CHAPTER 12

THE CALCULUS OF VARIATIONS

65 INTRODUCTION. SOME TYPICAL PROBLEMS OF THE SUBJECT

The calculus of variations has been one of the major branches of analysis for more than two centuries. It is a tool of great power that can be applied to a wide variety of problems in pure mathematics. It can also be used to express the basic principles of mathematical physics in forms of the utmost simplicity and elegance.

The flavor of the subject is easy to grasp by considering a few of its typical problems. Suppose that two points P and Q are given in a plane (Fig. 92). There are infinitely many curves joining these points, and we can ask which of these curves is the shortest. The intuitive answer is of course a straight line. We can also ask which curve will generate the surface of revolution of smallest area when revolved about the x-axis, and in this case the answer is far from clear. If we think of a typical curve as a frictionless wire in a vertical plane, then another nontrivial problem is that of finding the curve down which a bead will slide from P to Q in the shortest time. This is the famous brachistochrone problem of John Bernoulli, which we discussed in Section 6. Intuitive answers to such questions are quite rare, and the calculus of variations provides a uniform analytical method for dealing with situations of this kind.

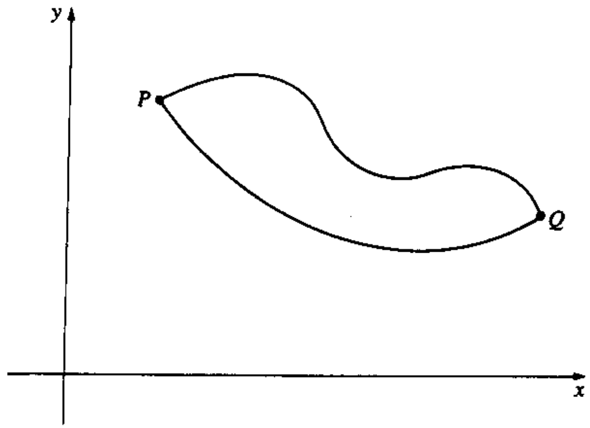

FIGURE 92

Every student of elementary calculus is familiar with the problem of finding points at which a function of a single variable has maximum or minimum values. The above problems show that in the calculus of variations we consider some quantity (arc length, surface area, time of descent) that depends on an entire curve, and we seek the curve that minimizes the quantity in question. The calculus of variations also deals with minimum problems depending on surfaces. For example, if a circular wire is bent in any manner and dipped into a soap solution, then the soap film spanning the wire will assume the shape of the surface of smallest area bounded by the wire. The mathematical problem is to find the surface from this minimum property and the known shape of the wire.

In addition, the calculus of variations has played an important role as a unifying influence in mechanics and as a guide in the mathematical interpretation of many physical phenomena. For instance, it has been found that if the configuration of a system of moving particles is governed by their mutual gravitational attractions, then their actual paths will be minimizing curves for the integral, with respect to time, of the difference between the kinetic and potential energies of the system. This far-reaching statement of classical mechanics is known as *Hamilton's principle* after its discoverer. Also, in modern physics, Einstein made extensive use of the calculus of variations in his work on general relativity, and Schrödinger used it to discover his famous wave equation, which is one of the cornerstones of quantum mechanics.

A few of the problems of the calculus of variations are very old, and were considered and partly solved by the ancient Greeks. The invention of ordinary calculus by Newton and Leibniz stimulated the study of a number of variational problems, and some of these were solved by ingenious special methods. However, the subject was launched as a coherent branch of analysis by Euler in 1744, with his discovery of the basic differential equation for a minimizing curve.

We shall discuss Euler's equation in the next section, but first we observe that each of the problems described in the second paragraph of this section is a special case of the following more general problem. Let P and Q have coordinates (x_1,y_1) and (x_2,y_2), and consider the family of functions

$$y = y(x) \tag{1}$$

that satisfy the boundary conditions $y(x_1) = y_1$ and $y(x_2) = y_2$—that is, the graph of (1) must join P and Q. Then we wish to find the function in this family that minimizes an integral of the form

$$I(y) = \int_{x_1}^{x_2} f(x,y,y')\,dx. \tag{2}$$

To see that this problem indeed contains the others, we note that the length of the curve (1) is

$$\int_{x_1}^{x_2} \sqrt{1 + (y')^2}\,dx, \tag{3}$$

and that the area of the surface of revolution obtained by revolving it about the x-axis is

$$\int_{x_1}^{x_2} 2\pi y\sqrt{1 + (y')^2}\,dx. \tag{4}$$

In the case of the curve of quickest descent, it is convenient to invert the coordinate system and take the point P at the origin, as in Fig. 93. Since the speed $v = ds/dt$ is given by $v = \sqrt{2gy}$, the total time of descent is

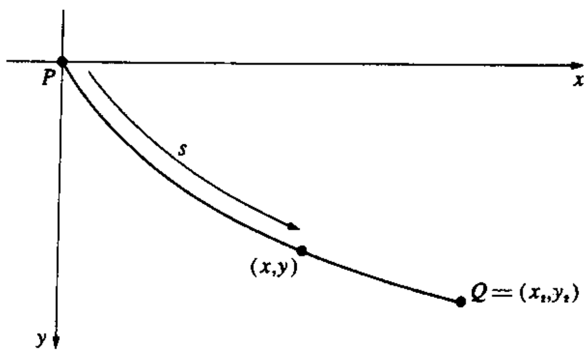

FIGURE 93

the integral of ds/v and the integral to be minimized is

$$\int_{x_1}^{x_2} \frac{\sqrt{1 + (y')^2}}{\sqrt{2gy}}\,dx. \tag{5}$$

Accordingly, the function $f(x,y,y')$ occurring in (2) has the respective forms $\sqrt{1 + (y')^2}$, $2\pi y\sqrt{1 + (y')^2}$ and $\sqrt{1 + (y')^2}/\sqrt{2gy}$ in our three problems.

It is necessary to be somewhat more precise in formulating the basic problem of minimizing the integral (2). First, we will always assume that the function $f(x,y,y')$ has continuous partial derivatives of the second order with respect to x, y, and y'. The next question is, What types of functions (1) are to be allowed? The integral (2) is a well-defined real number whenever the integrand is continuous as a function of x, and for this it suffices to assume that $y'(x)$ is continuous. However, in order to guarantee the validity of the operations we will want to perform, it is convenient to restrict outselves once and for all to considering only unknown functions $y(x)$ that have continuous second derivatives and satisfy the given boundary conditions $y(x_1) = y_1$ and $y(x_2) = y_2$. Functions of this kind will be called *admissible*. We can imagine a competition which only admissible functions are allowed to enter, and the problem is to select from this family the function or functions that yield the smallest value for I.

In spite of these remarks, we will not be seriously concerned with issues of mathematical rigor. Our point of view is deliberately naive, and our sole purpose is to reach the interesting applications as quickly and simply as possible. The reader who wishes to explore the very extensive theory of the subject can readily do so in the systematic treatises.[1]

66 EULER'S DIFFERENTIAL EQUATION FOR AN EXTREMAL

Assuming that there exists an admissible function $y(x)$ that minimizes the integral

$$I = \int_{x_1}^{x_2} f(x,y,y')\,dx, \tag{1}$$

how do we find this function? We shall obtain a differential equation for

[1] See, for example, I. M. Gelfand and S. V. Fomin, *Calculus of Variations*, Prentice-Hall, Englewood Cliffs, N.J., 1963; G. M. Ewing, *Calculus of Variations with Applications*, Norton, New York, 1969; or C. Carathéodory, *Calculus of Variations and Partial Differential Equations of the First Order, Part II: Calculus of Variations*, Holden-Day, San Francisco, 1967.

$y(x)$ by comparing the values of I that correspond to neighboring admissible functions. The central idea is that since $y(x)$ gives a minimum value to I, I will increase if we "disturb" $y(x)$ slightly. These disturbed functions are constructed as follows.

Let $\eta(x)$ be any function with the properties that $\eta''(x)$ is continuous and

$$\eta(x_1) = \eta(x_2) = 0. \tag{2}$$

If α is a small parameter, then

$$\bar{y}(x) = y(x) + \alpha\eta(x) \tag{3}$$

represents a one-parameter family of admissible functions. The vertical deviation of a curve in this family from the minimizing curve $y(x)$ is $\alpha\eta(x)$, as shown in Fig. 94.[2] The significance of (3) lies in the fact that for each family of this type, that is, for each choice of the function $\eta(x)$, the minimizing function $y(x)$ belongs to the family and corresponds to the value of the parameter $\alpha = 0$.

Now, with $\eta(x)$ fixed, we substitute $\bar{y}(x) = y(x) + \alpha\eta(x)$ and

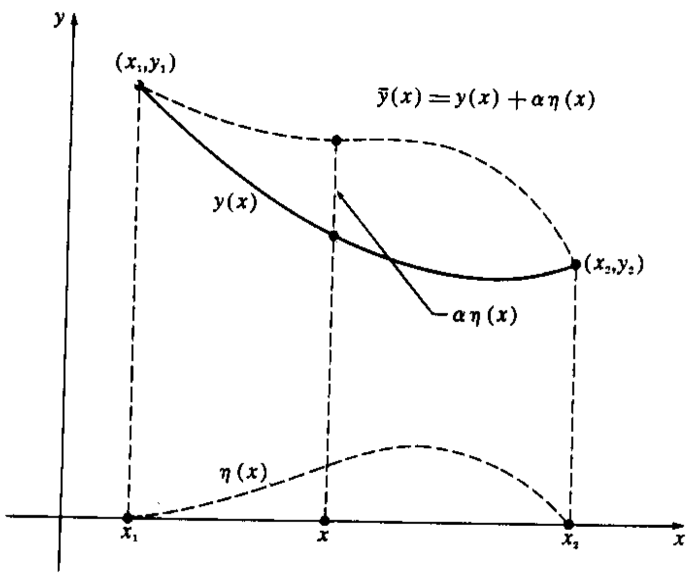

FIGURE 94

[2] The difference $\bar{y} - y = \alpha\eta$ is called the *variation* of the function y and is usually denoted by δy. This notation can be developed into a useful formalism (which we do not discuss) and is the source of the name *calculus of variations.*

$\bar{y}'(x) = y'(x) + \alpha\eta'(x)$ into the integral (1), and get a function of α,

$$\begin{aligned} I(\alpha) &= \int_{x_1}^{x_2} f(x,\bar{y},\bar{y}')\,dx \\ &= \int_{x_1}^{x_2} f[x,y(x) + \alpha\eta(x), y'(x) + \alpha\eta'(x)]\,dx. \end{aligned} \tag{4}$$

When $\alpha = 0$, formula (3) yields $\bar{y}(x) = y(x)$; and since $y(x)$ minimizes the integral, we know that $I(\alpha)$ must have a minimum when $\alpha = 0$. By elementary calculus, a necessary condition for this is the vanishing of the derivative $I'(\alpha)$ when $\alpha = 0$: $I'(0) = 0$. The derivative $I'(\alpha)$ can be computed by differentiating (4) under the integral sign, that is,

$$I'(\alpha) = \int_{x_1}^{x_2} \frac{\partial}{\partial\alpha} f(x,\bar{y},\bar{y}')\,dx. \tag{5}$$

By the chain rule for differentiating functions of several variables, we have

$$\begin{aligned} \frac{\partial}{\partial\alpha} f(x,\bar{y},\bar{y}') &= \frac{\partial f}{\partial x}\frac{\partial x}{\partial\alpha} + \frac{\partial f}{\partial\bar{y}}\frac{\partial\bar{y}}{\partial\alpha} + \frac{\partial f}{\partial\bar{y}'}\frac{\partial\bar{y}'}{\partial\alpha} \\ &= \frac{\partial f}{\partial\bar{y}}\eta(x) + \frac{\partial f}{\partial\bar{y}'}\eta'(x), \end{aligned}$$

so (5) can be written as

$$I'(\alpha) = \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial\bar{y}}\eta(x) + \frac{\partial f}{\partial\bar{y}'}\eta'(x)\right] dx. \tag{6}$$

Now $I'(0) = 0$, so putting $\alpha = 0$ in (6) yields

$$\int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y}\eta(x) + \frac{\partial f}{\partial y'}\eta'(x)\right] dx = 0. \tag{7}$$

In this equation the derivative $\eta'(x)$ appears along with the function $\eta(x)$. We can eliminate $\eta'(x)$ by integrating the second term by parts, which gives

$$\begin{aligned} \int_{x_1}^{x_2} \frac{\partial f}{\partial y'}\eta'(x)\,dx &= \left[\eta(x)\frac{\partial f}{\partial y'}\right]_{x_1}^{x_2} - \int_{x_1}^{x_2} \eta(x)\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx \\ &= -\int_{x_1}^{x_2} \eta(x)\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) dx \end{aligned}$$

by virtue of (2). We can therefore write (7) in the form

$$\int_{x_1}^{x_2} \eta(x)\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right] dx = 0. \tag{8}$$

Our reasoning up to this point is based on a fixed choice of the function $\eta(x)$. However, since the integral in (8) must vanish for *every* such function, we at once conclude that the expression in brackets must also vanish. This yields

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) - \frac{\partial f}{\partial y} = 0, \tag{9}$$

which is *Euler's equation.*[3]

It is important to have a clear understanding of the exact nature of our conclusion: namely, if $y(x)$ is an admissible function that minimizes the integral (1), then y satisfies Euler's equation. Suppose an admissible function y can be found that satisfies this equation. Does this mean that y minimizes I? Not necessarily. The situation is similar to that in elementary calculus, where a function $g(x)$ whose derivative is zero at a point x_0 may have a maximum, a minimum, or a point of inflection at x_0. When no distinctions are made, these cases are often called *stationary values* of $g(x)$, and the points x_0 at which they occur are *stationary points.* In the same way, the condition $I'(0) = 0$ can perfectly well indicate a maximum or point of inflection for $I(\alpha)$ at $\alpha = 0$, instead of a minimum. Thus it is customary to call any admissible solution of Euler's equation a *stationary function* or *stationary curve,* and to refer to the corresponding value of the integral (1) as a *stationary value* of this integral—without committing ourselves as to which of the several possibilities actually occurs. Furthermore, solutions of Euler's equation which are unrestricted by the boundary conditions are called *extremals.*

In calculus we use the second derivative to give sufficient conditions distinguishing one type of stationary value from another. Similar sufficient conditions are available in the calculus of variations, but since these are quite complicated, we will not consider them here. In actual practice, the geometry or physics of the problem under discussion often makes it possible to determine whether a particular stationary function maximizes or minimizes the integral (or neither). The reader who is interested in sufficient conditions and other theoretical problems will find adequate discussions in the books mentioned in Section 65.

As it stands, Euler's equation (9) is not very illuminating. In order to interpret it and convert it into a useful tool, we begin by emphasizing

[3] In more detail, the indirect argument leading to (9) is as follows. Assume that the bracketed function in (8) is not zero (say, positive) at some point $x = a$ in the interval. Since this function is continuous, it will be positive throughout some subinterval about $x = a$. Choose an $\eta(x)$ that is positive inside the subinterval and zero outside. For this $\eta(x)$, the integral in (8) will be positive—which is a contradiction. When this argument is formalized, the resulting statement is known as the *fundamental lemma of the calculus of variations.*

that the partial derivatives $\partial f/\partial y$ and $\partial f/\partial y'$ are computed by treating x, y, and y' as independent variables. In general, however, $\partial f/\partial y'$ is a function of x explicitly, and also implicitly through y and y', so the first term in (9) can be written in the expanded form

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial y'}\right) + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial y'}\right)\frac{dy}{dx} + \frac{\partial}{\partial y'}\left(\frac{\partial f}{\partial y'}\right)\frac{dy'}{dx}.$$

Accordingly, Euler's equation is

$$f_{y'y'}\frac{d^2y}{dx^2} + f_{y'y}\frac{dy}{dx} + (f_{y'x} - f_y) = 0. \tag{10}$$

This equation is of the second order unless $f_{y'y'} = 0$, so in general the extremals—its solutions—constitute a two-parameter family of curves; and among these, the stationary functions are those in which the two parameters are chosen to fit the given boundary conditions. A second order nonlinear equation like (10) is usually impossible to solve, but fortunately many applications lead to special cases that can be solved.

CASE A. If x and y are missing from the function f, then Euler's equation reduces to

$$f_{y'y'}\frac{d^2y}{dx^2} = 0;$$

and if $f_{y'y'} \neq 0$, we have $d^2y/dx^2 = 0$ and $y = c_1x + c_2$, so the extremals are all straight lines.

CASE B. If y is missing from the function f, then Euler's equation becomes

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

and this can be integrated at once to yield the first order equation

$$\frac{\partial f}{\partial y'} = c_1$$

for the extremals.

CASE C. If x is missing from the function f, then Euler's equation can be integrated to

$$\frac{\partial f}{\partial y'}y' - f = c_1.$$

This follows from the identity

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}y' - f\right) = y'\left[\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) - \frac{\partial f}{\partial y}\right] - \frac{\partial f}{\partial x},$$

since $\partial f/\partial x = 0$ and the expression in brackets on the right is zero by Euler's equation.

We now apply this machinery to the three problems formulated in Section 65.

Example 1. To find the shortest curve joining two points (x_1,y_1) and (x_2,y_2)—which we know intuitively to be a straight line—we must minimize the arc length integral

$$I = \int_{x_1}^{x_2} \sqrt{1 + (y')^2}\, dx.$$

The variables x and y are missing from $f(y') = \sqrt{1 + (y')^2}$, so this problem falls under Case A. Since

$$f_{y'y'} = \frac{\partial^2 f}{\partial y'^2} = \frac{1}{[1 + (y')^2]^{3/2}} \neq 0,$$

Case A tells us that the extremals are the two-parameter family of straight lines $y = c_1 x + c_2$. The boundary conditions yield

$$y - y_1 = \frac{y_2 - y_1}{x_2 - x_1}(x - x_1) \tag{11}$$

as the stationary curve, and this is of course the straight line joining the two points. It should be noted that this analysis shows only that if I has a stationary value, then the corresponding stationary curve must be the straight line (11). However, it is clear from the geometry that I has no maximizing curve but does have a minimizing curve, so we conclude in this way that (11) actually is the shortest curve joining our two points.

In this example we arrived at an obvious conclusion by analytical means. A much more difficult and interesting problem is that of finding the shortest curve joining two fixed points on a given surface and lying entirely on that surface. These curves are called *geodesics*, and the study of their properties is one of the focal points of the branch of mathematics known as differential geometry.

Example 2. To find the curve joining the points (x_1,y_1) and (x_2,y_2) that yields a surface of revolution of minimum area when revolved about the x-axis, we must minimize

$$I = \int_{x_1}^{x_2} 2\pi y\sqrt{1 + (y')^2}\, dx. \tag{12}$$

The variable x is missing from $f(y,y') = 2\pi y\sqrt{1 + (y')^2}$, so Case C tells us that Euler's equation becomes

$$\frac{y(y')^2}{\sqrt{1 + (y')^2}} - y\sqrt{1 + (y')^2} = c_1,$$

which simplifies to

$$c_1 y' = \sqrt{y^2 - c_1^2}.$$

On separating variables and integrating, we get

$$x = c_1 \int \frac{dy}{\sqrt{y^2 - c_1^2}} = c_1 \log\left(\frac{y + \sqrt{y^2 - c_1^2}}{c_1}\right) + c_2,$$

and solving for y gives

$$y = c_1 \cosh\left(\frac{x - c_2}{c_1}\right). \tag{13}$$

The extremals are therefore catenaries, and the required minimal surface—if it exists—must be obtained by revolving a catenary. The next problem is that of seeing whether the parameters c_1 and c_2 can indeed be chosen so that the curve (13) joins the points (x_1,y_1) and (x_2,y_2).

The choosing of these parameters turns out to be curiously complicated. If the curve (13) is made to pass through the first point (x_1,y_1), then one parameter is left free. Two members of this one-parameter family are shown in Fig. 95. It can be proved that all such curves are tangent to the

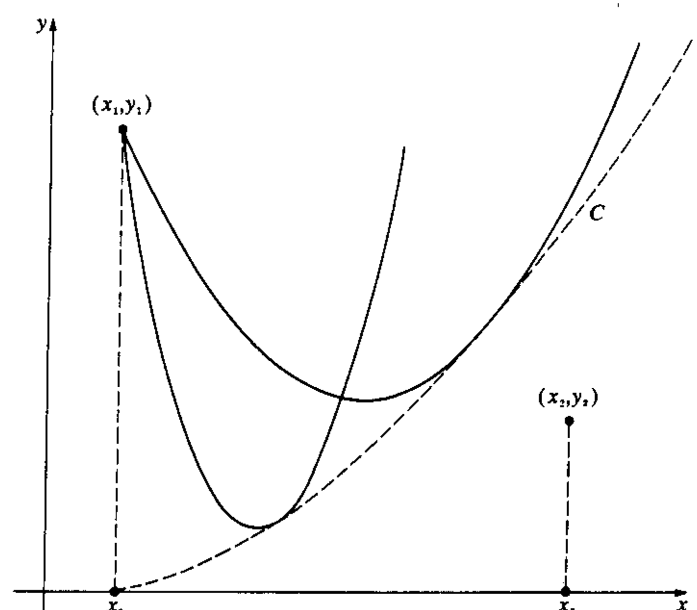

FIGURE 95

dashed curve C, so no curve in the family crosses C. Thus, when the second point (x_2,y_2) is below C, as in Fig. 95, there is no catenary through both points and no stationary function exists. In this case it is found that smaller and smaller surfaces are generated by curves that approach the dashed line from (x_1,y_1) to $(x_1,0)$ to $(x_2,0)$ to (x_2,y_2), so no admissible curve can generate a minimal surface. When the second point lies above C, there are two catenaries through the points, and hence two stationary functions, but only the upper catenary generates a minimal surface. Finally, when the second point is on C, there is only one stationary function but the surface it generates is not minimal.[4]

Example 3. To find the curve of quickest descent in Fig. 93, we must minimize

$$I = \int_{x_1}^{x_2} \frac{\sqrt{1+(y')^2}}{\sqrt{2gy}}\,dx.$$

Again the variable x is missing from the function $f(y,y') = \sqrt{1+(y')^2}/\sqrt{2gy}$, so by Case C, Euler's equation becomes

$$\frac{(y')^2}{\sqrt{y}\sqrt{1+(y')^2}} - \frac{\sqrt{1+(y')^2}}{\sqrt{y}} = c_1.$$

This reduces to

$$y[1+(y')^2] = c,$$

which is precisely the differential equation 6-(4) arrived at in our earlier discussion of this famous problem. Its solution is given in Section 6. The resulting stationary curve is the cycloid

$$x = a(\theta - \sin\theta) \qquad \text{and} \qquad y = a(1-\cos\theta) \tag{14}$$

generated by a circle of radius a rolling under the x axis, where a is chosen so that the first inverted arch passes through the point (x_2,y_2) in Fig. 93. As before, this argument shows only that if I has a minimum, then the corresponding stationary curve must be the cycloid (14). However, it is reasonably clear from physical considerations that I has no maximizing curve but does have a minimizing curve, so this cycloid actually minimizes the time of descent.

We conclude this section with an easy but important extension of our treatment of the integral (1). This integral represents variational problems of the simplest type because it involves only one unknown function. However, some of the situations we will encounter below are not quite so simple, for they lead to integrals depending on two or more unknown functions.

[4] A full discussion of these statements, with proofs, can be found in Chapter IV of G. A. Bliss's book *Calculus of Variations*, Carus Monograph no. 1, Mathematical Association of America, 1925.

For example, suppose we want to find conditions necessarily satisfied by two functions $y(x)$ and $z(x)$ that give a stationary value to the integral

$$I = \int_{x_1}^{x_2} f(x,y,z,y',z')\,dx, \tag{15}$$

where the boundary values $y(x_1)$, $z(x_1)$ and $y(x_2)$, $z(x_2)$ are specified in advance. Just as before, we introduce functions $\eta_1(x)$ and $\eta_2(x)$ that have continuous second derivatives and vanish at the endpoints. From these we form the neighboring functions $\bar{y}(x) = y(x) + \alpha\eta_1(x)$ and $\bar{z}(x) = z(x) + \alpha\eta_2(x)$, and then consider the function of α defined by

$$I(\alpha) = \int_{x_1}^{x_2} f(x, y+\alpha\eta_1, z+\alpha\eta_2, y'+\alpha\eta_1', z'+\alpha\eta_2')\,dx. \tag{16}$$

Again, if $y(x)$ and $z(x)$ are stationary functions we must have $I'(0) = 0$, so by computing the derivative of (16) and putting $\alpha = 0$ we get

$$\int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y}\eta_1 + \frac{\partial f}{\partial z}\eta_2 + \frac{\partial f}{\partial y'}\eta_1' + \frac{\partial f}{\partial z'}\eta_2'\right) dx = 0,$$

or, if the terms involving η_1' and η_2' are integrated by parts,

$$\int_{x_1}^{x_2} \left\{\eta_1(x)\left[\frac{\partial f}{\partial y} - \frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right)\right] + \eta_2(x)\left[\frac{\partial f}{\partial z} - \frac{d}{dx}\left(\frac{\partial f}{\partial z'}\right)\right]\right\} dx = 0. \tag{17}$$

Finally, since (17) must hold for all choices of the functions $\eta_1(x)$ and $\eta_2(x)$, we are led at once to Euler's equations

$$\frac{d}{dx}\left(\frac{\partial f}{\partial y'}\right) - \frac{\partial f}{\partial y} = 0 \qquad \text{and} \qquad \frac{d}{dx}\left(\frac{\partial f}{\partial z'}\right) - \frac{\partial f}{\partial z} = 0. \tag{18}$$

Thus, to find the extremals of our problem, we must solve the system (18). Needless to say, a system of intractable equations is harder to solve than only one; but if (18) can be solved, then the stationary functions are determined by fitting the resulting solutions to the given boundary conditions. Similar considerations apply without any essential change to integrals like (15) which involve more than two unknown functions.

PROBLEMS

1. Find the extremals for the integral (1) if the integrand is

(a) $\dfrac{\sqrt{1+(y')^2}}{y}$;

(b) $y^2 - (y')^2$.

2. Find the stationary function of

$$\int_0^4 [xy' - (y')^2]\,dx$$

which is determined by the boundary conditions $y(0) = 0$ and $y(4) = 3$.

3. When the integrand in (1) is of the form

$$a(x)(y')^2 + 2b(x)yy' + c(x)y^2,$$

show that Euler's equation is a second order linear differential equation.

4. If P and Q are two points in a plane, then in terms of polar coordinates, the length of a curve from P to Q is

$$\int_P^Q ds = \int_P^Q \sqrt{dr^2 + r^2\,d\theta^2}.$$

Find the polar equation of a straight line by minimizing this integral
(a) with θ as the independent variable;
(b) with r as the independent variable.

5. Consider two points P and Q on the surface of the sphere $x^2 + y^2 + z^2 = a^2$, and coordinatize this surface by means of the spherical coordinates θ and ϕ, where $x = a \sin\phi \cos\theta$, $y = a \sin\phi \sin\theta$, and $z = a\cos\phi$. Let $\theta = F(\phi)$ be a curve lying on the surface and joining P and Q. Show that the shortest such curve (a geodesic) is an arc of a great circle, that is, that it lies on a plane through the center. *Hint*: Express the length of the curve in the form

$$\int_P^Q ds = \int_P^Q \sqrt{dx^2 + dy^2 + dz^2}$$

$$= a\int_P^Q \sqrt{1 + \left(\frac{d\theta}{d\phi}\right)^2 \sin^2\phi}\, d\phi,$$

solve the corresponding Euler equation for θ, and convert the result back into rectangular coordinates.

6. Prove that any geodesic on the right circular cone $z^2 = a^2(x^2 + y^2)$, $z \geq 0$, has the following property: If the cone is cut along a generator and flattened into a plane, then the geodesic becomes a straight line. *Hint:* Represent the cone parametrically by means of the equations

$$x = \frac{r\cos(\theta\sqrt{1+a^2})}{\sqrt{1+a^2}}, \qquad y = \frac{r\sin(\theta\sqrt{1+a^2})}{\sqrt{1+a^2}}, \qquad z = \frac{ar}{\sqrt{1+a^2}};$$

show that the parameters r and θ represent ordinary polar coordinates on the flattened cone; and show that a geodesic $r = r(\theta)$ is a straight line in these polar coordinates.

7. If the curve $y = g(z)$ is revolved about the z-axis, then the resulting surface of revolution has $x^2 + y^2 = g(z)^2$ as its equation. A convenient parametric representation of this surface is given by

$$x = g(z)\cos\theta, \qquad y = g(z)\sin\theta, \qquad z = z,$$

where θ is the polar angle in the xy-plane. Show that a geodesic $\theta = \theta(z)$ on this surface has

$$\theta = c_1 \int \frac{\sqrt{1 + [g'(z)]^2}}{g(z)\sqrt{g(z)^2 - c_1^2}}\,dz + c_2$$

as its equation.

8. If the surface of revolution in Problem 7 is a right circular cylinder, show that every geodesic of the form $\theta = \theta(z)$ is a helix or a generator.

67 ISOPERIMETRIC PROBLEMS

The ancient Greeks proposed the problem of finding the closed plane curve of given length that encloses the largest area. They called this the *isoperimetric problem*, and were able to show in a more or less rigorous manner that the obvious answer—a circle—is correct.[5] If the curve is expressed parametrically by $x = x(t)$ and $y = y(t)$, and is traversed once counterclockwise as t increases from t_1 to t_2, then the enclosed area is known to be

$$A = \frac{1}{2}\int_{t_1}^{t_2} \left(x\frac{dy}{dt} - y\frac{dx}{dt}\right) dt, \tag{1}$$

which is an integral depending on two unknown functions.[6] Since the length of the curve is

$$L = \int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2}\, dt, \tag{2}$$

the problem is to maximize (1) subject to the side condition that (2) must have a constant value. The term *isoperimetric problem* is usually extended to include the general case of finding extremals for one integral subject to any constraint requiring a second integral to take on a prescribed value.

We will also consider finite side conditions, which do not involve integrals or derivatives. For example, if

$$G(x,y,z) = 0 \tag{3}$$

is a given surface, then a curve on this surface is determined parametrically by three functions $x = x(t)$, $y = y(t)$, and $z = z(t)$ that satisfy equation (3), and the problem of finding geodesics amounts to the

[5] See B. L. van der Waerden, *Science Awakening*, pp. 268–269, Oxford University Press, London, 1961; also, G. Polya, *Induction and Analogy in Mathematics*, Chapter 10, Princeton University Press, Princeton, N.J., 1954.

[6] Formula (1) is a special case of Green's theorem. Also, see Problem 1.

problem of minimizing the arc length integral

$$\int_{t_1}^{t_2} \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}\, dt \tag{4}$$

subject to the side condition (3).

Lagrange multipliers. It is necessary to begin by considering some problems in elementary calculus that are quite similar to isoperimetric problems. For example, suppose we want to find the points (x,y) that yield stationary values for a function $z = f(x,y)$, where, however, the variables x and y are not independent but are constrained by a side condition

$$g(x,y) = 0. \tag{5}$$

The usual procedure is to arbitrarily designate one of the variables x and y in (5) as independent, say x, and the other as dependent on it, so that dy/dx can be computed from

$$\frac{\partial g}{\partial x} + \frac{\partial g}{\partial y}\frac{dy}{dx} = 0.$$

We next use the fact that since z is now a function of x alone, $dz/dx = 0$ is a necessary condition for z to have a stationary value, so

$$\frac{dz}{dx} = \frac{\partial f}{\partial x} + \frac{\partial f}{\partial y}\frac{dy}{dx} = 0$$

or

$$\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}\frac{\partial g/\partial x}{\partial g/\partial y} = 0. \tag{6}$$

On solving (5) and (6) simultaneously, we obtain the required points (x,y).[7]

One drawback to this approach is that the variables x and y occur symmetrically but are treated unsymmetrically. It is possible to solve the same problem by a different and more elegant method that also has many practical advantages. We form the function

$$F(x,y,\lambda) = f(x,y) + \lambda g(x,y)$$

and investigate its *unconstrained* stationary values by means of the

[7] In very simple cases, of course, we can solve (5) for y as a function of x and insert this in $z = f(x,y)$, which gives z as an explicit function of x; and all that remains is to compute dz/dx, solve the equation $dz/dx = 0$, and find the corresponding y's.

necessary conditions

$$\begin{aligned} \frac{\partial F}{\partial x} &= \frac{\partial f}{\partial x} + \lambda\frac{\partial g}{\partial x} = 0,\\ \frac{\partial F}{\partial y} &= \frac{\partial f}{\partial y} + \lambda\frac{\partial g}{\partial y} = 0,\\ \frac{\partial F}{\partial \lambda} &= g(x,y) = 0. \end{aligned} \tag{7}$$

If λ is eliminated from the first two of these equations, then the system clearly reduces to

$$\frac{\partial f}{\partial x} - \frac{\partial f}{\partial y}\frac{\partial g/\partial x}{\partial g/\partial y} = 0 \qquad \text{and} \qquad g(x,y) = 0,$$

and this is the system obtained in the above paragraph. It should be observed that this technique (solving the system (7) for x and y) solves the given problem in a way that has two major features important for theoretical work: it does not disturb the symmetry of the problem by making an arbitrary choice of the independent variable; and it removes the side condition at the small expense of introducing λ as another variable. The parameter λ is called a *Lagrange multiplier*, and this method is known as the method of Lagrange multipliers.[8] This discussion extends in an obvious manner to problems involving functions of more than two variables with several side conditions.

Integral side conditions. Here we want to find the differential equation that must be satisfied by a function $y(x)$ that gives a stationary value to the integral

$$I = \int_{x_1}^{x_2} f(x,y,y')\, dx, \tag{8}$$

where y is subject to the side condition

$$J = \int_{x_1}^{x_2} g(x,y,y')\, dx = c \tag{9}$$

and assumes prescribed values $y(x_1) = y_1$ and $y(x_2) = y_2$ at the end-points. As before, we assume that $y(x)$ is the actual stationary function and disturb it slightly to find the desired analytic condition. However, this problem cannot be attacked by our earlier method of considering neighboring functions of the form $\bar{y}(x) = y(x) + \alpha\eta(x)$, for in general

[8] A brief account of Lagrange is given in Appendix A.

these will not maintain the second integral J at the constant value c. Instead, we consider a two-parameter family of neighboring functions

$$\bar{y}(x) = y(x) + \alpha_1\eta_1(x) + \alpha_2\eta_2(x), \tag{10}$$

where $\eta_1(x)$ and $\eta_2(x)$ have continuous second derivatives and vanish at the endpoints. The parameters α_1 and α_2 are not independent, but are related by the condition that

$$J(\alpha_1,\alpha_2) = \int_{x_1}^{x_2} g(x,\bar{y},\bar{y}')\,dx = c. \tag{11}$$

Our problem is then reduced to that of finding necessary conditions for the function

$$I(\alpha_1,\alpha_2) = \int_{x_1}^{x_2} f(x,\bar{y},\bar{y}')\,dx \tag{12}$$

to have a stationary value at $\alpha_1 = \alpha_2 = 0$, where α_1 and α_2 satisfy (11). This situation is made to order for the method of Lagrange multipliers. We therefore introduce the function

$$\begin{aligned} K(\alpha_1,\alpha_2,\lambda) &= I(\alpha_1,\alpha_2) + \lambda J(\alpha_1,\alpha_2) \\ &= \int_{x_1}^{x_2} F(x,\bar{y},\bar{y}')\,dx, \end{aligned} \tag{13}$$

where

$$F = f + \lambda g,$$

and investigate its unconstrained stationary value at $\alpha_1 = \alpha_2 = 0$ by means of the necessary conditions

$$\frac{\partial K}{\partial \alpha_1} = \frac{\partial K}{\partial \alpha_2} = 0 \qquad \text{when } \alpha_1 = \alpha_2 = 0. \tag{14}$$

If we differentiate (13) under the integral sign and use (10), we get

$$\frac{\partial K}{\partial \alpha_i} = \int_{x_1}^{x_2}\left[\frac{\partial F}{\partial \bar{y}}\eta_i(x) + \frac{\partial F}{\partial \bar{y}'}\eta_i'(x)\right]dx \qquad \text{for } i = 1, 2;$$

and setting $\alpha_1 = \alpha_2 = 0$ yields

$$\int_{x_1}^{x_2}\left[\frac{\partial F}{\partial y}\eta_i(x) + \frac{\partial F}{\partial y'}\eta_i'(x)\right]dx = 0$$

by virtue of (14). After the second term is integrated by parts, this becomes

$$\int_{x_1}^{x_2} \eta_i(x)\left[\frac{\partial F}{\partial y} - \frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right)\right]dx = 0. \tag{15}$$

Since $\eta_1(x)$ and $\eta_2(x)$ are both arbitrary, the two conditions embodied in (15) amount to only one condition, and as usual we conclude that the stationary function $y(x)$ must satisfy Euler's equation

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0. \tag{16}$$

The solutions of this equation (the extremals of our problem) involve three undetermined parameters: two constants of integration, and the Lagrange multiplier λ. The stationary function is then selected from these extremals by imposing the two boundary conditions and giving the integral J its prescribed value c.

In the case of integrals that depend on two or more functions, this result can be extended in the same way as in the previous section. For example, if

$$I = \int_{x_1}^{x_2} f(x,y,z,y',z')\,dx$$

has a stationary value subject to the side condition

$$J = \int_{x_1}^{x_2} g(x,y,z,y',z')\,dx = c,$$

then the stationary functions $y(x)$ and $z(x)$ must satisfy the system of equations

$$\frac{d}{dx}\left(\frac{\partial F}{\partial y'}\right) - \frac{\partial F}{\partial y} = 0 \qquad \text{and} \qquad \frac{d}{dx}\left(\frac{\partial F}{\partial z'}\right) - \frac{\partial F}{\partial z} = 0, \tag{17}$$

where $F = f + \lambda g$. The reasoning is similar to that already given, and we omit the details.

Example 1. We shall find the curve of fixed length L that joins the points (0,0) and (1,0), lies above the x-axis, and encloses the maximum area between itself and the x-axis. This is a restricted version of the original isoperimetric problem in which part of the curve surrounding the area to be maximized is required to be a line segment of length 1. Our problem is to maximize $\int_0^1 y\,dx$ subject to the side condition

$$\int_0^1 \sqrt{1 + (y')^2}\,dx = L$$

and the boundary conditions $y(0) = 0$ and $y(1) = 0$. Here we have $F = y + \lambda\sqrt{1 + (y')^2}$, so Euler's equation is

$$\frac{d}{dx}\left(\frac{\lambda y'}{\sqrt{1 + (y')^2}}\right) - 1 = 0, \tag{18}$$

or, after carrying out the differentiation,

$$\frac{y''}{[1 + (y')^2]^{3/2}} = \frac{1}{\lambda}. \tag{19}$$

In this case no integration is necessary, since (19) tells us at once that the curvature is constant and equals $1/\lambda$. It follows that the required maximizing curve is an arc of a circle (as might have been expected) with radius λ. As an alternate procedure, we can integrate (18) to get

$$\frac{y'}{\sqrt{1+(y')^2}} = \frac{x-c_1}{\lambda}.$$

On solving this for y' and integrating again, we obtain

$$(x-c_1)^2 + (y-c_2)^2 = \lambda^2, \tag{20}$$

which of course is the equation of a circle with radius λ.

Example 2. In Example 1 it is clearly necessary to have $L > 1$. Also, if $L > \pi/2$, the circular arc determined by (20) will not define $y > 0$ as a single-valued function of x. We can avoid these artificial issues by considering curves in parametric form $x = x(t)$ and $y = y(t)$ and by turning our attention to the original isoperimetric problem of maximizing

$$\frac{1}{2}\int_{t_1}^{t_2} (x\dot{y} - y\dot{x})\,dt$$

(where $\dot{x} = dx/dt$ and $\dot{y} = dy/dt$) with the side condition

$$\int_{t_1}^{t_2} \sqrt{\dot{x}^2+\dot{y}^2}\,dt = L.$$

Here we have

$$F = \frac{1}{2}(x\dot{y} + y\dot{x}) + \lambda\sqrt{\dot{x}^2+\dot{y}^2},$$

so the Euler equations (17) are

$$\frac{d}{dt}\left(-\frac{1}{2}y + \frac{\lambda\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right) - \frac{1}{2}\dot{y} = 0$$

and

$$\frac{d}{dt}\left(\frac{1}{2}x + \frac{\lambda\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}}\right) + \frac{1}{2}\dot{x} = 0.$$

These equations can be integrated directly, which yields

$$-y + \frac{\lambda\dot{x}}{\sqrt{\dot{x}^2+\dot{y}^2}} = -c_1 \qquad \text{and} \qquad x + \frac{\lambda\dot{y}}{\sqrt{\dot{x}^2+\dot{y}^2}} = c_2.$$

If we solve for $x - c_2$ and $y - c_1$, square, and add, then the result is

$$(x-c_2)^2 + (y-c_1)^2 = \lambda^2,$$

so the maximizing curve is a circle. This result can be expressed in the following way: if L is the length of a closed plane curve that encloses an area A, then $A \le L^2/4\pi$, with equality if and only if the curve is a circle. A relation of this kind is called an *isoperimetric inequality*.[9]

[9] Students of physics may be interested in the ideas discussed in G. Polya and G. Szegö, *Isoperimetric Inequalities in Mathematical Physics*, Princeton University Press, Princeton, N.J., 1951.

Finite side conditions. At the beginning of this section we formulated the problem of finding geodesics on a given surface

$$G(x,y,z) = 0. \tag{21}$$

We now consider the slightly more general problem of finding a space curve $x = x(t)$, $y = y(t)$, $z = z(t)$ that gives a stationary value to an integral of the form

$$\int_{t_1}^{t_2} f(\dot{x},\dot{y},\dot{z})\,dt, \tag{22}$$

where the curve is required to lie on the surface (21).

Our strategy is to eliminate the side condition (21), and to do this we proceed as follows. There is no loss of generality in assuming that the curve lies on a part of the surface where $G_z \neq 0$. On this part of the surface we can solve (21) for z, which gives $z = g(x,y)$ and

$$\dot{z} = \frac{\partial g}{\partial x}\dot{x} + \frac{\partial g}{\partial y}\dot{y}. \tag{23}$$

When (23) is inserted in (22), our problem is reduced to that of finding unconstrained stationary functions for the integral

$$\int_{t_1}^{t_2} f\left(\dot{x}, \dot{y}, \frac{\partial g}{\partial x}\dot{x} + \frac{\partial g}{\partial y}\dot{y}\right) dt.$$

We know from the previous section that the Euler equations 66-(18) for this problem are

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}} + \frac{\partial f}{\partial \dot{z}}\frac{\partial g}{\partial x}\right) - \frac{\partial f}{\partial \dot{z}}\frac{\partial \dot{z}}{\partial x} = 0,$$

and

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}} + \frac{\partial f}{\partial \dot{z}}\frac{\partial g}{\partial y}\right) - \frac{\partial f}{\partial \dot{z}}\frac{\partial \dot{z}}{\partial y} = 0.$$

It follows from (23) that

$$\frac{\partial \dot{z}}{\partial x} = \frac{d}{dt}\left(\frac{\partial g}{\partial x}\right) \qquad \text{and} \qquad \frac{\partial \dot{z}}{\partial y} = \frac{d}{dt}\left(\frac{\partial g}{\partial y}\right),$$

so the Euler equations can be written in the form

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) + \frac{\partial g}{\partial x}\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{z}}\right) = 0 \qquad \text{and} \qquad \frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}}\right) + \frac{\partial g}{\partial y}\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{z}}\right) = 0.$$

If we now define a function $\lambda(t)$ by

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{z}}\right) = \lambda(t)G_z, \tag{24}$$

and use the relations $\partial g/\partial x = -G_x/G_z$ and $\partial g/\partial y = -G_y/G_z$, then Euler's equations become

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{x}}\right) = \lambda(t)G_x, \tag{25}$$

and

$$\frac{d}{dt}\left(\frac{\partial f}{\partial \dot{y}}\right) = \lambda(t)G_y. \tag{26}$$

Thus a necessary condition for a stationary value is the existence of a function $\lambda(t)$ satisfying equations (24), (25), and (26). On eliminating $\lambda(t)$, we obtain the symmetric equations

$$\frac{(d/dt)(\partial f/\partial \dot{x})}{G_x} = \frac{(d/dt)(\partial f/\partial \dot{y})}{G_y} = \frac{(d/dt)(\partial f/\partial \dot{z})}{G_z}, \tag{27}$$

which together with (21) determine the extremals of the problem. It is worth remarking that equations (24), (25), and (26) can be regarded as the Euler equations for the problem of finding unconstrained stationary functions for the integral

$$\int_{t_1}^{t_2} [f(\dot{x},\dot{y},\dot{z}) + \lambda(t)G(x,y,z)]\, dt.$$

This is very similar to our conclusion for integral side conditions, except that here the multiplier is an undetermined function of t instead of an undetermined constant.

When we specialize this result to the problem of finding geodesics on the surface (21), we have

$$f = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2}.$$

The equations (27) become

$$\frac{(d/dt)(\dot{x}/f)}{G_x} = \frac{(d/dt)(\dot{y}/f)}{G_y} = \frac{(d/dt)(\dot{z}/f)}{G_z}, \tag{28}$$

and the problem is to extract information from this system.

Example 3. If we choose the surface (21) to be the sphere $x^2 + y^2 + z^2 = a^2$, then $G(x,y,z) = x^2 + y^2 + z^2 - a^2$ and (28) is

$$\frac{f\ddot{x} - \dot{x}\dot{f}}{2xf^2} = \frac{f\ddot{y} - \dot{y}\dot{f}}{2yf^2} = \frac{f\ddot{z} - \dot{z}\dot{f}}{2zf^2},$$

which can be rewritten in the form

$$\frac{x\ddot{y} - y\ddot{x}}{x\dot{y} - y\dot{x}} = \frac{\dot{f}}{f} = \frac{y\ddot{z} - z\ddot{y}}{y\dot{z} - z\dot{y}}.$$

If we ignore the middle term, this is

$$\frac{(d/dt)(x\dot{y} - y\dot{x})}{x\dot{y} - y\dot{x}} = \frac{(d/dt)(y\dot{z} - z\dot{y})}{y\dot{z} - z\dot{y}}.$$

One integration gives $x\dot{y} - y\dot{x} = c_1(y\dot{z} - z\dot{y})$ or

$$\frac{\dot{x} + c_1\dot{z}}{x + c_1 z} = \frac{\dot{y}}{y},$$

and a second yields $x + c_1 z = c_2 y$. This is the equation of a plane through the origin, so the geodesics on a sphere are arcs of great circles. A different method of arriving at this conclusion is given in Problem 66-5.

In this example we were able to solve equations (28) quite easily, but in general this task is extremely difficult. The main significance of these equations lies in their connection with the following very important result in mathematical physics: if a particle glides along a surface, free from the action of any external force, then its path is a geodesic. We shall prove this dynamical theorem in Appendix B. For the purpose of this argument it will be convenient to assume that the parameter t is the arc length s measured along the curve, so that $f = 1$ and equations (28) become

$$\frac{d^2x/ds^2}{G_x} = \frac{d^2y/ds^2}{G_y} = \frac{d^2z/ds^2}{G_z}. \tag{29}$$

PROBLEMS

1. Convince yourself of the validity of formula (1) for a closed convex curve like that shown in Fig. 96. *Hint:* What is the geometric meaning of

$$\int_P^Q y\, dx + \int_Q^P y\, dx,$$

where the first integral is taken from right to left along the upper part of the curve and the second from left to right along the lower part?

2. Verify formula (1) for the circle whose parametric equations are $x = a\cos t$ and $y = a\sin t$, $0 \le t \le 2\pi$.
3. Solve the following problems by the method of Lagrange multipliers.
   (a) Find the point on the plane $ax + by + cz = d$ that is nearest the origin. *Hint:* Minimize $w = x^2 + y^2 + z^2$ with the side condition $ax + by + cz - d = 0$.
   (b) Show that the triangle with greatest area A for a given perimeter is equilateral. *Hint:* If x, y, and z are the sides, then $A = \sqrt{s(s-x)(s-y)(s-z)}$ where $s = (x + y + z)/2$.
   (c) If the sum of n positive numbers $x_1, x_2, \ldots, x_n$ has a fixed value s, prove that their product $x_1x_2\cdots x_n$ has s^n/n^n as its maximum value, and

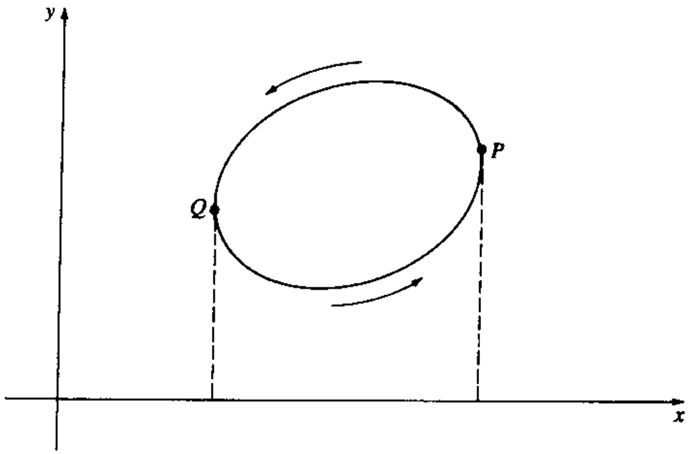

FIGURE 96

conclude from this that the geometric mean of n positive numbers can never exceed their arithmetic mean:

$$\sqrt[n]{x_1 x_2 \cdots x_n} \leq \frac{x_1 + x_2 + \cdots + x_n}{n}.$$

4. A curve in the first quadrant joins (0,0) and (1,0) and has a given area beneath it. Show that the shortest such curve is an arc of a circle.
5. A uniform flexible chain of given length hangs between two points. Find its shape if it hangs in such a way as to minimize its potential energy.
6. Solve the original isoperimetric problem (Example 2) by using polar coordinates. *Hint:* Choose the origin to be any point on the curve and the polar axis to be the tangent line at that point; then maximize

$$\frac{1}{2}\int_0^{\pi} r^2\, d\theta$$

with the side condition that

$$\int_0^{\pi} \sqrt{\left(\frac{dr}{d\theta}\right)^2 + r^2}\, d\theta$$

must be constant.
7. Show that the geodesics on any cylinder of the form $g(x,z) = 0$ make a constant angle with the y-axis.

APPENDIX A. LAGRANGE

Joseph Louis Lagrange (1736–1813) detested geometry but made outstanding discoveries in the calculus of variations and analytical mechanics. He also contributed to number theory and algebra, and fed the stream of thought that later nourished Gauss and Abel. His mathematical career can be viewed as a natural extension of the work of his older and greater contemporary, Euler, which in many respects he carried forward and refined.

Lagrange was born in Turin of mixed French–Italian ancestry. As a boy, his tastes were more classical than scientific; but his interest in mathematics was kindled while he was still in school by reading a paper by Edmund Halley on the uses of algebra in optics. He then began a course of independent study, and progressed so rapidly that at the age of nineteen he was appointed professor of mathematics at the Royal Artillery School in Turin.[10]

Lagrange's contributions to the calculus of variations were among his earliest and most important works. In 1755 he communicated to Euler his method of multipliers for solving isoperimetric problems. These problems had baffled Euler for years, since they lay beyond the reach of his own semigeometrical techniques. Euler was immediately able to answer many questions he had long contemplated; but he replied to Lagrange with admirable kindness and generosity, and withheld his own work from publication "so as not to deprive you of any part of the glory which is your due." Lagrange continued working for a number of years on his analytic version of the calculus of variations, and both he and Euler applied it to many new types of problems, especially in mechanics.

In 1766, when Euler left Berlin for St. Petersburg, he suggested to Frederick the Great that Lagrange be invited to take his place. Lagrange accepted and lived in Berlin for 20 years until Frederick's death in 1786. During this period he worked extensively in algebra and number theory and wrote his masterpiece, the treatise *Mécanique Analytique* (1788), in which he unified general mechanics and made of it, as Hamilton later said, "a kind of scientific poem." Among the enduring legacies of this work are Lagrange's equations of motion, generalized coordinates, and the concept of potential energy (which are all discussed in Appendix B).[11]

Men of science found the atmosphere of the Prussian court rather uncongenial after the death of Frederick, so Lagrange accepted an invitation from Louis XVI to move to Paris, where he was given

[10] See George Sarton's valuable essay, "Lagrange's Personality," *Proc. Am. Phil. Soc.*, vol. 88, pp. 457–496 (1944).

[11] For some interesting views on Lagrangian mechanics (and many other subjects), see S. Bochner, *The Role of Mathematics in the Rise of Science*, pp. 199–207, Princeton University Press, Princeton, N.J., 1966.

apartments in the Louvre. Lagrange was extremely modest and undogmatic for a man of his great gifts; and though he was a friend of aristocrats—and indeed an aristocrat himself—he was respected and held in affection by all parties throughout the turmoil of the French Revolution. His most important work during these years was his leading part in establishing the metric system of weights and measures. In mathematics, he tried to provide a satisfactory foundation for the basic processes of analysis, but these efforts were largely abortive. Toward the end of his life, Lagrange felt that mathematics had reached a dead end, and that chemistry, physics, biology, and other sciences would attract the ablest minds of the future. His pessimism might have been relieved if he had been able to forsee the coming of Gauss and his successors, who made the nineteenth century the richest in the long history of mathematics.

APPENDIX B. HAMILTON'S PRINCIPLE AND ITS IMPLICATIONS

One purpose of the mathematicians of the eighteenth century was to discover a general principle from which Newtonian mechanics could be deduced. In searching for clues, they noted a number of curious facts in elementary physics: for example, that a ray of light follows the quickest path through an optical medium; that the equilibrium shape of a hanging chain minimizes its potential energy; and that soap bubbles assume a shape having the least surface area for a given volume. These facts and others suggested to Euler that nature pursues its diverse ends by the most efficient and economical means, and that hidden simplicities underlie the apparent chaos of phenomena. It was this metaphysical idea that led him to create the calculus of variations as a tool for investigating such questions. Euler's dream was realized almost a century later by Hamilton.

Hamilton's principle. Consider a particle of mass m moving through space under the influence of a force

$$\mathbf{F} = F_1\mathbf{i} + F_2\mathbf{j} + F_3\mathbf{k},$$

and assume that this force is *conservative* in the sense that the work it does in moving the particle from one point to another is independent of the path. It is easy to show that there exists a scalar function $U(x,y,z)$ such that $\partial U/\partial x = F_1$, $\partial U/\partial y = F_2$, and $\partial U/\partial z = F_3$.[12] The function $V = -U$ is called the *potential energy* of the particle, since the change in

[12] In the language of vector analysis, $\mathbf{F}$ is the *gradient of U.*

its value from one point to another is the work done against $\mathbf{F}$ in moving the particle from the first point to the second. Furthermore, if $\mathbf{r}(t) = x(t)\mathbf{i} + y(t)\mathbf{j} + z(t)\mathbf{k}$ is the position vector of the particle, so that

$$\mathbf{v} = \frac{dx}{dt}\mathbf{i} + \frac{dy}{dt}\mathbf{j} + \frac{dz}{dt}\mathbf{k} \quad \text{and} \quad v = \sqrt{\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2}$$

are its velocity and speed, respectively, then $T = mv^2/2$ is its *kinetic energy*.

If the particle is at points P_1 and P_2 at times t_1 and t_2, then we are interested in the path it traverses in moving from P_1 to P_2. The *action* (or *Hamilton's integral*) is defined as

$$A = \int_{t_1}^{t_2} (T - V)\,dt,$$

and in general its value depends on the path along which the particle moves in passing from P_1 to P_2. We will show that the actual path of the particle is one that yields a stationary value for the action A.

The function $L = T - V$ is called the *Lagrangian*, and in the case under consideration it is given by

$$L = \frac{1}{2}m\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right] - V(x,y,z).$$

The integrand of the action is therefore a function of the form $f(x,y,z,dx/dt,dy/dt,dz/dt)$, and if the action has a stationary value, then Euler's equations must be satisfied. These equations are

$$m\frac{d^2x}{dt^2} + \frac{\partial V}{\partial x} = 0, \qquad m\frac{d^2y}{dt^2} + \frac{\partial V}{\partial y} = 0, \qquad m\frac{d^2z}{dt^2} + \frac{\partial V}{\partial z} = 0,$$

and can be written in the form

$$m\frac{d^2\mathbf{r}}{dt^2} = -\frac{\partial V}{\partial x}\mathbf{i} - \frac{\partial V}{\partial y}\mathbf{j} - \frac{\partial V}{\partial z}\mathbf{k} = \mathbf{F}.$$

This is precisely Newton's second law of motion. Thus Newton's law is a necessary condition for the action of the particle to have a stationary value. Since Newton's law governs the motion of the particle, we have the following conclusion.

> **Hamilton's principle.** *If a particle moves from a point P_1 to a point P_2 in a time interval $t_1 \le t \le t_2$, then the actual path it follows is one for which the action assumes a stationary value.*

It is quite easy to give simple examples in which the actual path of a particle maximizes the action. However, if the time interval is sufficiently short, then it can be shown that the action is necessarily a minimum. In this form, Hamilton's principle is sometimes called the *principle of least action*, and can be loosely interpreted as saying that nature tends to equalize the kinetic and potential energies throughout the motion.

In the above discussion we assumed Newton's law and deduced Hamilton's principle as a consequence. The same argument shows that Newton's law follows from Hamilton's principle, so these two approaches to the dynamics of a particle—the vectorial and the variational—are equivalent to one another. This result emphasizes the essential characteristic of variational principles in physics: they express the pertinent physical laws in terms of energy alone, without reference to any coordinate system.

The argument we have given extends at once to a system of n particles of masses m_i, with position vectors $\mathbf{r}_i(t) = x_i(t)\mathbf{i} + y_i(t)\mathbf{j} + z_i(t)\mathbf{k}$, which are moving under the influence of conservative forces $\mathbf{F}_i = F_{i1}\mathbf{i} + F_{i2}\mathbf{j} + F_{i3}\mathbf{k}$. Here the potential energy of the system is a function $V(x_1,y_1,z_1, \dots, x_n,y_n,z_n)$ such that

$$\frac{\partial V}{\partial x_i} = -F_{i1}, \qquad \frac{\partial V}{\partial y_i} = -F_{i2}, \qquad \frac{\partial V}{\partial z_i} = -F_{i3},$$

the kinetic energy is

$$T = \frac{1}{2}\sum_{i=1}^{n} m_i\left[\left(\frac{dx_i}{dt}\right)^2 + \left(\frac{dy_i}{dt}\right)^2 + \left(\frac{dz_i}{dt}\right)^2\right],$$

and the action over a time interval $t_1 \le t \le t_2$ is

$$A = \int_{t_1}^{t_2} (T - V)\, dt.$$

In just the same way as above, we see that Newton's equations of motion for the system,

$$m_i\frac{d^2\mathbf{r}_i}{dt^2} = \mathbf{F}_i,$$

are a necessary condition for the action to have a stationary value. Hamilton's principle therefore holds for any finite system of particles in which the forces are conservative. It applies equally well to more general dynamical systems involving constraints and rigid bodies, and also to continuous media.

In addition, Hamilton's principle can be made to yield the basic laws of electricity and magnetism, quantum theory, and relativity. Its influence is so profound and far-reaching that many scientists regard it as the most powerful single principle in mathematical physics and place it at the pinnacle of physical science. Max Planck, the founder of quantum theory, expressed this view as follows: "The highest and most coveted

aim of physical science is to condense all natural phenomena which have been observed and are still to be observed into one simple principle . . . . Amid the more or less general laws which mark the achievements of physical science during the course of the last centuries, the principle of least action is perhaps that which, as regards form and content, may claim to come nearest to this ideal final aim of theoretical research."

Example 1. If a particle of mass m is constrained to move on a given surface $G(x,y,z) = 0$, and if no force acts on it, then it glides along a geodesic. To establish this, we begin by observing that since no force is present we have $V = 0$, so the Lagrangian $L = T - V$ reduces to T where

$$T = \frac{1}{2}m\left[\left(\frac{dx}{dt}\right)^2 + \left(\frac{dy}{dt}\right)^2 + \left(\frac{dz}{dt}\right)^2\right].$$

We now apply Hamilton's principle, and require that the action

$$\int_{t_1}^{t_2} L\, dt = \int_{t_1}^{t_2} T\, dt$$

be stationary subject to the side condition $G(x,y,z) = 0$. By Section 67, this is equivalent to requiring that the integral

$$\int_{t_1}^{t_2} [T + \lambda(t)G(x,y,z)]\, dt$$

be stationary with no side condition, where $\lambda(t)$ is an undetermined function of t. Euler's equations for this unconstrained variational problem are

$$m\frac{d^2x}{dt^2} - \lambda G_x = 0, \qquad m\frac{d^2y}{dt^2} - \lambda G_y = 0, \qquad m\frac{d^2z}{dt^2} - \lambda G_z = 0.$$

When m and λ are eliminated, these equations become

$$\frac{d^2x/dt^2}{G_x} = \frac{d^2y/dt^2}{G_y} = \frac{d^2z/dt^2}{G_z}.$$

Now the total energy $T + V = T$ of the particle is constant (we prove this below), so its speed is also constant, and therefore $s = kt$ for some constant k if the arc length s is measured from a suitable point. This enables us to write our equations in the form

$$\frac{d^2x/ds^2}{G_x} = \frac{d^2y/ds^2}{G_y} = \frac{d^2z/ds^2}{G_z}.$$

These are precisely equations 67-(29), so the path of the particle is a geodesic on the surface, as stated.

Lagrange's equations. In classical mechanics, Hamilton's principle can be viewed as the source of Lagrange's equations of motion, which occupy a dominant position in this subject. In order to trace the connection, we must first understand what is meant by degrees of freedom and generalized coordinates.

A single particle moving freely in three-dimensional space is said to have three *degrees of freedom,* since its position can be specified by three independent coordinates x, y, and z. By constraining it to move on a surface $G(x,y,z) = 0$, we reduce its degrees of freedom to two, since one of its coordinates can be expressed in terms of the other two. Similarly, an unconstrained system of n particles has $3n$ degrees of freedom, and the effect of introducing constraints is to reduce the number of independent coordinates needed to describe the configurations of the system. If the rectangular coordinates of the particles are x_i, y_i, and z_i $(i = 1, 2, \ldots, n)$, and if the constraints are described by k consistent and independent equations of the form

$$G_j(x_1,y_1,z_1, \ldots, x_n,y_n,z_n) = 0, \qquad j = 1, 2, \ldots, k,$$

then the number of degrees of freedom is $m = 3n - k$. In principle these equations can be used to reduce the number of coordinates from $3n$ to m by expressing the $3n$ numbers x_i, y_i, and z_i $(i = 1, 2, \ldots, n)$ in terms of m of these numbers. It is more convenient, however, to introduce Lagrange's *generalized coordinates* $q_1, q_2, \ldots, q_m$, which are any m independent coordinates whatever whose values determine the configurations of the system. This allows us full freedom to choose any coordinate system adapted to the problem at hand—rectangular, cylindrical, spherical, or any other—and renders our analysis independent of any particular coordinate system. We now express the rectangular coordinates of the particles in terms of these generalized coordinates and note that the resulting formulas automatically include the constraints: $x_i = x_i(q_1, \ldots, q_m)$, $y_i = y_i(q_1, \ldots, q_m)$, and $z_i = z_i(q_1, \ldots, q_m)$, where $i = 1, 2, \ldots, n$.

If m_i is the mass of the ith particle, then the kinetic energy of the system is

$$T = \frac{1}{2}\sum_{i=1}^{n} m_i\left[\left(\frac{dx_i}{dt}\right)^2 + \left(\frac{dy_i}{dt}\right)^2 + \left(\frac{dz_i}{dt}\right)^2\right];$$

and in terms of the generalized coordinates this can be written as

$$T = \frac{1}{2}\sum_{i=1}^{n} m_i\left[\left(\sum_{j=1}^{m}\frac{\partial x_i}{\partial q_j}\dot{q}_j\right)^2 + \left(\sum_{j=1}^{m}\frac{\partial y_i}{\partial q_j}\dot{q}_j\right)^2 + \left(\sum_{j=1}^{m}\frac{\partial z_i}{\partial q_j}\dot{q}_j\right)^2\right], \tag{1}$$

where $\dot{q}_j = dq_j/dt$. For later use, we point out that T is a homogeneous function of degree 2 in the $\dot{q}_j$. The potential energy V of the system is assumed to be a function of the q_j alone, so the Lagrangian $L = T - V$ is a function of the form

$$L = L(q_1,q_2, \ldots, q_m,\dot{q}_1,\dot{q}_2, \ldots, \dot{q}_m).$$

Hamilton's principle tells us that the motion proceeds in such a way that

the action $\int_{t_1}^{t_2} L\,dt$ is stationary over any interval of time $t_1 \le t \le t_2$, so Euler's equations must be satisfied. In this case these are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j} = 0, \qquad j = 1, 2, \ldots, m, \tag{2}$$

which are called *Lagrange's equations.* They constitute a system of m second order differential equations whose solution yields the q_j as functions of t.

We shall draw only one general deduction from Lagrange's equations, namely, the *law of conservation of energy.*

The first step in the reasoning is to note the following identity, which holds for any function L of the variables $t, q_1, q_2, \ldots, q_m, \dot{q}_1, \dot{q}_2, \ldots, \dot{q}_m$:

$$\frac{d}{dt}\left[\sum_{j=1}^{m}\dot{q}_j\frac{\partial L}{\partial \dot{q}_j} - L\right] = \sum_{j=1}^{m}\dot{q}_j\left[\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{q}_j}\right) - \frac{\partial L}{\partial q_j}\right] - \frac{\partial L}{\partial t}. \tag{3}$$

Since the Lagrangian L of our system satisfies equations (2) and does not explicitly depend on t, the right side of (3) vanishes and we have

$$\sum_{j=1}^{m}\dot{q}_j\frac{\partial L}{\partial \dot{q}_j} - L = E \tag{4}$$

for some constant E. We next observe that $\partial V/\partial \dot{q}_j = 0$, so $\partial L/\partial \dot{q}_j = \partial T/\partial \dot{q}_j$. As we have already remarked, formula (1) shows that T is a homogeneous function of degree 2 in the $\dot{q}_j$, so

$$\sum_{j=1}^{m}\dot{q}_j\frac{\partial L}{\partial \dot{q}_j} = \sum_{j=1}^{m}\dot{q}_j\frac{\partial T}{\partial \dot{q}_j} = 2T$$

by Euler's theorem on homogeneous functions.[13] With this result, equation (4) becomes $2T - L = E$ or $2T - (T - V) = E$, so

$$T + V = E,$$

which states that during the motion, the sum of the kinetic and potential energies is constant.

[13] Recall that a function $f(x,y)$ is homogeneous of degree n in x and y if $f(kx,ky) = k^n f(x,y)$. If both sides of this are differentiated with respect to k and then k is set equal to 1, we obtain

$$x\frac{\partial f}{\partial x} + y\frac{\partial f}{\partial y} = nf(x,y),$$

which is Euler's theorem for this function. The same result holds for a homogeneous function of more than two variables.

In the following example we illustrate the way in which Lagrange's equations can be used in specific dynamical problems.

Example 2. If a particle of mass m moves in a plane under the influence of a gravitational force of magnitude km/r^2 directed toward the origin, then it is natural to choose polar coordinates as the generalized coordinates: $q_1 = r$ and $q_2 = \theta$. It is easy to see that $T = (m/2)(\dot{r}^2 + r^2\dot{\theta}^2)$ and $V = -km/r$, so the Lagrangian is

$$L = T - V = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2) + \frac{km}{r}$$

and Lagrange's equations are

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{r}}\right) - \frac{\partial L}{\partial r} = 0, \tag{5}$$

$$\frac{d}{dt}\left(\frac{\partial L}{\partial \dot{\theta}}\right) - \frac{\partial L}{\partial \theta} = 0. \tag{6}$$

Since L does not depend explicitly on θ, equation (6) shows that $\partial L/\partial\dot{\theta} = mr^2\dot{\theta}$ is constant, so

$$r^2\frac{d\theta}{dt} = h \tag{7}$$

for some constant h assumed to be positive. We next observe that (5) can easily be written in the form

$$\frac{d^2r}{dt^2} - r\left(\frac{d\theta}{dt}\right)^2 = -\frac{k}{r^2}.$$

This is precisely equation 21-(12), which we solved in Section 21 to obtain the conclusion that the path of the particle is a conic section.

Variational problems for double integrals. Our general method of finding necessary conditions for an integral to be stationary can be applied equally well to multiple integrals. For example, consider a region R in the xy-plane bounded by a closed curve C (Fig. 97). Let $z = z(x,y)$ be a function that is defined in R and assumes prescribed boundary values on C, but is otherwise arbitrary (except for the usual differentiability conditions). This function can be thought of as defining a variable surface fixed along its boundary in space. An integral of the form

$$I(z) = \iint_R f(x,y,z,z_x,z_y)\,dx\,dy \tag{8}$$

will have values that depend on the choice of z, and we can pose the problem of finding a function z (a stationary function) that gives a stationary value to this integral.

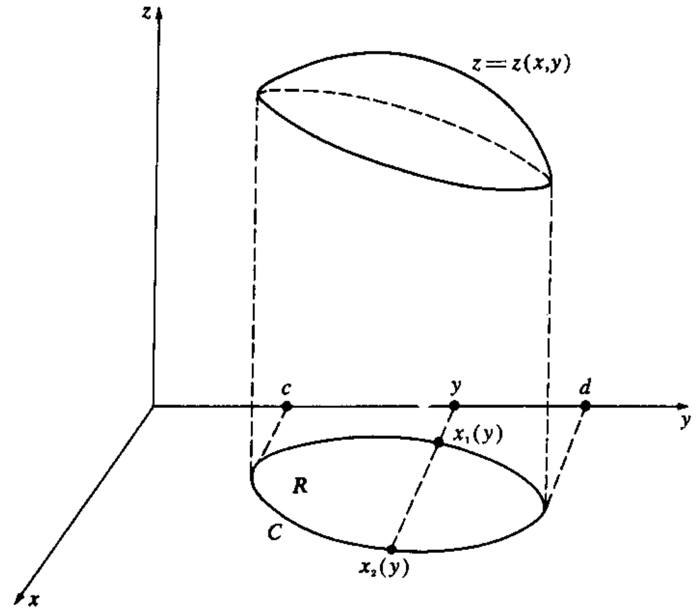

FIGURE 97

Our reasoning follows a familiar pattern. Assume that $z(x,y)$ is the desired stationary function and form the varied function $\bar{z}(x,y) = z(x,y) + \alpha\eta(x,y)$, where $\eta(x,y)$ vanishes on C. When $\bar{z}$ is substituted into the integral (8), we obtain a function $I(\alpha)$ of the parameter α, and just as before, the necessary condition $I'(0) = 0$ yields

$$\iint_R \left(\frac{\partial f}{\partial z}\eta + \frac{\partial f}{\partial z_x}\eta_x + \frac{\partial f}{\partial z_y}\eta_y\right) dx\,dy = 0. \tag{9}$$

To simplify the task of eliminating η_x and η_y, we now assume that the curve C has the property that each line in the xy-plane parallel to an axis intersects C in at most two points.[14] Then, regarding the double integral of the second term in parentheses in (9) as a repeated integral (see Fig. 97), we get

$$\iint_R \frac{\partial f}{\partial z_x}\eta_x\,dx\,dy = \int_c^d \int_{x_1(y)}^{x_2(y)} \frac{\partial f}{\partial z_x}\eta_x\,dx\,dy;$$

[14] This restriction is unnecessary, and can be avoided if we are willing to use Green's theorem.

and since

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial z_x} \eta_x \, dx = \eta \frac{\partial f}{\partial z_x}\Big]_{x_1}^{x_2} - \int_{x_1}^{x_2} \eta \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z_x}\right) dx$$
$$= -\int_{x_1}^{x_2} \eta \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z_x}\right) dx$$

because η vanishes on C, it follows that

$$\iint_R \frac{\partial f}{\partial z_x} \eta_x \, dx\, dy = -\iint_R \eta \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z_x}\right) dx\, dy.$$

The term containing η_y can be transformed by a similar procedure, and (9) becomes

$$\iint_R \eta\left[\frac{\partial f}{\partial z} - \frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z_x}\right) - \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z_y}\right)\right] dx\, dy = 0. \qquad (10)$$

We now conclude from the arbitrary nature of η that the bracketed expression in (10) must vanish, so

$$\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial z_x}\right) + \frac{\partial}{\partial y}\left(\frac{\partial f}{\partial z_y}\right) - \frac{\partial f}{\partial z} = 0 \qquad (11)$$

is Euler's equation for an extremal in this case. As before, a stationary function (if one exists) is an extremal that satisfies the given boundary conditions.

Example 3. In its simplest form, the *problem of minimal surfaces* was first proposed by Euler as follows: to find the surface of smallest area bounded by a given closed curve in space. If we assume that this curve projects down to a closed curve C surrounding a region R in the xy-plane, and also that the surface is expressible in the form $z = z(x,y)$, then the problem is to minimize the surface area integral

$$\iint_R \sqrt{1 + z_x^2 + z_y^2}\, dx\, dy$$

subject to the boundary condition that $z(x,y)$ must assume prescribed values on C. Euler's equation (11) for this integral is

$$\frac{\partial}{\partial x}\left(\frac{z_x}{\sqrt{1 + z_x^2 + z_y^2}}\right) + \frac{\partial}{\partial y}\left(\frac{z_y}{\sqrt{1 + z_x^2 + z_y^2}}\right) = 0,$$

which can be written in the form

$$z_{xx}(1 + z_y^2) - 2z_x z_y z_{xy} + z_{yy}(1 + z_x^2) = 0. \qquad (12)$$

This partial differential equation was discovered by Lagrange. Euler showed that every minimal surface not part of a plane must be saddle-shaped, and also that its mean curvature must be zero at every point.[15] The mathematical problem of proving that minimal surfaces exist, i.e., that (12) has a solution satisfying suitable boundary conditions, is extremely difficult. A complete solution was attained only in 1930 and 1931 by the independent work of T. Radó (Hungarian, 1895–1965) and J. Douglas (American, 1897–1965). An experimental method of finding minimal surfaces was devised by the blind Belgian physicist J. Plateau (1801–1883), who described it in his 1873 treatise on molecular forces in liquids. The essence of the matter is that if a piece of wire is bent into a closed curve and dipped in a soap solution, then the resulting soap film spanning the wire will assume the shape of a minimal surface in order to minimize the potential energy due to surface tension. Plateau performed many striking experiments of this kind, and since his time the problem of minimal surfaces has been known as *Plateau's problem.*[16]

Example 4. In Section 40 we obtained the one-dimensional wave equation from Newton's second law of motion. In this example we deduce it from Hamilton's principle with the aid of equation (11). Assume the following: a string of constant linear mass density m is stretched with a tension T and fastened to the x-axis at the points $x = 0$ and $x = \pi$; it is plucked and allowed to vibrate in the xy-plane; and its displacements $y(x,t)$ are relatively small, so that the tension remains essentially constant and powers of the slope higher than the second can be neglected. When the string is displaced, an element of length dx is stretched to a length ds, where

$$ds = \sqrt{1 + y_x^2}\, dx \cong \left(1 + \frac{1}{2}y_x^2\right) dx.$$

This approximation results from expanding $\sqrt{1 + y_x^2} = (1 + y_x^2)^{1/2}$ in the binomial series $1 + y_x^2/2 + \cdots$ and discarding all powers of y_x higher than the second. The work done on the element is $T(ds - dx) = \frac{1}{2}Ty_x^2\, dx$, so the potential energy of the whole string is

$$V = \frac{1}{2}T\int_0^\pi y_x^2\, dx.$$

The element has mass $m\, dx$ and velocity y_t, so its kinetic energy is $\frac{1}{2}my_t^2\, dx$,

[15] The *mean curvature* of a surface at a point is defined as follows. Consider the normal line to the surface at the point, and a plane containing this normal line. As this plane rotates about the line, the curvature of the curve in which it intersects the surface varies, and the mean curvature is one-half the sum of its maximum and minimum values.

[16] The standard mathematical work on this subject is R. Courant, *Dirichlet's Principle, Conformal Mapping, and Minimal Surfaces,* Interscience-Wiley, New York, 1950.

and for the whole string we have

$$T = \frac{1}{2} m \int_0^{\pi} y_t^2 \, dx.$$

The Lagrangian is therefore

$$L = T - V = \frac{1}{2} \int_0^{\pi} (my_t^2 - Ty_x^2) \, dx,$$

and the action, which must be stationary by Hamilton's principle, is

$$\frac{1}{2} \int_{t_1}^{t_2} \int_0^{\pi} (my_t^2 - Ty_x^2) \, dx \, dt.$$

In this case equation (11) becomes

$$\frac{T}{m} y_{xx} = y_{tt},$$

which we recognize as the wave equation 40-(8).

NOTE ON HAMILTON. The Irish mathematician and mathematical physicist William Rowan Hamilton (1805–1865) was a classic child prodigy. He was educated by an eccentric but learned clerical uncle. At the age of three he could read English; at four he began Greek, Latin, and Hebrew; at eight he added Italian and French; at ten he learned Sanskrit and Arabic; and at thirteen he is said to have mastered one language for each year he had lived. This forced flowering of linguistic futility was broken off at the age of fourteen, when he turned to mathematics, astronomy, and optics. At eighteen he published a paper correcting a mistake in Laplace's *Mécanique Céleste*; and while still an undergraduate at Trinity College in Dublin, he was appointed professor of astronomy at that institution and automatically became Astronomer Royal of Ireland.

His first important work was in geometrical optics. He became famous at twenty-seven as a result of his mathematical prediction of conical refraction. Even more significant was his demonstration that all optical problems can be solved by a single method that includes Fermat's principle of least time as a special case. He then extended this method to problems in mechanics, and by the age of thirty had arrived at a single principle (now called Hamilton's principle) that exhibits optics and mechanics as merely two aspects of the calculus of variations.

In 1835 he turned his attention to algebra, and constructed a rigorous theory of complex numbers based on the idea that a complex number is an ordered pair of real numbers. This work was done independently of Gauss, who had already published the same ideas in 1831, but with emphasis on the interpretation of complex numbers as points in the complex plane. Hamilton subsequently tried to extend the algebraic structure of the complex numbers, which can be thought of as vectors in a plane, to vectors in three-dimensional space. This project failed, but in 1843 his efforts led him to the discovery of quaternions. These are four-dimensional vectors that include the complex numbers as a subsystem; in modern terminology, they constitute the simplest

noncommutative linear algebra in which division is possible.[17] The remainder of Hamilton's life was devoted to the detailed elaboration of the theory and applications of quaternions, and to the production of massive indigestible treatises on the subject. This work had little effect on physics and geometry, and was supplanted by the more practical vector analysis of Willard Gibbs and the multilinear algebra of Grassmann and E. Cartan. The significant residue of Hamilton's labors on quaternions was the demonstrated existence of a consistent number system in which the commutative law of multiplication does not hold. This liberated algebra from some of the preconceptions that had paralyzed it, and encouraged other mathematicians of the late nineteenth and twentieth centuries to undertake broad investigations of linear algebras of all types.

Hamilton was also a bad poet and friend of Wordsworth and Coleridge, with whom he corresponded voluminously on science, literature, and philosophy.

[17] Fortunately Hamilton never learned that Gauss had discovered quaternions in 1819 but kept his ideas to himself. See Gauss, *Werke*, vol. VIII, pp. 357–362.