VARIATIONAL METHODS

G. W. F. DRAKE, *Department of Physics, University of Windsor, Windsor, Ontario, Canada*

INTRODUCTION

Variational principles derive from a certain aesthetic and metaphysical ideal of simplicity in the search for the principles underlying physical phenomena. The origins date back to the earliest Greek philosophers Thales (c. 600 B.C.) and Pythagoras (c. 550 B.C.). Aristotle (384–322 B.C.) clearly makes use of a variational principle to justify circular orbits for the planets when he says in *de Caelo II*

> Now of lines which return upon themselves, the line which bounds the circle is the shortest, and that movement is the swiftest which follows the shortest line.

This marks the first use of a "minimum" postulate, and the conclusion held sway until the time of Kepler (1571–1630). Hero of Alexandria (c. 125 B.C.) made the first rigorous use of a variational principle when he proved that when the angle of incidence equals the angle of reflection, the path taken by a ray of light from the object to the observer is shorter than any other possible path with fixed end points (see *Catoptrics* by Hero in

3-527-29476-7/98/$5.00 + .50

Cohen and Drabkin, 1958). This later became Fermat's principle of least time in geometrical optics.

The belief that nature is in some sense "simple" and can be explained by some economically small number of postulates pervades the works of Galileo (1564–1642), Newton (1642–1727), and Leibniz (1646–1716). Although some of the early conclusions turned out to be scientifically unfounded, the philosophical basis for variational principles has great antiquity. They continue to guide the development of new physical theories at the most fundamental level and to provide powerful methods of practical computation. Perhaps most importantly, they bring out the structural analogies between superficially different phenomena and allow techniques developed in one field to be readily applied in another. At a fundamental level, practically all physical phenomena can be expressed in terms of variational principles that have a striking similarity.

The *calculus of variations* provides the basic mathematical tool for formulating and analyzing variational principles. The purposes of this article are first to give an overview of the calculus of variations, and then to discuss its application to a variety of physical phenomena. There is a vast literature on both aspects, and only a few of the most important points can be covered in the space available. Only a few principal references are given, with further general references in the reading list at the end. The main emphasis is on applications to classical mechanics and bound-state problems in quantum mechanics. Except for a brief discussion of the Feynman path integral, applications to scattering problems are not covered. An informal and instructive introduction to variational methods can be found in Hildebrand and Tromba (1985).

1. TECHNIQUES OF THE CALCULUS OF VARIATIONS

In its simplest form, the calculus of variations addresses the problem of finding the function $y(x)$ for which the integral

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

is an extremum. The integrand $f(x,y,y_x)$ is some prescribed function of the indicated variables, where $y_x = dy/dx$, and x_1, x_2 are fixed end points. J is termed a *functional* of $y(x)$. As originally formulated by Euler, the problem is solved by considering infinitesimal variations $\delta y(x)$ about a particular path $y(x)$ connecting x_1 and x_2 (see Fig. 1) and demanding that the variation δJ induced in J vanish. For example, one might choose the variations to be $\delta y(x) = \epsilon\eta(x)$, where $\eta(x)$ is an arbitrary function such that $\eta(x_1) = \eta(x_2) = 0$ (to make the variation vanish at the end points), and ϵ is a small parameter controlling the size of the variation. Then $\delta y_x(x) = \epsilon d\eta(x)/dx$, and from a Taylor series expansion of $f(x,y,y_x)$ about $\epsilon = 0$ in Eq. (1), the induced variation in J is

$$\delta J = \epsilon \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y}\eta(x) + \frac{\partial f}{\partial y_x}\frac{d\eta(x)}{dx}\right]dx \tag{2}$$

up to terms of first order in ϵ. An integration of the second term by parts yields

$$\delta J = \epsilon \frac{\partial f}{\partial y_x}\eta(x)\Big|_{x_1}^{x_2} + \epsilon \int_{x_1}^{x_2} \left[\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x}\right]\eta(x)dx. \tag{3}$$

The first term vanishes by the assumption that $\eta(x) = 0$ at the end points. Since $\eta(x)$ is otherwise an arbitrary function, the condition $\delta J = 0$ can be fulfilled only if the integrand of the second term vanishes identically for $x_1 < x < x_2$; i.e.,

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x} = 0. \tag{4}$$

This is the basic Euler–Lagrange equation. For purposes of compactness, the functional

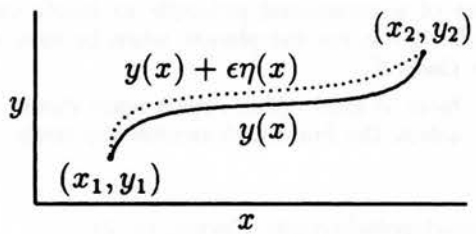

FIG. 1. Illustration of the actual path $y(x)$ and the varied path connecting fixed end points.

dependence of f on its various arguments is usually suppressed, as is done here.

Any solution to Eq. (4) satisfies the variational condition $\delta J = 0$. However, J itself could be a minimum, a maximum, or a point of inflection. One often knows from the geometrical nature of the problem being solved which case applies. Otherwise, it is necessary to extend the Taylor series expansion of $f(x,y,y_x)$ to terms of order ϵ^2 and determine the sign of the second-order variation $\delta^{(2)}J$ (see, e.g., Courant and Hilbert, 1966). If $\delta^{(2)}J < 0$, then J is a maximum; if $\delta^{(2)}J > 0$, then J is a minimum; if $\delta^{(2)}J = 0$, then J lies at a point of inflection.

If the end points are not fixed, then $\eta(x)$ does not vanish there. In this case, Eq. (4) still applies, subject to the condition that $\partial f/\partial y_x = 0$ at the end points (see Jeffreys and Jeffreys, 1972).

An important special case of Eq. (4) occurs if $f(x,y,y_x)$ does not depend explicitly on x because then the integrating factor is simply y_x. After multiplying Eq. (4) through by y_x and using

$$\frac{df}{dx} = \frac{\partial f}{\partial y} y_x + \frac{\partial f}{\partial y_x}\frac{dy_x}{dx} \tag{5}$$

(since $\partial f/\partial x = 0$ by assumption), the Euler–Lagrange equation becomes

$$\frac{d}{dx}\left(y_x \frac{\partial f}{\partial y_x} - f\right) = 0, \tag{6}$$

and so

$$y_x \frac{\partial f}{\partial y_x} - f = \text{const.} \tag{7}$$

A classic example is provided by the brachistochrone (shortest time) problem first propounded by John Bernoulli in 1696. It was solved by both him and his brother James, as well as by Newton and Leibnitz. Consider a bead sliding without friction on a wire of arbitrary length connecting two fixed points (x_1,y_1) and (x_2,y_2) in a vertical plane, as shown in Fig. 2. The problem is to find the shape that minimizes the travel time as the bead slides from rest under the force of gravity; i.e., to find the function $y = y(x)$ such that the integral

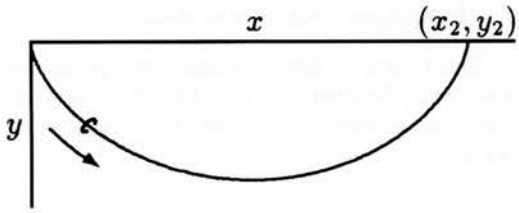

FIG. 2. The brachistochrone problem of a bead sliding without friction on a wire of arbitrary length in a vertical plane.

$$\tau = \int_{x_1}^{x_2} \frac{ds}{v} \tag{8}$$

for the travel time is a minimum. Here $ds = [(dy/dx)^2 + 1]^{1/2}dx$ is the element of arc length and v is the velocity. By conservation of energy, the velocity after falling a distance y (measured downward) is $v = \sqrt{2gy}$ (independent of x), where g is the acceleration due to gravity. The integral to be minimized is then

$$\tau = \int_{x_1}^{x_2} \frac{(y_x^2 + 1)^{1/2}}{(2gy)^{1/2}}\,dx. \tag{9}$$

With $f(x,y,y_x)$ defined by the above integrand, the Euler–Lagrange equation (7) then gives

$$\frac{y_x^2}{(y_x^2 + 1)^{1/2}y^{1/2}} - \frac{(y_x^2 + 1)^{1/2}}{y^{1/2}} = c^{1/2}, \tag{10}$$

and hence

$$y_x^{-1} \equiv \frac{dx}{dy} = \left(\frac{cy}{1 - cy}\right)^{1/2}, \tag{11}$$

where c is a constant of integration that determines the distance scale. Assuming that the bead starts from the origin, integration of this equation yields

$$cx = \sin^{-1}(cy)^{1/2} - (cy - c^2y^2)^{1/2}. \tag{12}$$

This is the equation of a cycloid symmetric about the minimum at $cy = 1$. The scale factor c is determined by the condition that the curve pass through the second terminus (x_2,y_2). If $y_2 = 0$ (as in Fig. 2), then $c = \pi/x_2$.

1.1 Variations with Constraints

There are many classes of problems where the functional J must be made an extremum subject to a subsidiary condition of the form

$$\int_{x_1}^{x_2} f_1(x,y,y_x)dx = \text{const.} \tag{13}$$

The example of a hanging chain of fixed length is discussed in the following paragraph. Such problems can be handled by applying the variational procedure to the functional

$$g(x,y,y_x) = f(x,y,y_x) + \lambda_1 f_1(x,y,x_y), \tag{14}$$

where λ_1 is called a *Lagrange undetermined multiplier.* The resulting Euler–Lagrange equation is

$$\frac{\partial f}{\partial y} - \frac{d}{dx}\frac{\partial f}{\partial y_x} = -\lambda_1\left(\frac{\partial f_1}{\partial y} - \frac{d}{dx}\frac{\partial f_1}{\partial y_x}\right), \tag{15}$$

or, if f and f_1 are independent of x, the first integral is [cf. Eq. (7)]

$$y_x\frac{\partial f}{\partial y_x} - f + \lambda_1\left(y_x\frac{\partial f_1}{\partial y_x} - f_1\right) = \text{const.} \tag{16}$$

The idea is to solve Eq. (15) or Eq. (16) for a fixed but arbitrary value of λ_1. The equation of constraint provides the additional condition to determine λ_1 at the end of the problem, together with the two constants of integration.

As an example, consider the problem of finding the shape of a uniform hanging chain with both ends fixed. The shape is such that the potential energy due to gravity is a minimum, and so the quantity to be minimized is

$$J = \int_{x_1}^{x_2} \mu g y(y_x^2 + 1)^{1/2}dx, \tag{17}$$

subject to the constraint

$$\int_0^L ds \equiv \int_{x_1}^{x_2} (y_x^2 + 1)^{1/2}dx = L, \tag{18}$$

where μ is the mass per unit length and L the length. Thus $f = \mu g y(y_x^2 + 1)^{1/2}$ and $f_1 = (y_x^2 + 1)^{1/2}$. After dividing by μg and defining $\lambda = \lambda_1/\mu g$, Eq. (16) gives

$$(y + \lambda)\left[\frac{y_x^2}{(y_x^2 + 1)^{1/2}} - (y_x^2 + 1)^{1/2}\right] = -\frac{1}{c}. \tag{19}$$

As for the brachistochrone problem, this equation can be solved for $y_x^{-1} \equiv dx/dy$ and the result integrated to obtain

$$cy = -c\lambda + \cosh[c(x - a)], \tag{20}$$

where a is the second constant of integration. The three parameters a, c, and λ are determined by the three conditions that the curve pass through the points (x_1,y_1) and (x_2,y_2) at the ends of the chain, together with the equation of constraint. For example, if the two points are (0,0) and $(x_2,0)$, then $a = x_2/2$, $c\lambda = \cosh(cx_2/2)$, and the equation of constraint becomes

$$L = \int_0^{x_2} [\sinh^2[c(x - \tfrac{1}{2}x_2)] + 1]^{1/2}dx = \frac{2}{c}\sinh(cx_2/2). \tag{21}$$

Solving this transcendental equation determines the remaining constant c. The quantity $T_0 = \mu g/c$ is the force of tension in the chain at the lowest point, where $y_x = 0$.

1.2 Generalizations

In applications to mechanics (see Sec. 2.1), the time t plays the role of the independent variable x marking the evolution of the system, but there are typically several dependent generalized coordinates q_i and $\dot{q}_i \equiv dq_i/dt$, $i = 1, \ldots, N$, in place of y and y_x, respectively. With this change in notation, f becomes a function of all the q_i's, $\dot{q}_i$'s, and t, and the generalization of Eq. (4) is

$$\frac{\partial f}{\partial q_i} - \frac{d}{dt}\frac{\partial f}{\partial \dot{q}_i} = 0, \quad i = 1, \ldots, N. \tag{22}$$

There are thus N coupled Euler–Lagrange equations, one for each degree of freedom of the system.

If in addition there are several independent variables $t_1, t_2, \ldots, t_r$, then Eq. (22) is further generalized to read

$$\frac{\partial f}{\partial q_i} - \sum_{j=1}^{r} \frac{\partial}{\partial t_j} \frac{\partial f}{\partial(\partial q_i/\partial t_j)} = 0, \quad i = 1, \ldots, N. \tag{23}$$

There can also be several equations of constraint of the form of Eq. (13) with integrands $f_k(t_j,q_i,\dot{q}_i)$, $k = 1, \ldots, m$. In this case, m Lagrange undetermined multipliers λ_k are introduced, and the function f in Eq. (22) or (23) is replaced by

$$g(t_j,q_i,\dot{q}_i) = f(t_j,q_i,\dot{q}_i) + \sum_{k=0}^{m} \lambda_k f_k(t_j,q_i,\dot{q}_i). \tag{24}$$

Constraints that can be expressed in integrated form, such as Eq. (13), are said to be *holonomic* (wholly named or specified). However, problems often arise in mechanics involving *nonholonomic* constraints that can only be expressed in differential form, such as a relation between velocities. An example is the problem of a vertical disk of radius R rolling without slipping on a plane. Four coordinates are required—the (x,y) Cartesian coordinates of the point of contact between the disk and the plane, a spinning angle of rotation θ about a vertical axis, and a rolling angle ϕ about an axis perpendicular to the disk. If the plane of the disk is initially perpendicular to the x axis (i.e., $\theta = 0$), then the constraint of "not slipping" corresponds to the differential relations

$$dx = R \sin\theta \, d\phi, \tag{25}$$

$$dy = -R \cos\theta \, d\phi. \tag{26}$$

These equations cannot be integrated without knowing in advance θ and ϕ as a function of t. However, the method of Lagrange undetermined multipliers can still be applied. If the general form of the differential constraints is written as

$$\sum_{i=1}^{N} a_{k,i} dq_i + a_{k,t} dt = 0, \quad k = 1, \ldots, s, \tag{27}$$

where the coefficients $a_{k,i}$ are, in general, functions of the q_i's and $\dot{q}_i$'s, then the generalization of Eq. (22) for nonholonomic systems is

$$\frac{\partial f}{\partial q_i} - \frac{d}{dt}\frac{\partial f}{\partial \dot{q}_i} + \sum_{k=1}^{s} \lambda_k a_{k,i} = 0, \quad i = 1, \ldots, N. \tag{28}$$

This equation can still be used even if the constraints are in fact holonomic. The term $\sum_{k=1}^{s} \lambda_k a_{k,i}$ can be identified with the generalized forces of constraint. In general, the λ_k are now functions of the q_i's and $\dot{q}_i$'s. The term $a_{k,t}dt$ in Eq. (27) does not contribute to Eq. (28) because the variations δq_i from the actual path are considered to occur at a particular instant of time (see Sec. 2.1).

As a simple illustration of the versatility of the method, consider the problem of finding the function $u(x,y,z)$ such that the square of its gradient is a minimum in a given volume of space. The problem is then to minimize

$$J = \iiint f dx dy dz, \tag{29}$$

with

$$f = (\nabla u)^2 = \left(\frac{\partial u}{\partial x}\right)^2 + \left(\frac{\partial u}{\partial y}\right)^2 + \left(\frac{\partial u}{\partial z}\right)^2. \tag{30}$$

This can be regarded as an application of Eq. (23) with the three independent variables $t_1 = x$, $t_2 = y$, and $t_3 = z$, and a single degree of freedom ($N = 1$) with $q_1 = u$. Equation (23) then immediately gives

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = 0, \tag{31}$$

which is Laplace's equation. The term $\partial f/\partial q_1 \equiv \partial f/\partial u$ does not contribute because f does not depend explicitly on u, only on the partial derivatives $\partial u/\partial x$, $\partial u/\partial y$, and $\partial u/\partial z$. This problem clearly illustrates the way in which functional derivatives are to be interpreted and should be carefully studied.

2. APPLICATIONS TO CLASSICAL MECHANICS

The elements and techniques of classical mechanics are covered under MECHANICS, CLASSICAL. This section draws together the connections with variational principles.

2.1 Introductory Concepts

Consider a classical system of n interacting particles having masses m_s located at positions $\mathbf{r}_s = (x_{s,1}, x_{s,2}, x_{s,3})$, and acted on by

forces $\mathbf{F}_s = (F_{s,1},F_{s,2},F_{s,3})$ due to the other particles and any external forces, including forces of constraint. The evolution of the system is obtained by solving Newton's equations of motion

$$m_s\ddot{x}_{s,j} = F_{s,j}, \quad s = 1, \ldots, n; j = 1, 2, 3. \tag{32}$$

These equations are completely general. However, it is often more convenient to replace the $3n$ Cartesian coordinates $x_{s,j}$ by *generalized coordinates* q_i defined through a system of connection equations of the form

$$x_{s,j} = x_{s,j}(q_1,q_2, \ldots ,q_{3n}). \tag{33}$$

The use of generalized coordinates is particularly effective in problems involving constraints. If the generalized coordinates are chosen such that their variations δq_i do no *virtual work* against the forces of constraint (i.e., δq_i is perpendicular to the *instantaneous* forces of constraint), then the number of independent q_i needed is reduced from $3n$ to $3n - m$, where m is the number of constraints.

To make these ideas concrete, consider the example of a bead sliding on a vertical wire hoop of radius R that is itself constrained to rotate about the z axis with angular velocity ω. In terms of the polar angles θ and $\phi = \omega t$, the connection equations are

$$\begin{aligned} x &= \sin\theta \cos\omega t, \\ y &= \sin\theta \sin\omega t, \\ z &= \cos\theta. \end{aligned} \tag{34}$$

The constraint of sliding on the hoop has reduced the number of independent coordinates from three to a single azimuthal angle θ. A variation $\delta\theta$ generates a displacement of the bead consistent with the instantaneous orientation of the hoop, but not consistent with the actual time evolution of the system, which includes the rotation of the hoop. Such variations are said to do no virtual work, and are called *virtual displacements*. From the connection equations (34), the kinetic energy of the bead is

$$\begin{aligned} T &= \tfrac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) \\ &= \tfrac{1}{2}m[(R\dot{\theta})^2 + (\omega R \sin\theta)^2]. \end{aligned} \tag{35}$$

If the system is conservative with a potential-energy function $V(x,y,z)$, then V can similarly be expressed in terms of θ.

2.2 Hamilton's Principle

In the absence of constraints, a direct transformation of Newton's equations of motion (32) from the $3n$ Cartesian coordinates $x_{s,j}$ to the $3n$ generalized coordinates q_i yields Lagrange's equations of motion

$$\frac{d}{dt}\frac{\partial T}{\partial \dot{q}_i} - \frac{\partial T}{\partial q_i} = Q_i, \tag{36}$$

where the Q_i are the generalized forces defined by

$$Q_i = \sum_{s=1}^{n} \mathbf{F}_s \cdot \frac{\partial \mathbf{r}_s}{\partial q_i}. \tag{37}$$

A comparison with Eqs. (3) and (22) shows immediately that with the identification $f = -T$, Lagrange's equations correspond to the variational condition

$$\int_{t_1}^{t_2} \left(\delta T + \sum_{i=1}^{3n} Q_i \delta q_i\right) dt = 0. \tag{38}$$

This is the most general form of Hamilton's principle in classical dynamics. The advantage gained is that $3n$ equations of motion have been consolidated into a single scalar variational condition that is invariant under coordinate transformation. In the absence of constraints, all the q_i can be varied independently so that each coefficient of δq_i must vanish separately, and Lagrange's equations are recovered. If m constraints are present, one need keep only $3n - m$ of the q_i whose virtual displacements δq_i are consistent with the instantaneous constraints, as in the rotating-hoop example of Sec. 2.1 where only a single parameter θ is required. One can still consider variations in the remaining q_i, even though they would violate the constraints. The only difference is that the corresponding Q_i are reinterpreted as the generalized forces required to maintain the constraints. They can be calculated by the method of Lagrange undetermined multipliers, as described in Sec. 1.2 [see especially Eq. (28)]. In this way, Lagrange's equations apply to the entire set, whether or not constraints are present. The invariance of La-

grange's equations and Hamilton's principle under coordinate transformation guarantees that if they are correct in Cartesian coordinates (as can easily be checked), they are correct in any other system of generalized coordinates.

2.2.1 Conservative Systems and First Integrals If the system is conservative, then the Q_i are derivable from a potential function $V(q_i)$ according to $Q_i = -\partial V/\partial q_i$, and Lagrange's equations reduce to

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \tag{39}$$

where $L = T - V$ is the Lagrangian. It is then immediately evident from Eq. (4) that Hamilton's principle becomes

$$\delta \int_{t_1}^{t_2} L\, dt = 0. \tag{40}$$

This is the most useful form of Hamilton's principle for theoretical discussion.

If T and V do not involve time explicitly, then the first integral corresponding to Eq. (6) is

$$\sum_i \dot{q}_i \frac{\partial T}{\partial \dot{q}_i} - T + V = \text{const.} \tag{41}$$

If, further, T is a homogeneous quadratic function of the $\dot{q}_i$, then the above becomes

$$T + V = \text{const.}, \tag{42}$$

where the constant can now be identified with the conserved energy E of the system.

For the example of a bead on a rotating hoop discussed in Sec. 2.1, the first integral gives

$$\begin{aligned} &\tfrac{1}{2}m[(R\dot{\theta})^2 - (\omega R \sin\theta)^2] + V(\theta) \\ &\quad = T + V - m(\omega R \sin\theta)^2 = \text{const.} \end{aligned} \tag{43}$$

The constant of the motion here differs from $E = T + V$ because the time-dependent forces of constraint do work on the system. Overall conservation of energy is recovered only when the work required to keep the hoop rotating at a constant rate is included. If N_z is the required torque, then, with the use of Eq. (43), the rate at which it does work is

$$N_z\omega = \frac{d}{dt}(T + V) = \frac{d}{dt} m(\omega R \sin\theta)^2 = \omega \frac{dl_z}{dt}, \tag{44}$$

where $l_z = m\omega(R \sin\theta)^2$ is the angular momentum of the bead about the z axis. Thus $N_z = dl_z/dt$ as expected.

2.3 The Hamilton–Jacobi Equation

The variations considered thus far are taken between fixed end points t_1 and t_2 and, therefore, necessarily do not correspond to a possible dynamical evolution of the system. The actual evolution between fixed end points is uniquely defined, at least with respect to local variations, and so the varied path is unphysical. (But it is not necessarily so for nonlocal variations. For example, two points on a Kepler orbit are connected by two paths, depending on which way around the particle goes.)

The Hamilton–Jacobi equation comes from consideration of a different kind of variation Δq_i along a possible dynamical path between points allowed to vary in both space and time. In this case, it is necessary to keep the integrated term in Eq. (3) (or its generalizations). Suppose that t_1 and t_2 are replaced by $t_1 + \delta t_1$ and $t_2 + \delta t_2$. There is then a corresponding variation δt in the arrival time at each point along the path, so that $\Delta q_i(t) \equiv \delta q_i(t + \delta t)$ is the variation in path evaluated at the modified arrival time.

For definiteness, suppose that the paths are parameterized according to

$$q_i(\epsilon,t) = q_i(0,t) + \epsilon\eta_i(t), \tag{45}$$

where $q_i(0,t)$ is the actual path and $\eta_i(t)$ is an arbitrary differentiable function not assumed to vanish at the end points. Then $\delta q_i(t) = \epsilon\eta_i(t)$ and

$$\begin{aligned} \delta q_i(t + \delta t) &= q_i(\epsilon, t + \delta t) - q_i(0,t) \\ &\simeq q_i(0, t + \delta t) - q_i(0,t) + \epsilon\eta_i(t) \\ &= \dot{q}_i \delta t + \delta q_i(t) \end{aligned} \tag{46}$$

up to terms of first order in ϵ and δt. Application of the Δ variation to the integral in Hamilton's principle then yields

$$\begin{aligned}\Delta \int_{t_1}^{t_2} L\,dt &= \int_{t_1+\delta t_1}^{t_2+\delta t_2} L\,dt - \int_{t_1}^{t_2} L\,dt + \int_{t_1}^{t_2} \delta L\,dt \\ &= L\delta t\Big|_{t_1+\delta t_1}^{t_2+\delta t_2} + \sum_i \frac{\partial L}{\partial \dot{q}_i}\,\delta q_i\Big|_{t_1}^{t_2} \\ &\quad + \int_{t_1}^{t_2} \sum_i \left(\frac{\partial L}{\partial q_i} - \frac{d}{dt}\frac{\partial L}{\partial \dot{q}_i}\right)\delta q_i dt. \end{aligned} \tag{47}$$

The last term vanishes by the assumption that the varied path is a possible dynamical path, and so Lagrange's equations of motion are satisfied. Interest therefore centers on the integrated terms. The first term is simply $L(\delta t_2 - \delta t_1)$. Using Eq. (46) to replace $\delta q_i(t)$ by $\delta q_i(t + \delta t)$ in the second term, these terms become

$$\Delta \int_{t_1}^{t_2} L\,dt = \left[\left(L - \sum_i \dot{q}_i \frac{\partial L}{\partial \dot{q}_i}\right)\delta t + \sum_i \frac{\partial L}{\partial \dot{q}_i}\,\delta q_i\right]_{t_1+\delta t_1}^{t_2+\delta t_2}. \tag{48}$$

With the definitions

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \tag{49}$$

$$H = \sum_i \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} - L, \tag{50}$$

$$S = \int_{t_1}^{t_2} L\,dt, \tag{51}$$

where the integral for S is taken along any dynamical path, Eq. (48) can be written in the form

$$\Delta S = \left[-H\delta t + \sum_i p_i \delta q_i\right]_{t_1+\delta t_1}^{t_2+\delta t_2}. \tag{52}$$

H is called the *Hamiltonian* for the system, p_i is the *canonical momentum,* and S is called *Hamilton's principal function.* For an ordinary conservative system, $H = T + V$ is the total energy.

Equation (52) may be interpreted as follows. Since by assumption the system evolves along a possible dynamical path from an initial configuration $(q_i)_1$ at time t_1 to a final configuration $(q_i)_2$ at time t_2, there will in general be only a single set of initial velocities $(\dot{q}_i)_1$ that satisfy these requirements (at least with respect to small variations). The same applies to the initial momenta $(p_i)_1$ since they are connected to the $(\dot{q}_i)_1$ through Eq. (49), and hence the $(\dot{q}_i)_1$ can be eliminated from the problem. The p_i's as well as the q_i's are therefore uniquely determined from the initial conditions at time t_1. Since Eq. (52) remains true as t_2 is varied, it follows that

$$\frac{\partial S}{\partial q_i} = p_i, \tag{53}$$

$$\frac{\partial S}{\partial t} = -H(q_i,p_i,t) = -H\left(q_i,\frac{\partial S}{\partial q_i},t\right). \tag{54}$$

Equation (54) is called the *Hamilton–Jacobi* equation. It is a first-order partial differential equation in $N + 1$ variables and does not involve S explicitly. As it stands, a complete solution involving $N + 1$ constants of integration is determined from the initial conditions

$$\left(\frac{\partial S}{\partial t}\right)_{t_1} = (H)_{t_1}, \quad \left(\frac{\partial S}{\partial q_i}\right)_{t_1} = -(p_i)_{t_1}. \tag{55}$$

The real significance of the Hamilton–Jacobi equation comes not from its practical utility in solving mechanical problems, but for the insight it gives into the structure of mechanics, and for applications involving the use of perturbation theory. The significance is further elaborated after a brief discussion of the principle of least action, Hamilton's equations of motion, and the theory of canonical transformations.

2.3.1 The Principle of Least Action Assume that the system is conservative and holonomic so that H is a constant. The Maupertuis principle of least action follows by considering a restricted class of variations Δ_t that are identical to the Δ variations of Sec. 2.3, except that $\Delta_t q_i = \delta q_i(t + \delta t)$ is assumed to vanish at the end points; i.e., the system arrives at the same end point, but at the varied time $t + \delta t$, and H has the same constant value on the varied path. The varied path could be the same as the actual path except that it is traversed at a different rate. Under these conditions, Eq. (52) reduces immediately to

$$\Delta_t S = -H(\delta t_2 - \delta t)_1. \tag{56}$$

However, with the use of Eq. (50) to replace L by H, a direct evaluation of $\Delta_t S$ from Eq. (51) yields

$$\Delta_t S = \Delta_t \int_{t_2}^{t_2} \sum_i p_i \dot{q}_i dt - H(\delta t_2 - \delta t_1), \tag{57}$$

from which it follows that

$$\Delta_t \int_{t_1}^{t_2} \sum_i p_i \dot{q}_i dt = 0. \tag{58}$$

The quantity

$$W = \int_{t_1}^{t_2} \sum_i p_i \dot{q}_i dt \tag{59}$$

defines the classical action.

This is one of many possible ways of expressing the principle of least action. A purely geometrical form in which t is eliminated can be obtained as follows. If V is velocity-independent and T is a homogeneous quadratic function of the $\dot{q}_i$, then

$$\sum_i p_i \dot{q}_i = 2T = \sum_{j,k} M_{j,k}(q) \dot{q}_j \dot{q}_k, \tag{60}$$

where the $M_{j,k}(q)$ are the coefficients appearing in the kinetic-energy expression. The $M_{j,k}(q)$ can be regarded as the elements of a metric tensor in a curvilinear coordinate space such that the element of path length is

$$(d\rho)^2 = \sum_{j,k} M_{j,k}(q) dq_j dq_k. \tag{61}$$

The principle of least action can then be written in the form

$$\Delta_t \int_{t_1}^{t_2} T dt = 0 = \Delta \int_{\rho_1}^{\rho_2} \sqrt{T}\, d\rho, \tag{62}$$

or equivalently,

$$\Delta \int_{\rho_1}^{\rho_2} \sqrt{H - V}\, d\rho = 0. \tag{63}$$

For a single particle moving in a potential V, $d\rho$ is simply the element of arc length ds along the trajectory. In this form, the principle of least action is formally identical to Fermat's principle of geometrical optics. One need merely identify $\sqrt{H - V}$ with a variable index of refraction $n(s) = c_{\text{vac}}/c(s)$, which is inversely proportional to the velocity of light in the medium. The path taken by a beam of light is then such that the travel time given by

$$\tau = \frac{1}{c_{\text{vac}}} \int_{s_1}^{s_2} n(s) ds \tag{64}$$

is a minimum (or, more strictly, an extremum). This justifies the solution to the brachistochrone problem in Sec. 1, and it demonstrates the formal equivalence between geometrical optics and the dynamics of conservative systems.

In a recent article, Gray *et al.* (1996) show that the Maupertuis principle of least action can be generalized to a form in which W is stationary with respect to varied paths on which H is held constant only *in the mean* and that there exists a reciprocal principle in which the mean value $\bar{H}$ is made stationary with respect to varied paths of constant W. They also show that the reciprocal principle can be derived directly from the classical limit of the Schrödinger variational principle of quantum mechanics (see Sec. 3.2). A similar reciprocity theorem for Hamilton's principle provides a set of four variational principles analogous to the four equilibrium principles of thermodynamics.

2.3.2 Hamilton's Equations of Motion

The basic approach in Lagrangian mechanics is to regard the N generalized coordinates $q_i(t)$ as the independent variables whose time dependence is determined by the N second-order Lagrangian equations of motion expressed by Eq. (39). The velocities $\dot{q}_i(t)$ enter only as derived quantities whose initial values, together with the initial $q_i(t)$, determine the required $2N$ constants of integration. The Hamiltonian approach differs in that the $\dot{q}_i(t)$ are eliminated in favor of the canonical momenta $p_i(t)$ defined by Eq. (49). The $p_i(t)$ are then elevated to an equal footing with the $q_i(t)$ so that the set $\{q_i, p_i | i = 1, \ldots, N\}$ forms a set of $2N$ independent variables satisfying a set of $2N$ coupled first-order differential equations called Hamilton's equations of motion. In what follows, we adopt the convention that a summation over repeated subscripts is implied, and q or p without a subscript stands for the entire set.

Hamilton's equations of motion can be