then the integral

$$I[y(x)] = \int_a^b F(x, y, y')\,dx = \int_a^b (M\,dx + N\,dy)$$

becomes independent of the path of integration. Thus the functional is the same on all admissible curves leading to a meaningless variational problem.

1.3 Variational Problem for Functionals of the Form

$$\int_a^b F(x, y_1(x), y_2(x), \ldots, y_n(x), y_1'(x), y_2'(x), \ldots, y_n'(x))\,dx,$$

where the function F is differentiable three times with respect to all its arguments.

To find the necessary conditions for the extremum of the above functional, we consider the following boundary conditions for $y_1(x), y_2(x), \ldots, y_n(x)$:

$$y_1(a) = Y_1,\ y_2(a) = Y_2,\ \ldots,\ y_n(a) = Y_n \tag{1.37a}$$

$$y_1(b) = Z_1,\ y_2(b) = Z_2,\ \ldots,\ y_n(b) = Z_n \tag{1.37b}$$

where $Y_1, Y_2, \ldots, Z_1, Z_2 \ldots$ are constants.

We vary only one of the functions $y_j(x)$ $(j = 1, 2, \ldots, n)$, keeping the others fixed. Then the above functional reduces to a functional dependent on, say, only one of the functions $y_i(x)$. Thus the function $y_i(x)$ having a continuous derivative must satisfy Euler's equation

$$F_{y_i} - \frac{d}{dx} F_{y_i'} = 0$$

where the boundary conditions on $y_i(x)$ at $x = a$ and $x = b$ are utilized from (1.37a) and (1.37b).

Since this argument applies to any function $y_i(x)$ ($i = 1, 2, \ldots, n$), we obtain a system of second-order differential equations

$$F_{y_i} - \frac{d}{dx} F_{y_i'} = 0 \qquad (i = 1, 2, \ldots, n). \tag{1.38a}$$

These define, in general, a $2n$-parameter family of curves in the space $x, y_1, y_2, \ldots, y_n$ and provide the family of extremals for the given variational problem.

Let us illustrate the above principle by considering a problem from optics.

***Example* 8.** Derive the differential equations of the lines of propagation of light in an optically non-homogeneous medium with the speed of light $C(x, y, z)$.

***Solution*.** According to well known Fermat's law, light propagates from one point to another point along a curve, for which, the time T of passage of light will be minimum.

If the equation of the desired path of the light ray be $y = y(x)$ and $z = z(x)$, then clearly,

$$T = \int_{x_1}^{x_2} \frac{ds}{C} = \int_{x_1}^{x_2} \frac{\sqrt{1 + y'^2 + z'^2}}{C(x, y, z)} dx \tag{1.38b}$$

where ds is a line element on the path.

Using (1.5), one gets the system of Euler's equations

$$\frac{\sqrt{1 + y'^2 + z'^2}}{C^2} \frac{\partial C}{\partial y} + \frac{d}{dx} \left[\frac{y'}{C\sqrt{1 + y'^2 + z'^2}} \right] = 0$$

$$\frac{\sqrt{1 + y'^2 + z'^2}}{C^2} \frac{\partial C}{\partial z} + \frac{d}{dx} \left[\frac{z'}{C\sqrt{1 + y'^2 + z'^2}} \right] = 0$$

which determine the path of the light propagation.

It should be noted, however, that in the above form, the principle cannot always be applied. Let P_1 be the centre of a hemispherical mirror. The length of the path of the ray emerging from P_1 and reflected by the mirror at its pole p to a point P_2 on the straight line pP_1 will be longer than the path P_1QP_2, consisting of two rectilinear segments QP_2 and P_1Q, corresponding to a reflection by the mirror at a point Q distinct from p. This difficulty can be circumvented by removing from the formulation the specific mention of fixed end points. A better formulation is as follows: A curve can represent the path of a ray of light if and only if, each point P on Γ, is an interior point of a segment P_1P_2 of Γ which possesses the property that the integral (1.38b) for T taken along the segment P_1P_2 of Γ has a smaller value than that taken along any other curve of light from a point source $P_1(t_1, x_1, y_1, z_1)$. After a given time T_0, such a disturbance will be seen on a surface $F(T_0)$ which, according to Fermat's principle, is such that each point $P_2(t_2, x_2, y_2, z_2)$ is joined to P_1 by an extremal for which the integral (1.38b) takes the value T_0, this value being common to all points of $F(T_0)$. The surface $F(T_0)$ is a wave front and for various values of T_0, a succession or family of such wave fronts is obtained.

One can show that the family of wave fronts corresponding to the emission from a point source at P_1 is identical to the family of concentric geodesic spheres centred at P_1, a problem of the calculus of variations and determined by the integral (1.38b).

Remark 1. Certain interesting results follow if we consider the problem of propagation of a light ray in an inhomogeneous two-dimensional medium with the velocity of light, proportional to y (see Fig. 1.8). In this case the light rays are the extremals of the functional

$$I[y(x)] = \int_a^b \frac{(1 + y'^2)^{1/2}}{y}\, dx$$

Here the integral of Euler's equation gives $y(1 + y'^2)^{1/2} = \bar{C}_1$, whose integration leads to

$$(x + C_2)^2 + y^2 = \bar{C}_1^2$$

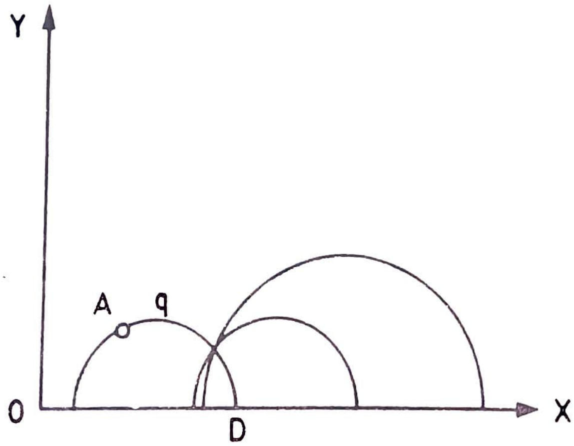

Fig. 1.8 Path of light ray propagation in an inhomogeneous medium.

This is a family of circles centred on the x-axis. The desired extremal is the one which passes through given points. This problem has a unique solution, since only one semicircle centred on the x-axis, passes through any two points lying in the upper half plane.

Consider the curve q. The optical path length q is the time $T(q)$, during which the curve is traversed with velocity of light $v(x, y) = y$. It may be shown that one end of the part AD of the semicircle q, which lies on the x-axis, has an infinite optical path length. Hence we call the points on the x-axis as infinite points. We shall consider the semicircles with centres on the x-axis to be straight lines, and the optical path lengths of the arcs of such semicircles, to be their lengths, and the angles between the tangents to the semicircles at their intersections to be the angles between such straight lines. Thus we derive a flat geometry in which many of the postulates of ordinary geometry remain valid. For example, only one straight line can be drawn through two points (only one semi-circle centred on the x-axis can be drawn through two points on the semicircle). Two straight lines will be deemed as parallel if they have a common point at infinity (i.e., two semicircles touch each other at a certain point B lying on the x-axis as shown in Fig. 1.9). Further it is possible to draw through a given point A, not lying on the straight line q, two straight lines q_1 and q_2 parallel to q.

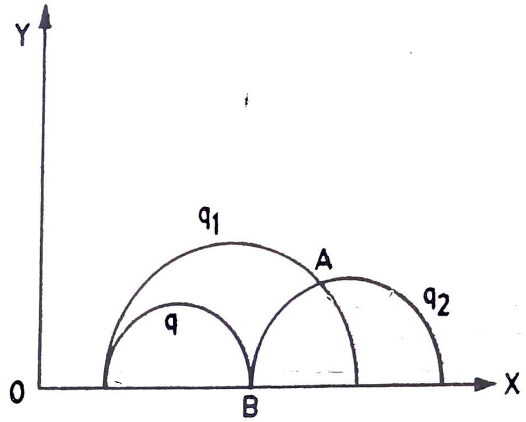

Fig. 1.9 Optical path in an inhomogeneous medium.

We have thus obtained an interesting new geometry, which is called the *Poincare model of Lobachevskian geometry* in the plane.

Remark 2. The foregoing remarks at once lead to the question of the possibility of drawing an extremal through just any two points with distinct abscissae. An answer to this question can sometimes be found from the following theorem due to Bernstein [14] (proof omitted):

Consider the equation

$$y'' = F(x, y, y'). \tag{1.38c}$$

If F, F_y and $F_{y'}$ are continuous at each end point (x, y) for every finite y' and if there exist a constant $k > 0$ and functions

$$\alpha = \alpha(x, y) \geq 0, \qquad \beta = \beta(x, y) \geq 0$$

bounded in every finite portion of the plane such that

$$F_y(x, y, y') > k, \qquad |F(x, y, y')| \leq \alpha y'^2 + \beta,$$

then one and only one integral curve $y = \phi(x)$ of (1.38c), passes through any two points (a, A) and (b, B) of the plane, with distinct abscissae $(a \neq b)$.

Consider, for example, the functional

$$I = \int e^{-2y^2} (y'^2 - 1)\, dx$$

Its Euler equation is

$$y'' = 2y(1 + y'^2)$$

Since $F(x, y, y') = 2y(1 + y'^2)$, we have

$$F_y = 2(1 + y'^2) \geq 2 = k \text{ and, further,}$$

$$|F(x, y, y')| = |2y(1 + y'^2)| \leq 2\,|y|\,y'^2 + 2\,|y|$$

so that $\alpha = \beta = 2\,|y| \geq 0$. Hence by Bernstein's theorem, there exists an extremal through any two points with distinct abscissae.

On the other hand, it can be shown that it is not possible, to draw an extremal of the functional

$$I[y(x)] = \int [y^2 + \sqrt{1 + y'^2}\,]\, dx$$

through just any two points of a plane having distinct abscissae.

1.4 Functionals Dependent on Higher-Order Derivatives

Let us now consider the extremum of a functional of the form

$$\int_a^b F(x, y(x), y'(x), \ldots, y^{(n)}(x))\, dx, \tag{1.39}$$

where we assume F to be differentiable $n + 2$ times with respect to all its arguments. The boundary conditions are taken in the form

$$y(x_0) = y_0,\ y'(x_0) = y_0',\ \ldots,\ y^{(n-1)}(x_0) = y_0^{(n-1)}, \tag{1.40a}$$

$$y(x_1) = y_1,\ y'(x_1) = y_1',\ \ldots,\ y^{(n-1)}(x_1) = y_1^{(n-1)}. \tag{1.40b}$$

This implies that at the boundary points the values of y together with all their derivatives upto the order $n - 1$ (inclusive) are prescribed. We further assume that the extremum of the functional I is attained on a curve $y = y(x)$ which is differentiable $2n$ times, and any admissible comparison curve $y = \bar{y}(x)$ is also $2n$ times differentiable. It is clear that both $y = y(x)$ and $y = \bar{y}(x)$ can be included in a one-parameter family of curves

$$y(x, \alpha) = y(x) + \alpha[\bar{y}(x) - y(x)]$$

such that $y(x, \alpha) = y(x)$ for $\alpha = 0$ and $y(x, \alpha) = \bar{y}(x)$ for $\alpha = 1$.

Now on the curves of the above family, the functional (1.39) reduces to a function of α, say, $\Psi(\alpha)$. Since the extremizing curve corresponds to $\alpha = 0$, we must have $\Psi'(\alpha) = 0$ at $\alpha = 0$. This gives, as in Section 1.2, for an extremum, the relation

$$\left[\frac{d}{d\alpha}\int_a^b F(x, y(x, \alpha), y'(x, \alpha), \ldots, y^{(n)}(x, \alpha))\, d\alpha\right]_{\alpha=0}$$

$$= \int_a^b (F_y\,\delta y + F_{y'}\,\delta y' + \ldots + F_{y^{(n)}}\,\delta y^{(n)})\, dx = 0 \tag{1.41}$$

Integrate by parts the second term on the right-hand side once and the third term twice, yielding

$$\int_a^b F_{y'}\,\delta y' dx = [F_{y'}\,\delta y]_a^b - \int_a^b \left(\frac{d}{dx}F_{y'}\right)\delta y\, dx$$

$$\int_a^b F_{y''}\,\delta y''\, dx = [F_{y''}\,\delta y']_a^b - \left[\left(\frac{d}{dx}F_{y'}\right)\delta y\right]_a^b + \int_a^b \left(\frac{d^2}{dx^2}F_{y''}\right)\delta y\, dx$$

and so on. The last term on the right-hand side of (1.41) can be written by successive integration by parts as

$$\int_a^b F_{y^{(n)}}\,\delta y^{(n)}\, dx = [F_{y^{(n)}}\,\delta y^{(n-1)}]_a^b - \left[\left(\frac{d}{dx}F_{y^{(n)}}\right)\delta y^{(n-2)}\right]_a^b$$

$$+ \ldots + (-1)^n \int_a^b \left(\frac{d^n}{dx^n}F_{y^{(n)}}\right)\delta y\, dx$$

By virtue of the boundary conditions (1.40a) and (1.40b), the integrated parts in all the above expressions on the right side vanish. Thus from (1.41), we find that on the extremizing curve

$$\int_a^b \left(F_y - \frac{d}{dx}F_{y'} + \frac{d^2}{dx^2}F_{y''} - \ldots + (-1)^n \frac{d^n}{dx^n}F_{y^{(n)}}\right)\delta y\, dx = 0$$

for an arbitrary choice of δy. Due to conditions of continuity imposed on F, the first factor in the foregoing integral is a continuous function of x in $[a, b]$. Thus invoking the fundamental lemma of Section 1.2, the function $y = y(x)$, which extremizes I satisfies

$$F_y - \frac{d}{dx}F_{y'} + \frac{d^2}{dx^2}F_{y''} - \ldots + (-1)^n \frac{d^n}{dx^n}F_{y^{(n)}} = 0 \tag{1.42}$$

which is known as the Euler-Poisson equation.

Clearly this is a differential equation of the order $2n$ and hence its solution involves $2n$ arbitrary constants. These are found by using the $2n$ boundary conditions (1.40a) and (1.40b).

Example 9. Determine the extremal of the functional

$$I[y(x)] = \int_{-l}^{l} \left(\frac{1}{2}\mu y''^2 + \rho y \right) dx$$

subject to

$$y(-l) = 0, \qquad y'(-l) = 0, \qquad y(l) = 0, \qquad y'(l) = 0.$$

Solution. This variational problem arises in finding the axis of a flexibly bent cylindrical beam clamped at the ends. If the beam is homogeneous, ρ and μ are constants. Then Euler-Poisson's equation (1.42) becomes

$$\rho + \frac{d^2}{dx^2}(\mu y'') = 0$$

whose solution satisfying the prescribed boundary conditions is

$$y = -\frac{\rho}{24\,\mu}(x^4 - 2l^2x^2 + l^4)$$

1.5 Functionals Dependent on Functions of Several Independent Variables

In the variational problems, considered so far, Euler's equations for determining extremals, are ordinary differential equations. We now extend this to the problem of determining the extrema of functionals involving multiple integrals leading to one or more partial differential equations. Consider, for example, the problem of finding an extremum of the functional

$$J[u(x, y)] = \iint_G F(x, y, u, u_x, u_y)\, dx\, dy \tag{1.43}$$

over a region of integration G by determining u which is continuous and has continuous derivatives upto the second order, and takes on prescribed values on the boundary of G. We further assume that F is thrice differentiable.

Let the extremizing surface be $u = u(x, y)$ so that an admissible one-parameter surface is taken as

$$u(x, y, \alpha) = u(x, y) + \alpha\eta(x, y)$$

where $\eta(x, y) = 0$ on the boundary of G. Then the necessary condition for an extremum is the vanishing of the first variation

$$\delta J = \left(\frac{\partial}{\partial \alpha} J[u + \alpha\eta] \right)_{\alpha = 0}$$

This implies from (1.43)

$$\iint_G (F_u \eta + F_{u_x} \eta_x + F_{u_y} \eta_y)\, dx\, dy = 0 \tag{1.44}$$

which may be again transformed by integration by parts. We assume that the boundary Γ of G admits of a tangent, which turns piecewise continuously. Then using the familiar Green's theorem, we have

$$\iint_G (\eta_x F_{u_x} + \eta_y F_{u_y})\, dx\, dy = \int_\Gamma \eta(F_{u_x}\, dy - F_{u_y}\, dx)$$

$$- \iint_G \eta\left(\frac{\partial}{\partial x} F_{u_x} + \frac{\partial}{\partial y} F_{u_y}\right) dx\, dy$$

Thus from (1.44),

$$\iint_G \left[F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y}\right] \eta\, dx\, dy + \int_\Gamma \eta(F_{u_x}\, dy - F_{u_y}\, dx) = 0.$$

Since $\eta = 0$ on Γ and the above relation holds for any arbitrary continuously differentiable function η, it follows from above that by using the generalization of the fundamental lemma of Section 1.2 that

$$F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} = 0 \tag{1.45}$$

The extremizing function $u(x, y)$ is determined from the solution of the second-order partial differential equation (1.45) which is known as *Euler-Ostrogradsky equation.* If the integrand of a functional J contains derivatives of order higher than two, then by a straightforward extension of the above analysis, we may derive a modified Euler-Ostrogradsky equation for determining extremals. For example, in the case of the functional

$$J[u(x, y)] = \iint_G F(x, y, u, u_x, u_y, u_{xx}, u_{xy}, u_{yy})\, dx\, dy$$

we get the equation for extremals as

$$F_u - \frac{\partial}{\partial x} F_{u_x} - \frac{\partial}{\partial y} F_{u_y} + \frac{\partial^2}{\partial x^2} F_{u_{xx}} + \frac{\partial^2}{\partial x \partial y} F_{u_{xy}} + \frac{\partial^2}{\partial y^2} F_{u_{yy}} = 0$$

Example 10. Find the Euler-Ostrogradsky equation for

$$I[u(x, y)] = \iint_D \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2\right] dx\, dy$$

where the values of u are prescribed on the boundary Γ of the domain D.

Solution. It clearly follows from (1.45) that the equation for extremals is

$$\nabla^2 u = \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = 0, \tag{1.46}$$

1.6 Variational Problems in Parametric Form

In many variational problems, it is more convenient and sometimes necessary to make use of a parametric representation of a line as follows:

$$x = \phi(t), \quad y = \Psi(t) \text{ for } t_0 \le t \le t_1 \tag{1.47}$$

Consider the functional

$$J[x(t), y(t)] = \int_{t_0}^{t_1} F(t, x, y, \dot{x}, \dot{y})\, dt, \tag{1.48}$$

where the integration is along the line (1.47) and a dot denotes derivative with respect to t. In order that the values of the functional (1.48) depend only on the line, and not on the parametrization (which can be accomplished in a number of ways), it is both necessary and sufficient that the integrand in (1.48) does not contain t explicitly and that it is homogeneous of the first degree in $\dot{x}$ and $\dot{y}$. Thus

$$F(x, y, k\dot{x}, k\dot{y}) = kF(x, y, \dot{x}, \dot{y}), \quad k > 0. \tag{1.49}$$

Take, for example,

$$I[x(t), y(t)] = \int_{t_0}^{t_1} \phi\,(x(t), y(t), \dot{x}(t), \dot{y}(t))\, dt,$$

where ϕ satisfies the homogeneity condition (1.49). Let us now consider a new parametric representation

$$\tau = \zeta(t)(\dot{\zeta}(t) \neq 0), \quad x = x\,(\tau), \quad y = y(\tau).$$

Then

$$\int_{t_0}^{t_1} \phi(x(t), y(t), \dot{x}(t), \dot{y}(t))\, dt = \int_{\tau_0}^{\tau_1} \phi(x(\tau), y(\tau), x_\tau(\tau)\dot{\zeta}(t), y_\tau(\tau)\, \dot{\zeta}(\tau))\, \frac{d\tau}{\dot{\tau}}.$$

But since ϕ is a homogeneous function of first degree in $\dot{x}$ and $\dot{y}$,

$$\phi(x, y, \dot{x}_\tau \dot{\zeta}, \dot{y}_\tau \dot{\zeta}) = \dot{\zeta}\phi(x, y, \dot{x}_\tau, \dot{y}_\tau).$$

This gives from the above equation

$$\int_{t_0}^{t_1} \phi(x, y, \dot{x}_t, \dot{y}_t)\, dt = \int_{\tau_0}^{\tau_1} \phi(x, y, \dot{x}_\tau, \dot{y}_\tau)\, d\tau.$$

Thus the integrand remains unchanged with a change in the parametric representation.

For example, the area bounded by a closed curve given by $\int_{t_1}^{t_2} (x\dot{y} - y\dot{x})\, dt$ is a functional which can be put in the form

$$I[x(t), y(t)] = \int_{t_1}^{t_2} \Phi(x, y, \dot{x}, \dot{y})\, dt,$$

where Φ is a homogeneous function of degree one in $\dot{x}$ and $\dot{y}$. Thus to find extremals for I, one has to solve Euler's equations

$$\Phi_x - \frac{d}{dt}\Phi_{\dot{x}} = 0, \quad \Phi_y - \frac{d}{dt}\Phi_{\dot{y}} = 0 \tag{1.50}$$

However, these equations are not independent, because these must be satisfied by a certain solution $x = x(t)$, $y = y(t)$, and any other pairs of functions with a different parametric representation of the same curve, which, in the case of Euler's equations being independent, would conflict with the theorem of existence and uniqueness of a solution of a system of differential equations. Thus we conclude that in (1.50), any one equation is a consequence of the other and to find the extremals, one has to solve any one of the equations (1.50) along with the equation $\dot{x}^2 + \dot{y}^2 = 1$, which shows that the arc length of the curve is taken as the parameter.

The Weirstrassian form of Euler equations (1.50) is

$$\frac{1}{r} = \frac{\Phi_{x\dot{y}} - \Phi_{y\dot{x}}}{\Phi_1(\dot{x}^2 + \dot{y}^2)^{3/2}} \tag{1.51}$$

where r is the radius of curvature of the extremal and Φ_1 is the common value of the ratios

$$\phi_1 = \phi_{\dot{x}\dot{x}}/\dot{y}^2 = \phi_{\dot{y}\dot{y}}/\dot{x}^2 = -\phi_{\dot{x}\dot{y}}/\dot{x}\dot{y}$$

For example, in finding the extremals of

$$I[x(t), y(t)] = \int_{t_0}^{t_1} [(\dot{x}^2 + \dot{y}^2)^{1/2} + a^2(x\dot{y} - y\dot{x})]\, dt$$

we first notice that the integrand $\Phi(x, y, \dot{x}, \dot{y})$ is homogeneous of degree one. Using Weirstrass form of Euler equations, we find that

$$F_{x\dot{y}} = a^2, \quad F_{y\dot{x}} = -a^2, \quad F_1 = \frac{1}{(\dot{x}^2 + \dot{y}^2)^{3/2}} \tag{1.52}$$

give $1/r = 2a^2$, which shows that the extremals are circles.

1.7 Some Applications to Problems of Mechanics

Hamilton's Principle

One of the most fundamental and important principles of mechanics and mathematical physics is the principle of least action due to Hamilton (William Rowan Hamilton (1805–1865), an Irish mathematician, also known for his invention of quaternions). Using this principle one can deduce the basic equations governing many physical phenomena. Let us formulate the principle for a dynamical system of particles and begin by considering the case of a single particle.

We consider a particle of mass m moving in a force field. If the position vector of the particle with respect to a fixed origin is denoted by $\mathbf{r}$, then by Newton's

1.6 Variational Problems in Parametric Form

In many variational problems, it is more convenient and sometimes necessary to make use of a parametric representation of a line as follows:

$$x = \phi(t), \quad y = \Psi(t) \text{ for } t_0 \le t \le t_1 \tag{1.47}$$

Consider the functional

$$J[x(t), y(t)] = \int_{t_0}^{t_1} F(t, x, y, \dot{x}, \dot{y})\, dt, \tag{1.48}$$

where the integration is along the line (1.47) and a dot denotes derivative with respect to t. In order that the values of the functional (1.48) depend only on the line, and not on the parametrization (which can be accomplished in a number of ways), it is both necessary and sufficient that the integrand in (1.48) does not contain t explicitly and that it is homogeneous of the first degree in $\dot{x}$ and $\dot{y}$. Thus

$$F(x, y, k\dot{x}, k\dot{y}) = kF(x, y, \dot{x}, \dot{y}), \quad k > 0. \tag{1.49}$$

Take, for example,

$$I[x(t), y(t)] = \int_{t_0}^{t_1} \phi\,(x(t), y(t), \dot{x}(t), \dot{y}(t))\, dt,$$

where ϕ satisfies the homogeneity condition (1.49). Let us now consider a new parametric representation

$$\tau = \zeta(t)(\dot{\zeta}(t) \neq 0), \quad x = x\,(\tau), \quad y = y(\tau).$$

Then

$$\int_{t_0}^{t_1} \phi(x(t), y(t), \dot{x}(t), \dot{y}(t))\, dt = \int_{\tau_0}^{\tau_1} \phi(x(\tau), y(\tau), x_\tau(\tau)\dot{\zeta}(t), y_\tau(\tau)\,\dot{\zeta}(\tau))\,\frac{d\tau}{\dot{\tau}}.$$

But since ϕ is a homogeneous function of first degree in $\dot{x}$ and $\dot{y}$,

$$\phi(x, y, \dot{x}_\tau\dot{\zeta}, \dot{y}_\tau\dot{\zeta}) = \dot{\zeta}\phi(x, y, \dot{x}_\tau, \dot{y}_\tau).$$

This gives from the above equation

$$\int_{t_0}^{t_1} \phi(x, y, \dot{x}_t, \dot{y}_t)\, dt = \int_{\tau_0}^{\tau_1} \phi(x, y, \dot{x}_\tau, \dot{y}_\tau)\, d\tau.$$

Thus the integrand remains unchanged with a change in the parametric representation.

For example, the area bounded by a closed curve given by $\int_{t_1}^{t_2} (x\dot{y} - y\dot{x})\, dt$ is a functional which can be put in the form

$$I[x(t), y(t)] = \int_{t_1}^{t_2} \Phi(x,y,\dot{x},\dot{y})\, dt,$$

where Φ is a homogeneous function of degree one in $\dot{x}$ and $\dot{y}$. Thus to find extremals for I, one has to solve Euler's equations

$$\Phi_x - \frac{d}{dt}\Phi_{\dot{x}} = 0, \quad \Phi_y - \frac{d}{dt}\Phi_{\dot{y}} = 0 \tag{1.50}$$

However, these equations are not independent, because these must be satisfied by a certain solution $x = x(t)$, $y = y(t)$, and any other pairs of functions with a different parametric representation of the same curve, which, in the case of Euler's equations being independent, would conflict with the theorem of existence and uniqueness of a solution of a system of differential equations. Thus we conclude that in (1.50), any one equation is a consequence of the other and to find the extremals, one has to solve any one of the equations (1.50) along with the equation $\dot{x}^2 + \dot{y}^2 = 1$, which shows that the arc length of the curve is taken as the parameter.

The Weirstrassian form of Euler equations (1.50) is

$$\frac{1}{r} = \frac{\Phi_{x\dot{y}} - \Phi_{y\dot{x}}}{\Phi_1(\dot{x}^2 + \dot{y}^2)^{3/2}} \tag{1.51}$$

where r is the radius of curvature of the extremal and Φ_1 is the common value of the ratios

$$\phi_1 = \phi_{\dot{x}\dot{x}}/\dot{y}^2 = \phi_{\dot{y}\dot{y}}/\dot{x}^2 = -\phi_{\dot{x}\dot{y}}/\dot{x}\dot{y}$$

For example, in finding the extremals of

$$I[x(t), y(t)] = \int_{t_0}^{t_1} [(\dot{x}^2 + \dot{y}^2)^{1/2} + a^2(x\dot{y} - y\dot{x})]\, dt$$

we first notice that the integrand $\Phi(x, y, \dot{x}, \dot{y})$ is homogeneous of degree one. Using Weirstrass form of Euler equations, we find that

$$F_{x\dot{y}} = a^2, \quad F_{y\dot{x}} = -a^2, \quad F_1 = \frac{1}{(\dot{x}^2 + \dot{y}^2)^{3/2}} \tag{1.52}$$

give $1/r = 2a^2$, which shows that the extremals are circles.

1.7 Some Applications to Problems of Mechanics

Hamilton's Principle

One of the most fundamental and important principles of mechanics and mathematical physics is the principle of least action due to Hamilton (William Rowan Hamilton (1805–1865), an Irish mathematician, also known for his invention of quaternions). Using this principle one can deduce the basic equations governing many physical phenomena. Let us formulate the principle for a dynamical system of particles and begin by considering the case of a single particle.

We consider a particle of mass m moving in a force field. If the position vector of the particle with respect to a fixed origin is denoted by $\mathbf{r}$, then by Newton's law of motion, the path of the particle is governed by the equation

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{f}, \tag{1.53}$$

where $\mathbf{f}$ is the force acting on the particle.

Now consider any other path $\mathbf{r} + \delta\mathbf{r}$ on the proviso that the true path and the varied path coincide at two distinct instants $t = t_1$ and $t = t_2$. This demands that

$$\delta\mathbf{r}\,|_{t_1} = \delta\mathbf{r}\,|_{t_2} = 0 \tag{1.54}$$

At any intermediate time t, we examine the true path $\mathbf{r}$ and varied path $\mathbf{r} + \delta\mathbf{r}$. Taking the dot product of variation $\delta\mathbf{r}$ into (1.53) and integrating the result with respect to time over the interval (t_1, t_2), we get

$$\int_{t_1}^{t_2}\left(m\frac{d^2\mathbf{r}}{dt^2}\cdot\delta\mathbf{r} - \mathbf{f}\cdot\delta\mathbf{r}\right)dt = 0 \tag{1.55}$$

Integrating the first term in (1.55) by parts, and by using (1.54) we find that

$$m\int_{t_1}^{t_2}\frac{d^2\mathbf{r}}{dt^2}\cdot\delta\mathbf{r}\,dt = m\left[\frac{d\mathbf{r}}{dt}\cdot\delta\mathbf{r}\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2}\frac{d\mathbf{r}}{dt}\cdot\delta\left(\frac{d\mathbf{r}}{dt}\right)dt\right]$$

$$= -\delta\int_{t_1}^{t_2}\frac{m}{2}\left(\frac{d\mathbf{r}}{dt}\right)^2 dt = -\delta\int_{t_1}^{t_2} T.dt$$

where T is the kinetic energy $\frac{1}{2}m(d\mathbf{r}/dt)^2$ of the particle. Substitution of the above relation in (1.55) then gives

$$\int_{t_1}^{t_2}(\delta T + \mathbf{f}\cdot\delta\mathbf{r})\,dt = 0 \tag{1.56}$$

This is Hamilton's principle in its most general form, for a single particle. Now consider the case when the force field $\mathbf{f}$ having components (X, Y, Z) is conservative which implies that

$$\mathbf{f}\cdot d\mathbf{r}\ (= X\,dx + Y\,dy + Z\,dz)$$

is the differential of a single-valued function $\Phi(x, y, z)$. This function is called the force potential and its negative, say V, is the potential energy of the particle. Thus

$$\mathbf{f}\cdot\delta\mathbf{r} = \delta\Phi = -\,\delta V$$

and its substitution in (1.56) gives

$$\delta\int_{t_1}^{t_2}(T - V)\,dt = 0 \tag{1.57}$$

which is Hamilton's principle. This states that the motion is such that the integral of the difference between the kinetic and potential energies is stationary for the actual path. The difference $T-V$ is known as the Lagrangian function. It can be shown further that this integral is a minimum for the true path, as compared with any neighbouring path having the same terminal configurations, at least over a sufficiently short time interval.

From Hamilton's principle, one can derive another important principle, known as Hertz's principle. This states that a particle (or a system) moving on a surface without external forces acting on it, follows a geodesic line.

A line q is called the geodesic on a surface if at each point of q, the principal normal coincides with the normal to the surface. As an example of Hertz's principle, we may note that a point on a spherical surface moves along a great circle if it is not acted by external forces. Similarly, a point on a cylindrical surface moves along a helix under the same circumstances.

By topological methods in variational problems (Lyusternik [2]), one can show that on any smooth closed surface, there are at least three closed geodesic lines. For example, an ellipsoid has three symmetry planes, and the ellipses, along which these three planes cut the ellipsoid, are closed geodesics.

If the force field is non-conservative (e.g., a dissipative system), the potential energy function does not exist, but (1.56) still holds and $\mathbf{f} \cdot \delta\mathbf{r}$ is the work done by the force $\mathbf{f}$ in a small displacement $\delta\mathbf{r}$. The foregoing study can be easily extended to a system of N particles by summation and to a continuous system by integration. Thus for N particles, the kinetic energy is

$$T = \sum_{k=1}^{N} \frac{1}{2} m_k \left(\frac{d\mathbf{r}_k}{dt} \right)^2$$

and the total work done by the forces is $\sum_{k=1}^{N} \mathbf{f}_k \cdot \delta\mathbf{r}_k$.

In fact, the principle applies equally well to a general dynamical system consisting of a system of particles and rigid bodies.

Let us now apply the above principle to derive the equation of vibration of a rectilinear bar. A displacement from the equilibrium position $u(x, t)$ will be a function of x (measured along the bar in the undisturbed position) and time t. Thus the kinetic energy of the bar of length l is

$$T = \frac{1}{2} \int_0^l \rho \left(\frac{\partial u}{\partial t} \right)^2 dx$$

We assume that the bar is slightly extensible. The potential energy of an elastic bar with a constant curvature is proportional to the square of the curvature. Thus the differential dV of the potential energy of the bar is

$$dV = \frac{1}{2} k \left[\frac{\partial^2 u}{\partial x^2} \Bigg/ \left[1 + \left(\frac{\partial u}{\partial x} \right)^2 \right]^{3/2} \right]^2 dx$$

law of motion, the path of the particle is governed by the equation

$$m\frac{d^2\mathbf{r}}{dt^2} = \mathbf{f}, \tag{1.53}$$

where $\mathbf{f}$ is the force acting on the particle.

Now consider any other path $\mathbf{r} + \delta\mathbf{r}$ on the proviso that the true path and the varied path coincide at two distinct instants $t = t_1$ and $t = t_2$. This demands that

$$\delta\mathbf{r}\,|_{t_1} = \delta\mathbf{r}\,|_{t_2} = 0 \tag{1.54}$$

At any intermediate time t, we examine the true path $\mathbf{r}$ and varied path $\mathbf{r} + \delta\mathbf{r}$. Taking the dot product of variation $\delta\mathbf{r}$ into (1.53) and integrating the result with respect to time over the interval (t_1, t_2), we get

$$\int_{t_1}^{t_2}\left(m\frac{d^2\mathbf{r}}{dt^2}\cdot\delta\mathbf{r} - \mathbf{f}\cdot\delta\mathbf{r}\right)dt = 0 \tag{1.55}$$

Integrating the first term in (1.55) by parts, and by using (1.54) we find that

$$m\int_{t_1}^{t_2}\frac{d^2\mathbf{r}}{dt^2}\cdot\delta\mathbf{r}\,dt = m\left[\frac{d\mathbf{r}}{dt}\cdot\delta\mathbf{r}\Big|_{t_1}^{t_2} - \int_{t_1}^{t_2}\frac{d\mathbf{r}}{dt}\cdot\delta\left(\frac{d\mathbf{r}}{dt}\right)dt\right]$$

$$= -\delta\int_{t_1}^{t_2}\frac{m}{2}\left(\frac{d\mathbf{r}}{dt}\right)^2 dt = -\delta\int_{t_1}^{t_2} T.dt$$

where T is the kinetic energy $\frac{1}{2}m(d\mathbf{r}/dt)^2$ of the particle. Substitution of the above relation in (1.55) then gives

$$\int_{t_1}^{t_2}(\delta T + \mathbf{f}\cdot\delta\mathbf{r})\,dt = 0 \tag{1.56}$$

This is Hamilton's principle in its most general form, for a single particle. Now consider the case when the force field $\mathbf{f}$ having components (X, Y, Z) is conservative which implies that

$$\mathbf{f}\cdot d\mathbf{r}\ (= X\,dx + Y\,dy + Z\,dz)$$

is the differential of a single-valued function $\Phi(x, y, z)$. This function is called the force potential and its negative, say V, is the potential energy of the particle. Thus

$$\mathbf{f}\cdot\delta\mathbf{r} = \delta\Phi = -\,\delta V$$

and its substitution in (1.56) gives

$$\delta\int_{t_1}^{t_2}(T - V)\,dt = 0 \tag{1.57}$$

which is Hamilton's principle. This states that the motion is such that the integral of the difference between the kinetic and potential energies is stationary for the actual path. The difference T-V is known as the Lagrangian function. It can be shown further that this integral is a minimum for the true path, as compared with any neighbouring path having the same terminal configurations, at least over a sufficiently short time interval.

From Hamilton's principle, one can derive another important principle, known as Hertz's principle. This states that a particle (or a system) moving on a surface without external forces acting on it, follows a geodesic line.

A line q is called the geodesic on a surface if at each point of q, the principal normal coincides with the normal to the surface. As an example of Hertz's principle, we may note that a point on a spherical surface moves along a great circle if it is not acted by external forces. Similarly, a point on a cylindrical surface moves along a helix under the same circumstances.

By topological methods in variational problems (Lyusternik [2]), one can show that on any smooth closed surface, there are at least three closed geodesic lines. For example, an ellipsoid has three symmetry planes, and the ellipses, along which these three planes cut the ellipsoid, are closed geodesics.

If the force field is non-conservative (e.g., a dissipative system), the potential energy function does not exist, but (1.56) still holds and $\mathbf{f} \cdot \delta\mathbf{r}$ is the work done by the force $\mathbf{f}$ in a small displacement $\delta\mathbf{r}$. The foregoing study can be easily extended to a system of N particles by summation and to a continuous system by integration. Thus for N particles, the kinetic energy is

$$T = \sum_{k=1}^{N} \frac{1}{2} m_k \left(\frac{d\mathbf{r}_k}{dt} \right)^2$$

and the total work done by the forces is $\sum_{k=1}^{N} \mathbf{f}_k \cdot \delta\mathbf{r}_k$.

In fact, the principle applies equally well to a general dynamical system consisting of a system of particles and rigid bodies.

Let us now apply the above principle to derive the equation of vibration of a rectilinear bar. A displacement from the equilibrium position $u(x, t)$ will be a function of x (measured along the bar in the undisturbed position) and time t. Thus the kinetic energy of the bar of length l is

$$T = \frac{1}{2} \int_0^l \rho \left(\frac{\partial u}{\partial t} \right)^2 dx$$

We assume that the bar is slightly extensible. The potential energy of an elastic bar with a constant curvature is proportional to the square of the curvature. Thus the differential dV of the potential energy of the bar is

$$dV = \frac{1}{2} k \left[\frac{\partial^2 u}{\partial x^2} \Big/ \left[1 + \left(\frac{\partial u}{\partial x} \right)^2 \right]^{3/2} \right]^2 dx$$

k being a constant. Thus the potential energy of the entire bar, whose axis-curvature is variable, is

$$U = \frac{1}{2}k \int_0^l \left[\frac{\partial^2 u}{\partial x^2} \Big/ \left[1 + \left(\frac{\partial u}{\partial x} \right)^2 \right]^{3/2} \right]^2 dx$$

According to our assumption of slight extensibility, the deviations of the bar from the equilibrium position are small, and the term $(\partial u/\partial x)^2$ may be ignored. Now by Hamilton's principle, the integral

$$\int_{t_1}^{t_2} \int_0^l \left[\frac{1}{2}\rho \left(\frac{\partial u}{\partial t} \right)^2 - \frac{1}{2}k \left(\frac{\partial^2 u}{\partial x^2} \right)^2 \right] dx\, dt$$

will be an extremum for fixed terminal times t_1 and t_2. The Euler-Ostogradsky equation then gives

$$\frac{\partial}{\partial t}\left(\rho \frac{\partial u}{\partial t} \right) + \frac{\partial^2}{\partial x^2}\left(k \frac{\partial^2 u}{\partial x^2} \right) = 0$$

which is the governing equation for displacement $u(x, t)$.

In a similar manner it can be shown that the differential equation for the displacement $u(x, t)$ of a highly flexible and almost inextensible homogeneous string from its equilibrium position is

$$\frac{\partial^2 u}{\partial t^2} = \frac{T}{\rho} \cdot \frac{\partial^2 u}{\partial x^2}$$

where T and ρ denote the tension and the line density of the material of the string.

There is an important generalization of the above equation, when the string is acted on by a uniformly distributed linear restoring force, directed towards the equilibrium position. This leads to adding a term of the form $-ku$ (k being a positive constant) to the right-hand side of the above equation. The new equation is known as the *Klein-Gordon equation* which, in its general form, is given by

$$\frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} = \nabla^2 u - ku, \quad c^2 = T/\rho$$

1.8 Variational Problems Leading to an Integral Equation or a Differential-Difference Equation

Thus far we have been concerned with variational problems involving functionals formed by integrating a certain differential expression in the argument function. But more general classes of functionals are often encountered in variational problems.

Let us consider the functional

$$I[\phi] = \iint K(s, t)\phi(s)\phi(t)\, ds\, dt + \int [\phi(s)]^2\, ds - 2\int \phi(s) f(s)\, ds,$$

which we want to extremize. Here $K(s, t)$ is a given continuous function with $K(s, t) = K(t, s)$, $f(s)$ is a given continuous function of s and $\phi(s)$ is the unknown continuous function. All integrations are confined to the interval $a \le s \le b$, $a \le t \le b$. We replace ϕ by $\phi + \varepsilon\zeta$ and consider $I(\phi + \varepsilon\zeta) = \Psi(\varepsilon)$. The variation δI given by $[d\Psi/d\varepsilon]_{\varepsilon=0}$ is obtained after some transformation as

$$2\int_a^b \zeta(t)\left[\int_a^b K(s,t)\phi(s)\,ds + \phi(t) - f(t)\right]dt.$$

Hence the requirement $\delta I = 0$ for an extremum leads to Fredholm integral equation

$$\int_a^b K(s,t)\phi(s)\,ds + \phi(t) = f(t).$$

as Euler equation for the problem.

Let us next consider another functional

$$I[\phi] = \int_{-\infty}^{\infty} [p(x)\,(\phi'(x))^2 + 2\phi(x+1)\,\phi(x-1) - \phi^2(x) - 2\phi(x)f(x)]\,dx$$

which is to be extremized. Here the argument function is continuous and has a piecewise continuous derivative in the entire interval $-\infty < x < \infty$. Now

$$\delta I = \left[\frac{d}{d\varepsilon} I(\phi + \varepsilon\zeta)\right]_{\varepsilon=0}$$

$$= 2\int_{-\infty}^{\infty} \zeta(x)\,[-(p\phi')' + \phi(x+2) + \phi(x-2) - \phi(x) - f(x)]\,dx$$

Now the vanishing of δI for arbitrary ζ gives

$$(p\phi')' - \phi(x+2) - \phi(x-2) + \phi(x) + f(x) = 0$$

which is a differential-difference equation for the argument function $\phi(x)$.

1.9 Theorem of du Bois-Reymond

It may be recalled that in deriving Euler's equation for the functional (1.8), it was assumed that the admissible functions admit of continuous first order derivative. However, a variational problem with integrand $F(x, y, y')$ is also meaningful when y' is required to be only piecewise continuous.

Consider first an actual minimum problem, such that $y(x)$ is that function with continuous first and second derivatives, which renders I in (1.8) a minimum. Then it can be shown that $y(x)$ yields a minimum if we expand the class of functions to include functions y^* which need not have second derivatives. In fact, according to Weirstrass approximation theorem, we can approximate the function y^* by a polynomial $p(x)$ and y^* by the derived polynomial $p'(x)$ as closely as we like, where $p(x)$ satisfies the boundary conditions $p(a) = y_1$ and $p(b) = y_2$ as in Section 1.2. Then clearly $I[p(x)]$ differs arbitrarily little from $I[y^*]$. But since

$p(x)$ is an admissible comparison function with continuous second derivatives, $I[p(x)] \geq I[y(x)]$ and therefore $I[y^*(x)] \geq I[y(x)]$. This proves the result.

We next prove our assertion in Section 1.2 that if in this variational problem $y(x)$ is the extremizing function (satisfying the boundary conditions) admitting first order continuous derivative, then $y(x)$ admits of a continuous second derivative also provided that $F_{y'y'} \neq 0$. This is the theorem of du Bois Reymond (see Courant and Hilbert [3a]).

We first prove the following lemma: If $\phi(x)$ is a piecewise continuous function in $[a, b]$ and if

$$\int_a^b \phi(x)\eta(x)\, dx = 0$$

holds for arbitrary continuous function $\eta(x)$ satisfying the condition

$$\int_a^b \eta(x)\, dx = 0, \tag{1.59}$$

then $\phi(x)$ is a constant.

To prove this lemma, we first note that the relation (1.58) is clearly satisfied for constant ϕ. We now fix a constant C such that $\int_a^b (\phi - C)\, dx = 0$ for the given ϕ. Then from (1.58) and (1.59), we have $\int_a^b (\phi - C)\eta\, dx = 0$. Setting $\eta = \phi - C$ in this relation gives $\int_a^b (\phi - C)^2\, dx = 0$ and this proves the lemma. This result can be generalized in the following manner: If $\phi(x)$ is a piecewise continuous function satisfying $\int_a^b \phi\eta\, dx = 0$ for all continuous functions $\eta(x)$ such that

$$\int_a^b \eta\, dx = 0, \int_a^b x\eta\, dx = 0, \ldots, \int_a^b x^n \eta\, dx = 0$$

then ϕ is a polynomial of nth degree.

To prove du Bois-Reymond's theorem we note from (1.13) that the relation

$$\int_a^b [F_y \zeta + F_{y'} \zeta']\, dx = 0$$

holds for any continuously differentiable function $\zeta(x)$, satisfying $\zeta(x_0) = \zeta(x_1) = 0$. Putting $F_y = A'$, $F_{y'} = B$, we obtain, after integration by parts, the relation

$$\int_{x_0}^{x_1} (A'\zeta + B\zeta')\, dx = \int_{x_0}^{x_1} \zeta'(B - A)\, dx = 0$$

We select an arbitrary function $\zeta' = \eta$ which is continuous and satisfies

$$\int_{x_0}^{x_1} \eta\, dx = \zeta(x_1) - \zeta(x_0) = 0$$

Applying the previous lemma, we obtain

$$B - A = F_{y'} - \int_{x_0}^{x_1} F_y \, dx = c \tag{1.60}$$

where c does not depend on x. Equation (1.60) takes the place of Euler's equation.

Now since $\int_{x_0}^{x_1} F_y \, dx$ is differentiable with respect to x, it follows from (1.60) that $F_{y'}$ also is differentiable. Hence Euler's equation

$$\frac{d}{dx} F_{y'} - F_y = 0$$

holds. Now if F is twice continuously differentiable with respect to its arguments, and further $F_{y'y'} \neq 0$ is satisfied, it follows that the piecewise continuous function y' is also continuous and has a continuous derivative. Because, if $F_{y'y'} \neq 0$, y' may be expressed as a continuously differentiable function $\phi(x, y, F_{y'})$. Further by virtue of (1.60), $F_{y'}$ is a continuous function of x and y' is also continuous. Hence the arguments y and $F_{y'}$ of ϕ are continuously differentiable and hence y' $(= \phi)$ is also continuously differentiable. This establishes Du Bois-Reymond's theorem which can be easily extended to an integrand of the form $F(x, y, y', \ldots, y^{(n)})$ by using the generalization of the above lemma.

1.10 Stochastic Calculus of Variations

We have already seen in Section 1.7 that in classical mechanics the dynamical laws of motion are represented by a variational principle given by Hamilton's principle of least action. In particular, for a dynamical system with f degrees of freedom, the possible motion is given by a flow in R^f which makes the functional

$$J = \int_a^b L(x(t), \dot{x}(t)) \, dt$$

stationary. Here $L \in C^1 (R^{2f} \to R)$ is the Lagrangian of the system, $x \in C^2 ([a, b] \to R^f)$ is the flow and $\dot{x}$ is the velocity dx/dt. In this case, Newton's dynamical law follows from the Euler-Lagrange equation

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} = 0$$

for the functional J.

In quantum mechanics, the dynamical law is given by Schrödinger equation

$$i\hbar \frac{\partial \Psi}{\partial t} = \left(- \frac{\hbar^2}{2m} \nabla \cdot \nabla + V \right) \Psi,$$

where h is the Planck constant divided by 2π, $\Psi \in L_2 (R^f \to C)$. Here L_2 is the space of square integrable functions in the Lebesgue sense. Motion of this system is determined by a one-parameter unitary flow in a Hilbert space $L_2(R^f \to C)$ generated by Schrödinger equation.

It has been opined that the dynamical law in quantum mechanics is radically different from that in classical mechanics. In particular there is no least action principle analogous to that of Hamilton in quantum mechanics, although Schwinger [4] derived a weak version of the principle.

Recently, Yasue [5] developed a theory of stochastic calculus of variations, which might be regarded as a generalization of the ordinary calculus of variations, to stochastic processes.

We now give, following Yasue, a brief exposition of this principle. Let (Ω, A, Pr) be a probability space, where Ω is a certain non-empty set, A is a σ-algebra of subsets of Ω and Pr is a probability measure defined on A. A mapping x from an open time interval I into a Hilbert space $H = L_2((\Omega, Pr) \to R^f)$ is called a stochastic process of second order in R^f if $t \to x(t)$ is continuous from I into H. Now let $P = \{p_t\}_{t \in I}$ and $F = \{f_t\}_{t \in I}$ be an increasing family and a decreasing family of σ-algebras, respectively, such that $x(t)$ is p_t-measurable and f_t-measurable.

If

$$Dx(t) = \lim_{h \to 0+} E\left[\frac{x(t+h) - x(t)}{h}\bigg|_{p_t}\right]$$

exists as a limit in H for each t in I and $t \to D\,x(t)$ is continuous from I into H, then the stochastic process is said to be mean forward differentiable. Further, if

$$D_* x(t) = \lim_{h \to 0+} E\left[\frac{x(t) - x(t-h)}{h}\bigg|_{f_t}\right]$$

exists, we say that $x(t)$ is mean backward differentiable. In the above definitions, $E[\cdot \mid B]$ denotes the conditional expectation with respect to a σ-algebra $B \subset A$. Further, Dx and D_*x are said to be mean forward and backward derivatives.

We now denote, by $C^1(I \to H)$, the totality of mean forward and backward differentiable stochastic processes of the second order adapted to P and F. The completion of C^1 in the norm

$$|||x||| = \sup_{t \in I} (\|x(t)\|_H + \|Dx(t)\|_H + \|D_* x(t)\|_H)$$

is also denoted by $C^1(I \to H)$, where $\|\cdot\|_H$ is the norm of H.

Let $L \in C^1(R^{3f} \to R)$ and we consider a functional defined on $C^1(I \to H)$,

$$J_{ab} = E\left[\int_a^b L(x(t), Dx(t), D_* x(t))\, dt\right]$$

where E denotes the expectation and $a, b \in I$ with $a < b$. A functional J defined on $C^1(I \to H)$ is said to be differentiable at $x \in C^1(I \to H)$ if

$$J(x + z) - J(x) = dJ(x, z) + R(x, z)$$

Here dJ is a linear functional of $z \in C^1(I \to H)$ and $R(x, z) = 0\,(\|z\|)$. The linear functional dJ on $C^1(I \to H)$ is called the variation of the functional J at $x \in C^1(I \to H)$.

Kunio Yasue made some regularity assumptions on $x \in C^1(I \to H)$ and $L \in C^1(R^N \to R)$, viz. $\partial L/\partial Dx(t)$ is adapted to F and mean backward differentiable, and $\partial L/\partial D_*$ is adapted to P and mean forward differentiable. A stochastic process $x \in C^1(I \to H)$ satisfying these regularity assumptions is said to be *L*-adapted. Kunio Yasue then proved the following theorem of stochastic calculus of variations (proof omitted). The functional J_{ab} given above is differentiable at any *L*-adapted process $x \in C^1(I \to H)$, and its variation is given by

$$dJ_{ab} = E\left[\int_a^b \left[\frac{\partial L}{\partial x(t)} - D_*\left(\frac{\partial L}{\partial Dx(t)}\right) - D\left(\frac{\partial L}{\partial D_* x(t)}\right)\right] z(t)\, dt\right]$$

$$+ E\left[\left(\frac{\partial L}{\partial Dx(t)} + \frac{\partial L}{\partial D_* x(t)}\right) \cdot z(t)\Big|_a^b\right]$$

An *L*-adapted process $x \in C^1(I \to H)$ is called a stationary point or an extremal of the functional J_{ab} if it destroys the differential at x, i.e., $dJ_{ab} = 0$.

We now state the following fundamental theorem: A necessary and sufficient condition, for an *L*-adapted process to be an extremal of the functional J_{ab} given before with fixed end points $x(a) = x_a$ and $x(b) = x_b$ is that it satisfies

$$\frac{\partial L}{\partial x(t)} - D_*\left(\frac{\partial L}{\partial Dx(t)}\right) - D\left(\frac{\partial L}{\partial D_* x(t)}\right) = 0 \qquad (1.61)$$

almost surely (a.s.), where x_a and x_b belong to H.

To prove this theorem we note that since $z(a) = z(b) = 0$,

$$dJ_{ab} = E\left[\int_a^b \left[\frac{\partial L}{\partial x(t)} - D_*\left(\frac{\partial L}{\partial Dx(t)}\right) - D\left(\frac{\partial L}{\partial D_* x(t)}\right)\right] z(t)\, dt\right]$$

It suffices to show that for a stochastic process of second order y,

$$E\left[\int_a^b y(t)\, z(t)\, dt\right] = 0$$

for any $z \in C^1(I \to H)$ iff $y = 0$ (a.s.). Assume $y(u) > 0$ (a.s.) for $u \in (a, b)$. Then by continuity $y(t) > C > 0$ (a.s.) in a neighbourhood of u; $a < u - d < t < u + d < b$, $d > 0$. But since $z \in C^1(I \to H)$ is arbitrary, we may choose a mean square differentiable process z such that $z(t) = 0$ (a.s.) for $a \le t \le u - d$, $u + d \le t \le b$, $z(t) > 0$ (a.s.) for $u - d < t < u + d$ and $z(t) = 1$ (a.s.) for $u - \frac{d}{2} < t < u + \frac{d}{2}$. Then $\int_a^b y(t) \cdot z(t)\, dt \ge Cd > 0$ (a.s.) and we get a contradiction. This proves the theorem.

To give an example, we consider a Markov process $x \in C^1(I \to H)$ and L is given by

$$L(x, Dx, D_*x) = L_1(Dx, D_*x) - V(x)$$

$$= \frac{1}{2}\left(\frac{1}{2} m\,|Dx|^2 + \frac{1}{2} m\,|D_*x|^2\right) - V(x)$$

where m is the particle mass and L_1 and V correspond to the kinetic and potential energy. Here the Euler equation (1.61) becomes

$$\frac{m}{2}[DD_*x + D_*Dx] = -\nabla V(x)$$

almost surely, which might be regarded as the generalization of Newton's dynamical law to a stochastic process.

1.11 Supplementary Remarks

Variational Principle for the Equation $y''(x) = f(x, y, y')$

It can be shown that any equation of the above form is a Euler equation for some functional

$$I[y(x)] = \int F(x, y, y')\, dx.$$

This can be established by seeking the functional for which the Euler's equation

$$F_y - F_{y'x} - F_{y'y}\, y' - F_{y'y'}\, y'' = 0$$

coincides with the above second-order differential equation. This means that there must be an identity with respect to x, y, y'

$$F_y - F_{y'x} - F_{y'y}y' - F_{y'y'} \cdot f(x, y, y') \equiv 0.$$

Differentiating this with respect to y', we get

$$F_{y'y'x} + F_{y'y'y}\, y' + F_{y'y'y'} \cdot f + F_{y'y'} \cdot f_{y'} = 0.$$

Setting $u = F_{y'y'}$, we obtain the partial differential equation (PDE)

$$\frac{\partial u}{\partial x} + y' \frac{\partial u}{\partial y} + f \frac{\partial u}{\partial y'} + f_{y'} \cdot u = 0.$$

Hence finding the functional, that is, finding the function $F(x, y, y')$, reduces to the solution of the above PDE, and to subsequent quadrature.

Existence of an Extremum

We conclude this chapter by pointing out a characteristic difficulty in the solution of variational problems. In problems involving ordinary maxima or minima of a function, the existence of a solution is guaranteed by the fundamental theorem of Weirstrass. However variational problems, even if they are meaningfully formulated, may not have solutions. This stems from the fact that it is not, in general, possible to choose the domain of admissible functions as a 'compact set' for which the principle of point of accumulation holds.

For definiteness, let us consider the functional

$$I[y(x)] = \int_a^b [P(x)y'^2 + Q(x)y^2]\,dx \tag{1.62}$$

with the boundary conditions

$$y(a) = y(b) = 0. \tag{1.63}$$

To fix ideas we also assume an integral constraint

$$\int_a^b y^2\,dx = 1. \tag{1.64}$$

We suppose that $P(x)$ and $Q(x)$ are continuous in $[a, b]$ with $P(x) > 0$. It then follows from (1.62) that

$$I[y] \geq \int_a^b Q(x)y^2\,dx \geq \min_{a \leq x \leq b} Q(x) \int_a^b y^2\,dx = \min_{a \leq x \leq b} Q(x).$$

Thus the above functional is bounded below. Now, even if the least value of $I[y]$ is not attained, we can construct a sequence of functions $Y_n(x)$ satisfying (1.63) and (1.64) such that $\lim_{n \to \infty} I[y_n(x)]$ is equal to the greatest lower bound. Such a sequence is known as the minimizing sequence for the functional I.

If a minimizing sequence $\{Y_n(x)\}$ (or its subsequence) is convergent (say, uniformly convergent) to a limit function $Y(x)$, then the functional attains a minimum value for $y = Y(x)$. In fact, we have

$$\int_a^b Y^2\,dx = \lim_{n \to \infty} \int_a^b Y_n^2(x)\,dx = 1$$

$$\int_a^b Q(x)Y^2(x)\,dx = \lim_{n \to \infty} \int_a^b Q(x)Y_n^2(x)\,dx.$$

But

$$\int_a^b P(x)Y'^2(x)\,dx \leq \lim_{n \to \infty} \int_a^b P(x)Y_n'^2(x)\,dx.$$

The last result may appear somewhat strange because one may think that the left hand side equals the right hand side. However, there are cases when the right hand side exceeds the left hand side. Both possibilities are demonstrated in Figs. 1.10(a) and 1.10(b).

Thus we have

$$I[Y] \leq \lim_{n \to \infty} I[Y_n(x)]$$

The sequence $\{Y_n(x)\}$ being a minimizing sequence, we conclude that the quantity $I[Y]$ is the least value and the existence of a minimum is therefore established.

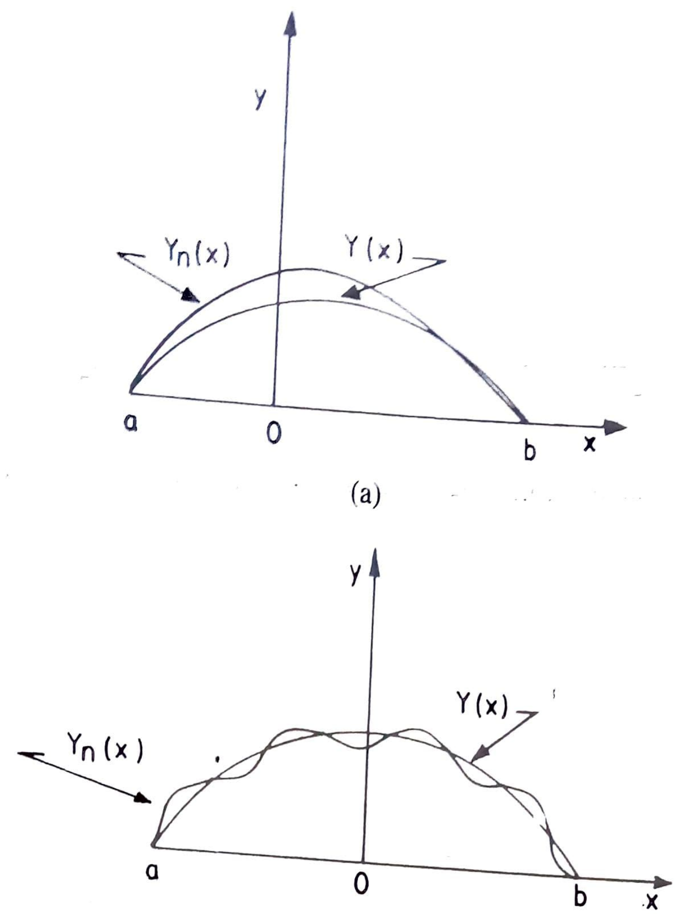

Fig. 1.10 (a) The minimizing function $Y_n(x)$ and limit function $Y(x)$ in the case when the left side of the above inequality equals the right side, and (b) The minimizing function $Y_n(x)$ and limit function $Y(x)$ in the case when the left side of the above inequality is less than the right side.

Conditions for the existence of a convergent sequence $\{Y_n(x)\}$ involve the concept of compactness, which plays a vital role in these and many other problems. A set M of points (elements) of a normed space R (or a more general metric space (R)) is said to be compact in (R), if every infinite sequence of points belonging to M has at least one convergent subsequence. Using this notion we may rephrase the Bolzano-Weirstrass theorem as: every bounded set of points belonging to a finite-dimensional Euclidean space is compact. But every unbounded set of a finite-dimensional Euclidean space is noncompact.

An important feature of infinite-dimensional spaces is that their bounded subsets are not necessarily compact, and therefore, the investigation of these spaces presents difficulties. For example, the sequence of functions $\sin x$, $\sin 2x$, $\sin 3x$, ... is bounded, but noncompact as a subset of the space $C[a, b]$ for fixed a and b. This is due to the fact that the values of $\sin nx$, $n = 1, 2, 3, \ldots$ oscillate between -1 and 1 with increasing frequency as $n \to \infty$ and, therefore, the subsequences of this sequence do not converge uniformly. But it is also possible to prove that every minimizing sequence $\{Y_n\}$ mentioned above is not only bounded, but also compact in the space $C[a, b]$ because, in this case, the absence of uniformly convergent subsequences would imply that there is a subsequence of the number sequence $\left\{\int_a^b Y_n'^2\, dx\right\}$, which increases indefinitely and hence the sequence $I\{Y_n\}$ becomes unbounded as well. These properties ensure the existence of a uniformly convergent subsequence of the sequence Y_n, whose limit function $Y(x)$ renders the functional minimum.

An alternative approach to the above problem of existence is due to Tonelli (see Young [6]) and is based on the notion of semi-continuity. A function $F(P)$, defined in a set of P, in which limits have a meaning, is said to be lower semi-continuous at P_0, if it satisfies the following conditions: (a) $F(P)$ is extended real-valued, i.e., its values are real with the possible addition of $+\infty$ and $-\infty$, (b) $F(P_0)$ is defined and $F(P_0) \neq -\infty$, (c) $F(P_0) \leq \liminf F(P)$ as $P \to P_0$. If these conditions are satisfied for each P_0 of the set, we say that $F(P)$ is lower semi-continuous.

Now consider the variational problem for the functional

$$I[C] = \int_a^b f(x, y(x), y'(x))\, dx$$

where C is a curve $y = y(x)$ joining two given points a and b. Then it was shown by Tonelli that $I[C]$ is lower semi-continuous for C belonging to any class K of curves of uniformly bounded lengths situated in a cube if $f(x, y(x), y'(x))$ is convex in $y'(x)$. Further $I[C]$ attains its minimum in K if K is closed.

The link between the proof of existence based on compactness and Tonelli's approach lies in the fact that a closed bounded set in a finite-dimensional space is compact and the well known Weirstrass principle (valid for a continuous function on a closed compact set) remains valid for a lower semi-continuous function, defined on a sequentially compact closed set.

The general scheme presented above involves most of the basic techniques for solving variational problems. These schemes reduce to testing that the functional is bounded below, selecting a minimizing sequence, and proving the compactness of the sequence in an appropriately chosen function space.

There is yet another way of overcoming the difficulty of not having a convergent minimizing sequence, to ensure the existence of an extremum as suggested by Young [6]. It adapts the idea of L. Schwartz for constructing a dual space, and embedding the original space (of admissible curves in our variational problem) into its dual. To fix ideas, consider the problem of minimizing a functional $\int F(\mathbf{x}(t), \mathbf{x}'(t), t), dt,\ t_1 \leq t \leq t_2$, where $\mathbf{x}(t)$ varies along a parametrized arc Γ of finite length. In this context F and Γ are elements of dual spaces. Let the function F belong to some normed space B, so that $\int_\Gamma F\, dt \equiv \langle F, \Gamma\rangle$ clearly defines a linear functional in the dual space B^*. Defining the operation $\oplus$ as union of arcs, we have $\langle F, (\Gamma_1 \oplus \Gamma_2)\rangle = \langle F, \Gamma_1\rangle + \langle F, \Gamma_2\rangle$ and $\langle F, c\Gamma\rangle = c\,\langle F, \Gamma\rangle = \langle cF, \Gamma\rangle$ for any real number c. We see that this functional is indeed bilinear and bounded.

Thus Γ is an element of B^*. Now following the ideas of L. Schwartz, one can define weak convergence in B^*. Such weak limits are called generalized arcs. However the limitations of this approach are similar to those of the theory of distributions.

From the foregoing discussion, it may appear that the non existence of solution owes its origin to the lack of compactness or closure properties of the family of admissible solutions of a variational problem. However, an entirely different reason for nonexistence of solutions of a variational problem, may also arise.

We pose the following question: When are the initial or boundary conditions well-posed (see ref. [3b]) for a certain class of boundary value problems, involving differential equations? This well-posedness in the sense of Hadamard implies that the solution exists, is unique and depends continuously on the initial/boundary data. Thus, if a differential system is not well-posed, the corresponding variational problem is also not well-posed. The third requirement above, which is particularly incisive, is necessary, if the mathematical formulation in the form of differential equations is to describe observable natural phenomena. Data in nature cannot be assumed to be rigidly fixed. The mere process of measuring them involves small errors. Thus a poor mathematical modelling of a real life problem may lead to ill-posed differential system so that the corresponding variational problem may also be ill-posed, as the following example due to Caratheodory shows.

A variational problem leads to the following Euler-Lagrange equation as the necessary condition for the existence of an extremum

$$y_t = (1 + x_t^2)^{1/2}$$

with the boundary conditions

$$x(0) = y(0) = 0, \quad x(1) = y(1) = 1.$$

It can be shown that the above boundary value problem is not well-posed so that no differentiable solution satisfies the above system.

A simple geometrical example of non-existence of solution can be given as follows: Two points of the x-axis are to be joined by the shortest possible line of continuous curvature which is perpendicular to the x-axis at the end-points. Clearly this problem has no solution. In fact, the length of such a line is always greater than that of the straight line, joining the two end points, but it may approximate this length as closely as desired. Hence there exists a greatest lower bound, but no minimum for admissible curves.

We may now sum up the foregoing considerations about the existence of an extremum in a variational problem. The characterization of such an extremum, in the absence of an existence proof, may turn out to be a nonsense, as in the following example of Oscar Perron. Let us assume that there exists a largest positive integer N. Thus $N \geq n$ for any positive integer n. If $N > 1$, then clearly $N^2 > N$. But $N \geq N^2$ by our hypothesis leads to $N^2 = N$. This gives $N = 1$. Although there is nothing wrong in the proof, the nonsense arises from our original assumption of the existence of a largest positive integer. One may arrive at a similar nonsense if necessary or sufficient conditions for the extremum of a functional are derived, without first checking whether such an extremum is attained in the class of admissible functions. This has a bearing on a basic difficulty arising in the modelling of a physical phenomenon. The existence of solutions in a mathematical model must realistically reflect our physical experience.

PROBLEMS

1. Test for an extremum the functional

$$I[y(x)] = \int_0^1 (xy + y^2 - 2y^2 y')\, dx,\ y(0) = 1,\ y(1) = 2.$$

Ans. An extremum is not achieved on the class of continuous functions.

2. Find the extremals of the functional

$$I[y(x)] = \int_{x_0}^{x_1} \frac{(1 + y^2)}{y'^2}\, dx.$$

Ans. $y = \sinh (C_1 x + C_2)$

3. Find the extremals of the functional

$$I[y(x)] = \int_{x_0}^{x_1} (2xy + y'''^2)\, dx.$$

Ans. $y = \frac{x^7}{7!} + C_1 x^5 + C_2 x^4 + C_3 x^3 + C_4 x^2 + C_5 x + C_6.$

4. Find the Euler-Ostrogradsky equation for the functional

$$I[u(x, y, z)] = \iiint_D \left[\left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2 + 2uf\right] dx\, dy\, dz.$$

Ans. $\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = f(x, y, z).$