Physics 221A
Fall 2007
Appendix B
Classical Mechanics

1. Introduction

In this Appendix we summarize some aspects of classical mechanics that will be relevant for our course in quantum mechanics. We concentrate on the Lagrangian and Hamiltonian formulation of classical mechanics.

2. Lagrangians, Dissipation, and Entropy

In this Appendix we consider exclusively classical systems that can be described by a Lagrangian. This means that we neglect friction and other dissipative processes that cause an increase in entropy. Dissipation is always present in macroscopic systems, so its neglect is an approximation that is more or less good depending on circumstances.

The concept of entropy applies when making a statistical description of a system, which is often necessary when the number of degrees of freedom is large. Entropy provides a measure of our ignorance of the state of the system, that is, of the values of the dynamical variables that describe the system.

In some macroscopic systems it is a good approximation to concentrate on just a few degrees of freedom that we are interested in, and to neglect interactions with a large number of other degrees of freedom that we do not care about. For example, in the motion of a planet around the sun we may wish to describe the position of the center of mass of the planet in detail while ignoring its rotations as well as the internal stresses inside the solid body of the planet, the bulk motions of fluids in its oceans and atmospheres, etc. The latter are macroscopic degrees of freedom that are in turn coupled to microscopic ones, the motions of the individual atoms and molecules that make up the planet. There is a chain of couplings leading from macroscopic degrees of freedom down to microscopic ones, and a corresponding transport of energy from large motions down to small ones, ultimately resulting in the production of heat.

If the couplings at any stage are small it may be possible to neglect them, to obtain a closed system with few degrees of freedom. This would certainly simplify the problem and may be valid over short periods of time. But over longer periods dissipation will be important and cannot be ignored. For example, ancient dissipative processes inside the body of the moon have caused it to settle into a state in which its rotation is synchronized with its revolution, so that we see only one face of it from the earth. And the tides in the earth's oceans lead to dissipation of energy in the earth-moon system, causing the day to become longer while the moon recedes from the earth. (The energy required to move the moon away from the earth comes from the kinetic energy of the earth's rotation. Conservation of angular momentum requires that the moon recede in this process.) The earth-moon system is heading toward a state in which the length of the day and the length of the month are equal.

Similar concepts apply to a block sliding across a table. Friction causes the kinetic energy of the block to be converted into heat in the block and the table, that is, into the energy of motion of the individual atoms and molecules of the block and table. If friction is small, we may neglect it and treat the block as if its center of mass or center of mass and orientation were isolated degrees of freedom, not interacting with any other.

If we do wish to take friction into account in the motion of a block sliding across a table, we may do so while limiting ourselves to a small number of degrees of freedom by using a phenomenological force law, for example, one that says that the frictional force is $-k|\mathbf{v}|^\alpha \mathbf{v}$, where $k > 0$ and α are constants and where $\mathbf{v}$ is the velocity of the block. Such force laws may have some theoretical justification or may be just fits to experimental data. They amount to taking into account the large number of internal degrees of freedom of the block and table in an average way. It is straightforward to write down Newton's laws $\mathbf{F} = m\mathbf{a}$, including such frictional forces, to obtain differential equations for the motion of the few degrees of freedom we are interested in. Such frictional force laws, however, preclude a Lagrangian formulation of the system.

In this course we will be especially interested in microscopic systems with a small number of degrees of freedom, whose dynamics we will follow in detail, that is, without using statistics. Such systems are sometimes isolated from their environment, so it is not hard to write down equations of motion for their internal dynamics. An example is an atom or molecule in a low-density gas between collisions. Since such systems only have a small number of degrees of freedom, there is no question of any coupling to a large number of "internal" degrees of freedom, as in the case of the block. If the system is not isolated, however, then it does interact with "external" degrees of freedom, that is with external fields (gravitational, electromagnetic, etc).

In such cases we will assume that the external fields are simply given as a function of space and time. In reality, external fields are never known precisely, for example, a magnetic field produced by a current in some coils has fluctuations due to the finite temperature of the source of the current and of the coils themselves. In this course, we shall ignore such effects.

In summary, we shall ignore dissipative effects in this course, so that a Lagrangian description is possible. Systems that can be described by a Lagrangian include classical systems of point particles, fluids, elastic media and fields, both in nonrelativistic and relativistic mechanics. We begin with the nonrelativistic mechanics of systems of point particles.

3. Constrained Systems, Configuration Space, and Generalized Coordinates

In a system with N particles, the positions of all the particles relative to an inertial frame are specified by $3N$ coordinates, $(x_1, y_1, z_1, \ldots, x_N, y_N, z_N)$, or $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ for short. Sometimes the positions of the particles are subject to *constraints*, for example, if a particle is sliding on the surface of a sphere of radius a, its position coordinates are subject to the constraint $x^2 + y^2 + z^2 = a^2$. Or if the particles make up a rigid body, then the distances between any two particles is constant,

$$|\mathbf{x}_i - \mathbf{x}_j| = \text{const}, \tag{1}$$

for all $i, j = 1, \ldots, N$. Constraints appear both in classical mechanics and in quantum mechanics, for example, many small molecules are approximately rigid bodies.

Constrained systems are idealizations of systems with a larger number of degrees of freedom, in which strong confining forces cause the particles to move on or near the constraint surface. For example, a particle sliding on the surface of a sphere of radius a can be thought of as a particle that feels a strong restoring force if it moves away from the surface $r = a$ in three-dimensional space.

Interesting issues arise when exploring such models in detail to see how the simple picture of constrained dynamics emerges from a more realistic model with strong constraining forces. For example, one finds that if the initial position is on the constraint surface and the initial velocity is tangent to it, then the position remains on the constraint surface at later times, to a good approximation when the constraining forces are strong. When the initial position is not on the constraint surface or the initial velocity is not tangent to it, then there is a high-frequency oscillation about the constraint surface and the dynamics along the constraint surface are modified. This is an interesting chapter in adiabatic theory that we shall not pursue here, although we will touch on adiabatic problems in the course (see Notes 35).

In the case of constraints, the positions of the N particles are described by some number $n < 3N$ of coordinates, call them $q_1, \ldots, q_n$. For example, in the case of the particle sliding on the sphere we have $n = 2$ and we can take $(q_1, q_2) = (\theta, \phi)$, the usual spherical coordinates on the sphere. In the case of the rigid body with one point fixed we have $n = 3$ and we can take $(q_1, q_2, q_3) = (\alpha, \beta, \gamma)$, the Euler angles of a rotation that maps a standard orientation of the rigid body into the actual orientation. See Sec. 11.11 for the definition of Euler angles. In the case of a rigid body whose center of mass is free to move (for example, in a model of a planet), we have $n = 6$ and coordinates $(q_1, \ldots, q_6) = (\alpha, \beta, \gamma, X, Y, Z)$, where (X, Y, Z) are the coordinates of the center of mass. In general, the positions $(\mathbf{x}_1, \ldots, \mathbf{x}_N)$ of the particles are functions of the q's,

$$\mathbf{x}_i = \mathbf{x}_i(q_1, \ldots, q_n, t), \qquad i = 1, \ldots, N. \tag{2}$$

If the constraints are time-dependent, then these functions depend on time as indicated. For example, in the case of a particle sliding on a sphere of radius a we have

$$\begin{aligned} x &= a \sin\theta \cos\phi, \\ y &= a \sin\theta \sin\phi, \\ z &= a \cos\theta. \end{aligned} \tag{3}$$

If there are no constraints, then we have $n = 3N$ and we can take $(q_1, \ldots, q_{3N}) = (x_1, y_1, z_1, \ldots, x_N, y_N, z_N)$, that is, we can let the q's be the rectangular coordinates of the particles. We may choose not to do this, however, for example, in the case of a single particle moving in three-dimensional space we may wish to use spherical coordinates and choose $(q_1, q_2, q_3) = (r, \theta, \phi)$.

In general (with or without constraints), the q's are coordinates on *configuration space*, an abstract space in which a single point represents the positions of all the particles of the system. The number of the q's, which is the dimension of configuration space, is called the *number of degrees of freedom*. In the case of a single particle moving in three-dimensional space, configuration space is the same as physical space, but in general configuration space is best regarded as an abstract space. Here are two examples of configuration spaces. The configuration space of a system of N unconstrained particles is $\mathbb{R}^{3N}$, while the configuration space of a rigid body with one point fixed is the rotation group manifold $SO(3)$ (see Notes 11).

The coordinates q_i, $i = 1, \ldots, n$, are called *generalized coordinates*. The word "generalized" simply means that the coordinates need not be rectangular coordinates. For constrained systems such as the particle sliding on a sphere, rectangular coordinates are not an option and all coordinates are "generalized."

4. Variational Principles in Optics

It has been known for a long time that certain problems in optics can be expressed in terms of a variational principle. Consider, for example, the laws of reflection from a mirror, which state that the angle of incidence is equal to the angle of reflection. See Fig. 1, in which a light ray leaves point A, reflects off the mirror at point R, and then reaches point B. The actual path taken by the light ray on going from A to B, path ARB in the figure, is the one that minimizes the total distance of travel. For example, the dotted path $AR'B$ is longer. But the shortest path is also the one for which the angle of incidence is equal to the angle of reflection ($\angle ARC = \angle CRB$ in the figure).

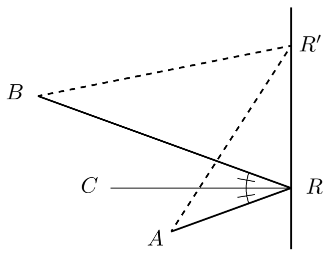

Fig. 1. The actual path taken by a light ray when reflecting from a mirror, ARB, is the one that minimizes the total distance traveled.

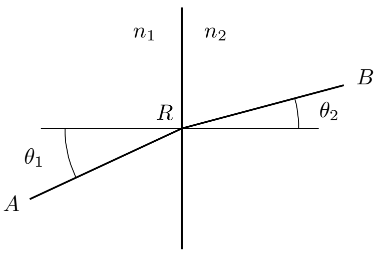

Fig. 2. The path taken by the light ray on passing from one medium with index of refraction n_1 to another with index of refraction n_2 minimizes the quantity $\ell_1 n_1 + \ell_2 n_2$.

Similarly, the laws of refraction (Snell's law) can be cast into a variational form. See Fig. 2, in which there are two media with indices of refraction n_1 and n_2. In this case it is not the total distance that is minimized, but the distance multiplied by the index of refraction. That is, the actual path, ARB in the figure, is the one that minimizes $\ell_1 n_1 + \ell_2 n_2$, where ℓ_1 is the distance from A to a point on the interface, and ℓ_2 is the distance from that point to B. It can be shown that this path is one that satisfies Snell's law,

$$n_1 \sin\theta_1 = n_2 \sin\theta_2. \tag{4}$$

In both these cases there is a space of paths over which the minimization takes place. In the case of the mirror, the space of paths consists of all paths that connect A to a point on the mirror by a straight line, and then connect that point to B by another straight line. This is a 2-parameter space of paths, because the mirror is a 2-dimensional surface (the parameters are the coordinates on the surface of the mirror). But the space of paths can be enlarged to include arbitrary, curved paths that join A to B, staying on one side of the mirror and touching it once, as in Fig. 3. The actual path still has the minimum length, because all curved paths between any two points are longer than the straight path between those points. The enlarged path space is infinite-dimensional. Similarly, in the refraction problem illustrated in Fig. 2, the space of paths can be enlarged to include curved paths that cross the interface only once.

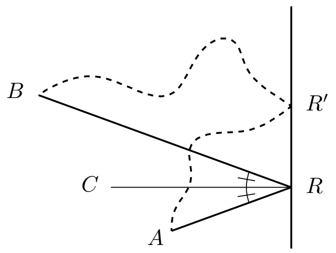

Fig. 3. The space of paths reflecting from the mirror can be enlarged to include curved paths, and the actual path still minimizes the distance.

A more general situation is one in which the index of refraction is a function of position, $n = n(\mathbf{x})$, for example, when a light ray passes through the earth's atmosphere. In this case it can be shown that the actual path taken by the light ray is one that minimizes the integral,

$$A[\mathbf{x}(s)] = \int n(\mathbf{x})\, ds, \tag{5}$$

where the path is parameterized by the arc length s. This integral is the generalization of the sum $\ell_1 n_1 + \ell_2 n_2$ used in the refraction example above. In this case the space of paths consists of all paths that join given points $\mathbf{x}_0$ and $\mathbf{x}_1$. The actual path satisfies the differential equation,

$$\frac{d}{ds}\Big(n(\mathbf{x})\frac{d\mathbf{x}}{ds}\Big) = \nabla n, \tag{6}$$

which is the Euler-Lagrange equation (see Sec. 7) for the variational problem defined by Eq. (5).

The quantity $A[\mathbf{x}(s)]$ in Eq. (5) is a functional of the path, and it is not obvious why it should be minimized along the actual path. A justification can be given by applying the short wavelength or WKB approximation to Maxwell's equations in a slowly varying medium, but that does not make the form of the functional $A[\mathbf{x}(s)]$ intuitive, nor does it explain why nonphysical paths should enter into a physical formulation of the problem. That is, the question arises, if the actual path followed by the light ray minimizes a certain quantity, does nature somehow know about the nonphysical paths in order to reject them? There is no answer purely within classical theory. As discussed in Notes 9, however, a more satisfactory justification comes from quantum mechanics. Suffice it to say here that in books the functional $A[\mathbf{x}(s)]$ is often multiplied by $1/c$, whereupon it becomes the time required for a particle to move along the light ray at the phase velocity $v = c/n$. That is, it becomes a time functional,

$$T[\mathbf{x}(s)] = \frac{1}{c} A[\mathbf{x}(s)] = \int \frac{n}{c}\, ds = \int \frac{ds}{v} = \int dt. \tag{7}$$

Thus, one can say that the actual path minimizes the time, defined in this way. Why the phase velocity should appear here rather than the group velocity (which is the actual velocity at which signals propagate) is however a mystery.

Actually, the physical path does not always minimize the time. More generally it is a critical or stationary point of the time functional, that is, the physical path satisfies

$$\frac{\delta T}{\delta \mathbf{x}(s)} = 0. \tag{8}$$

The first functional derivative of the time functional with respect to the path vanishes on the physical path.

5. Critical Points

In elementary calculus, if we wish to find the extrema of a function $f(x_1, \ldots, x_n)$ of several variables, we first find the roots of

$$\frac{\partial f}{\partial x_i}(x_{01}, \ldots, x_{0n}) = 0, \qquad i = 1, \ldots, n, \tag{9}$$

that is, we look for the places where the gradient of f vanishes. These roots $(x_{01}, \ldots, x_{0n})$ are called the *critical points* of the function f. We can describe a critical point in words by saying small variations about the critical point give rise to only second order variations in the value of the function (the first order variations vanish).

Critical points are candidates for extrema. To find out if one is an extremum and of what type, we may examine the second derivative matrix evaluated at the critical point,

$$\frac{\partial^2 f}{\partial x_i \partial x_j}(x_{01}, \ldots, x_{0n}). \tag{10}$$

This matrix is real and symmetric, so its eigenvalues are real. If the eigenvalues are all positive, then the critical point is a minimum, that is, small variations about the critical point can only increase the value of the function. In general the minimum is only a local minimum, that is, there may be other minima with a smaller value of f. If all the eigenvalues of the matrix (10) are negative, then the critical point is a (generally local) maximum. If some are negative and some are positive, then the critical point is a saddle. And if some eigenvalues are zero, then any test based on second derivatives alone is inconclusive.

Similarly, the condition (7) can be stated by saying that the physical path is a critical point (in path space) of the time functional, that is, the functional is stationary on the physical paths.

6. Hamilton's Principle

There is also a variational formulation of nondissipative systems in classical mechanics. It says that the functional

$$A[q(t)] = \int_{t_0}^{t_1} L(q, \dot{q}, t)\, dt \tag{11}$$

is stationary on the physical paths, where L is the Lagrangian function. Here q is short for $(q_1, \ldots, q_n)$, the generalized coordinates, and n is the number of degrees of freedom. Equivalently,

$$\frac{\delta A}{\delta q_i(t)} = 0 \tag{12}$$

on the physical paths, for $i = 1, \ldots, n$. This is called *Hamilton's principle*. The quantity $A[q(t)]$ is called the *action* associated with the path $q(t)$. In some cases, the action is actually minimum along the physical path, but in general the physical path is just a critical point of the action functional.

The space of paths that enters into Hamilton's principle is the space of all smooth paths $q(t)$ satisfying

$$q(t_0) = q_0, \qquad q(t_1) = q_1, \tag{13}$$

for fixed values of t_0, t_1, q_0 and q_1. Here as above q stands for all the generalized coordinates $(q_1, \ldots, q_n)$. The endtimes t_0 and t_1 are the limits of the integral (11). If q is one-dimensional, paths in this path space may be visualized in the q-t plane as in Fig. 4.

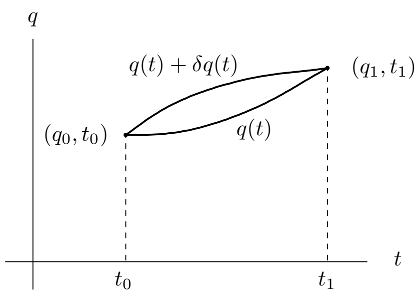

Fig. 4. A path $q(t)$ satisfying fixed boundary conditions $q(t_0) = q_0$ and $q(t_1) = q_1$. Also shown is a modified path $q(t) + \delta q(t)$, satisfying the same endpoint and endtime conditions.

7. The Euler-Lagrange Equations

The functional derivative (12) can be computed as follows. Let $q(t)$ be any path in the path space, that is, any path satisfying the boundary conditions. It need not be a physical path, that is, one that satisfies Newton's laws. Let $\delta q(t)$ be a small variation that takes the path $q(t)$ into a nearby path $q(t) + \delta q(t)$, as in Fig. 4. The new path is also required to be in the path space, so $\delta q(t_0) = \delta q(t_1) = 0$. The velocity along the original path is $\dot{q}(t)$ and that along the modified path is $\dot{q}(t) + \delta\dot{q}(t)$, where

$$\delta\dot{q}(t) = \frac{d\delta q(t)}{dt}. \tag{14}$$

Then the variation in the action is

$$\delta A = A[q(t) + \delta q(t)] - A[q(t)] = \int_{t_0}^{t_1} dt \sum_{i=0}^{n} \Big[\frac{\partial L}{\partial q_i}\,\delta q_i(t) + \frac{\partial L}{\partial \dot{q}_i}\,\delta\dot{q}_i(t)\Big], \tag{15}$$

dropping terms that are second order or higher in δq. But the second term on the right can be integrated by parts,

$$\int_{t_0}^{t_1} dt \sum_{i=0}^{n} \frac{\partial L}{\partial \dot{q}_i}\,\delta\dot{q}_i = \sum_{i=0}^{n} \frac{\partial L}{\partial \dot{q}_i}\,\delta q_i\Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} dt \sum_{i=0}^{n} \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{q}_i}\Big)\,\delta q_i(t), \tag{16}$$

where the first major term on the right vanishes because of the boundary conditions $\delta q(t_0) = \delta q(t_1) = 0$. Altogether, we have

$$\delta A = \int_{t_0}^{t_1} dt \sum_{i=0}^{n} \Big[\frac{\partial L}{\partial q_i} - \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{q}_i}\Big)\Big]\delta q_i(t). \tag{17}$$

Thus by the definition of the functional derivative,

$$\frac{\delta A}{\delta q_i(t)} = \frac{\partial L}{\partial q_i} - \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{q}_i}\Big). \tag{18}$$

This is the quantity which vanishes along the physical paths, so the equations of motion satisfied by the physical paths are

$$\frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{q}_i}\Big) = \frac{\partial L}{\partial q_i}, \tag{19}$$

for $i = 1, \ldots, n$. These are the *Euler-Lagrange equations.*

An ordinary function over some domain may have any number of critical points, from zero to infinity. Similarly, for a given path space, that is, for given values of t_0, t_1, q_0 and q_1, the action functional (11) may have any number of critical points. That is, for the given endpoints and endtimes, there may be any number of physical paths, from zero to infinity. Many books erroneously state that the physical path is unique for given endpoints and endtimes.

8. The Lagrangian, and the Case $L = T - V$

To say that a classical system admits a Lagrangian formulation means that there exists a function $L(q, \dot{q}, t)$ such that the Euler-Lagrange equations (19) are equivalent to Newton's laws. As mentioned, such a formulation exists whenever the system is nondissipative. This is shown by actually finding the Lagrangian function in different cases. We will not go through the proof of equivalence to Newton's laws here, instead we will just quote the Lagrangian function for the cases of greatest interest to us. The actual proof is straightforward for unconstrained systems, but more elaborate in the case of constraints, especially time-dependent ones.

The most important case for this course is that of nonrelativistic systems in which the forces can be obtained from a scalar potential. Let $V(\mathbf{x}_1, \ldots, \mathbf{x}_N, t)$ be the potential energy of the system, expressed as a function of the positions of the N particles and possibly the time. It is assumed that there is a force on particle i given by

$$\mathbf{F}_i = -\frac{\partial V(\mathbf{x}_1, \ldots, \mathbf{x}_N, t)}{\partial \mathbf{x}_i}, \qquad i = 1, \ldots, N. \tag{20}$$

In addition, if there are constraints, then there are also forces of constraint acting on the particles. We do not need to take explicit account of the forces of constraint in a Lagrangian formulation, but they must be dealt with when using Newton's laws. We assume that the forces (20) and the forces of constraint are the only forces acting on the particles. By using the relations (2), the potential energy can be expressed in terms of the generalized coordinates and possibly the time,

$$V = V(q_1, \ldots, q_n, t). \tag{21}$$

Similarly, by differentiating (2) with respect to time and using the chain rule, the kinetic energy T can be expressed as a function of the q's, the $\dot{q}$'s and possibly the time,

$$T = \frac{1}{2}\sum_{i=1}^{N} m_i |\dot{\mathbf{x}}|^2 = \frac{1}{2}\sum_{i=1}^{N} m_i \left| \frac{\partial \mathbf{x}_i}{\partial t} + \sum_{k=1}^{n} \frac{\partial \mathbf{x}_i}{\partial q_k} \dot{q}_k \right|^2 = T(q, \dot{q}, t), \tag{22}$$

where m_i is the mass of particle i and where n is the number of degrees of freedom.

In this case the Lagrangian is the difference between the kinetic and potential energies,

$$L(q, \dot{q}, t) = T(q, \dot{q}, t) - V(q, t). \tag{23}$$

An important case in practice is that of unconstrained particles interacting by means of electrostatic forces, including "external" electrostatic fields. In this case the potential energy of the N particles is

$$V(\mathbf{x}_1, \ldots, \mathbf{x}_N, t) = \frac{1}{2}\sum_{i \neq j} \frac{q_i q_j}{|\mathbf{x}_i - \mathbf{x}_j|} + \sum_{i=1}^{N} q_i \, \phi_{\text{ext}}(\mathbf{x}_i, t), \tag{24}$$

where q_i is the charge of particle i and where $\phi_{\text{ext}}(\mathbf{x}, t)$ is the electrostatic potential energy per unit charge due to external fields. The first sum on the right is the electrostatic energy of interaction of the particles among themselves. The sum is taken over all pairs (i,j), omitting the diagonal terms $i=j$ which represent the infinite self-energy of the charged particles. The second term is the external potential, corresponding to an external electric field,

$$\mathbf{E}_{\text{ext}}(\mathbf{r},t) = -\nabla\phi_{\text{ext}}(\mathbf{r},t). \tag{25}$$

Here we use $\mathbf{r}$ for an arbitrary field point at which the field may be evaluated.

The external potential may depend on time if the charges producing the external fields are moving. Think, for example, of an atom between the plates of a capacitor to which an AC voltage is applied. In this case the external electric field may be written,

$$\mathbf{E}_{\text{ext}} = E_0\hat{\mathbf{z}}\cos\omega t, \tag{26}$$

where E_0 is the amplitude of the electric field, assumed to be pointing in the z-direction, and ω is the frequency. Then the external potential is

$$\phi_{\text{ext}} = -E_0 z\cos\omega t. \tag{27}$$

The potential (24) only takes into account the interactions of the charged particles in the electrostatic approximation, which is valid for nonrelativistic systems of small spatial extent. That is, the equations of motion are equivalent to

$$m_i\ddot{\mathbf{x}}_i = q_i[\mathbf{E}_{\text{int}}(\mathbf{x}_1,\ldots,\mathbf{x}_N) + \mathbf{E}_{\text{ext}}(\mathbf{x}_i,t)]. \tag{28}$$

This approximation neglects magnetic effects as well as the effects of retardation and radiation. See Notes 36 for further discussion.

9. Magnetic Fields

The Lagrangian formulation may also be applied to problems with magnetic fields. The simplest case is one in which the magnetic fields are external, that is, we ignore any magnetic fields arising from the motion of the charged particles of the system itself. We write the external magnetic field in the form

$$\mathbf{B}_{\text{ext}}(\mathbf{r},t) = \nabla\times\mathbf{A}_{\text{ext}}(\mathbf{r},t), \tag{29}$$

where $\mathbf{A}_{\text{ext}}$ is the vector potential. As indicated, the external magnetic field is allowed to depend on time. Then the Lagrangian is

$$L = T - V + \sum_{i=1}^{N}\frac{q_i}{c}\dot{\mathbf{x}}_i\cdot\mathbf{A}_{\text{ext}}(\mathbf{x}_i,t). \tag{30}$$

If V is given by Eq. (24), then the Euler-Lagrange equations are equivalent to

$$m_i\ddot{\mathbf{x}}_i = q_i\Big[\mathbf{E}_{\text{int}}(\mathbf{x}_1,\ldots,\mathbf{x}_N) + \mathbf{E}_{\text{ext}}(\mathbf{x}_i,t) + \frac{1}{c}\dot{\mathbf{x}}\times\mathbf{B}_{\text{ext}}(\mathbf{x}_i,t)\Big]. \tag{31}$$

If it is desired to take into account the internal magnetic fields produced by the moving charges of the system itself, this may be done in an approximate way by adding extra terms to obtain the *Darwin Lagrangian*. See the discussion in Landau and Lifshitz or Jackson. The Darwin Lagrangian has more physics in it than the Lagrangian (30), but it still omits the effects of retardation and radiation. If these are important, one must include the infinite degrees of freedom of the electromagnetic field. See the discussion in Notes 36.

10. Advantages of Lagrangian Mechanics

The Lagrangian formulation of mechanics affords several advantages compared to Newton's laws. First, the Lagrangian is a single (scalar) function, which implicitly contains within it all the equations of motion. Thus, if we wish to transform to a new coordinate system, it is much easier to transform the (single, scalar) Lagrangian than the (vector, multicomponent) equations of motion. The Lagrangian formulation is the easiest way to write down the classical equations of motion in curvilinear coordinates.

In addition, if there are constraints, the Lagrangian formulation automatically gives us the correct equations of motion without having to take the forces of constraint into account. This is considerably simpler than working with Newton's laws, where the forces of constraint must be eliminated.

Third, there is a general relationship in Lagrangian mechanics between symmetries and invariants. That is, every symmetry of the system corresponds to a conservation law. This is called *Noether's theorem*.

Finally, the Lagrangian formulation of a classical system leads to the Hamiltonian formulation, which is the point at which we may pass to quantum mechanics.

The main disadvantage of the Lagrangian formulation is that it is unable to account for dissipative effects.

11. Lagrangians for Relativistic Problems

In relativistic mechanics we require that the equations of motion be covariant, that is, that they have the same form in all Lorentz frames. The easiest way to guarantee Lorentz covariance is to require that the action $A[q(t)]$ be a Lorentz scalar, see Eq. (11). This in turn implies that $L\,dt$ be a Lorentz scalar. But

$$L\,dt = L\frac{dt}{d\tau}\,d\tau = L\gamma\,d\tau, \tag{32}$$

where τ is proper time and where γ is the usual relativistic factor of time dilation,

$$\gamma = \frac{1}{\sqrt{1-(v/c)^2}}. \tag{33}$$

Since $d\tau$ is a Lorentz scalar, the quantity $L\gamma$ must also be a Lorentz scalar.

This is a powerful constraint on the form of the Lagrangian which determines it almost uniquely. Take, for example, the case of a free particle. The Lagrangian cannot depend on the particle 4-vector x^μ if it is to be invariant under space-time translations. As for the first derivatives of x^μ, the only Lorentz scalar that can be constructed out of these is

$$\frac{dx^\mu}{d\tau}\frac{dx_\mu}{d\tau} = c^2. \tag{34}$$

Here we use Jackson's conventions for special relativity, in which $x^\mu = (x^0, x^1, x^2, x^3) = (ct, x, y, z)$, and $g_{\mu\nu} = \mathrm{diag}(+1, -1, -1, -1)$. Thus $L\gamma$ for a free relativistic particle must be a function of the invariant scalar product (34). We take it to be

$$L\gamma = -mc\Big(\frac{dx^\mu}{d\tau}\frac{dx_\mu}{d\tau}\Big)^{1/2} = -mc^2, \tag{35}$$

or,

$$L = -\frac{mc^2}{\gamma} = -mc^2\sqrt{1 - \frac{|\dot{\mathbf{x}}|^2}{c^2}}. \tag{36}$$

We choose the form (35) because it has the right dimensions for a Lagrangian (dimensions of energy) and because it reduces to the correct nonrelativistic limit when the velocity is small. We see this if we expand out the square root in Eq. (36),

$$L = -mc^2\Big(1 - \frac{|\dot{\mathbf{x}}|^2}{2c^2} + \ldots\Big) = -mc^2 + \frac{m}{2}|\dot{\mathbf{x}}|^2 + \ldots \tag{37}$$

The constant $-mc^2$ has no effect on the equations of motion and can be dropped, leaving the nonrelativistic Lagrangian for a free particle. The Lagrangian for the free particle is the kinetic energy only in the nonrelativistic approximation (through order $(v/c)^2$).

Similarly, we may derive the Lagrangian for a relativistic charged particle interacting with a given electromagnetic field. This field corresponds to the "external" field in Secs. 8 and 9, but we drop the "ext" subscripts. The electric and magnetic fields can be derived from scalar and vector potentials ϕ and $\mathbf{A}$,

$$\begin{aligned} \mathbf{E} &= -\nabla\phi - \frac{1}{c}\frac{\partial\mathbf{A}}{\partial t}, \\ \mathbf{B} &= \nabla\times\mathbf{A}, \end{aligned} \tag{38}$$

or, in relativistic notation,

$$F_{\mu\nu} = \frac{\partial A_\nu}{\partial x^\mu} - \frac{\partial A_\mu}{\partial x^\nu}, \tag{39}$$

where

$$A^\mu = (A^0, A^1, A^2, A^3) = (\phi, \mathbf{A}). \tag{40}$$

The simplest Lorentz scalar that can be composed out of the particle 4-velocity $dx^\mu/d\tau$ and the electromagnetic field is

$$\frac{dx^\mu}{d\tau}A_\mu, \tag{41}$$

so we guess that $L\gamma$ for a particle interacting with a given electromagnetic field is the free particle expression (35) plus a term proportional to the term (41). Adjusting the proportionality constant to get the right equations of motion, we find

$$L\gamma = -mc\Big(\frac{dx^\mu}{d\tau}\frac{dx_\mu}{d\tau}\Big)^{1/2} - \frac{q}{c}\frac{dx^\mu}{d\tau}A_\mu, \tag{42}$$

or,

$$L = -mc^2\sqrt{1-\frac{|\dot{\mathbf{x}}|^2}{c^2}} - q\phi(\mathbf{x},t) + \frac{q}{c}\dot{\mathbf{x}}\cdot\mathbf{A}(\mathbf{x},t). \tag{43}$$

Expanding this out to lowest order in $(v/c)^2$ and dropping the constant $-mc^2$, we obtain the nonrelativistic Lagrangian for a charged particle interacting with an electromagnetic field,

$$L = \frac{m}{2}|\dot{\mathbf{x}}|^2 - q\phi(\mathbf{x},t) + \frac{q}{c}\dot{\mathbf{x}}\cdot\mathbf{A}(\mathbf{x},t). \tag{44}$$

This is a single-particle version of Eq. (30).

We do not attempt to write down a relativistic Lagrangian for a system of more than one particle interacting by means of electromagnetic forces, since in many cases it is impossible to do this without including the infinite number of degrees of freedom of the field. The electrostatic or Darwin approximations discussed above are not valid when the relative velocities of particles are relativistic.

12. The Momentum

Let us return to the Lagrangian $L(q,\dot{q},t)$, expressed in terms of generalized coordinates. We define the quantity

$$p_i = \frac{\partial L}{\partial \dot{q}_i}, \tag{45}$$

which we call the *generalized momentum conjugate to* q_i. In terms of p_i, the Euler-Lagrange equations (19) can be written,

$$\frac{dp_i}{dt} = \frac{\partial L}{\partial q_i}. \tag{46}$$

The physical meaning of the generalized momentum depends on the coordinates used and the nature of the forces.

Consider first Cartesian coordinates, for which we write the momentum conjugate to $\mathbf{x}_i$ as $\mathbf{p}_i$ (a 3-vector),

$$\mathbf{p}_i = \frac{\partial L}{\partial \dot{\mathbf{x}}_i}. \tag{47}$$

Using the nonrelativistic Lagrangian (44) for a single charged particle interacting with a given (external) electromagnetic field, we find

$$\mathbf{p} = m\dot{\mathbf{x}} + \frac{q}{c}\mathbf{A}(\mathbf{x},t). \tag{48}$$

The first term, $m\dot{\mathbf{x}}$, is the usual definition of the momentum in elementary mechanics. We call this term alone the *kinetic momentum*. In the presence of a nonzero vector potential, however, there is a second term. We call the total momentum of Eq. (48) the *canonical momentum*. The canonical momentum depends on the choice of gauge for the vector potential, so it is not as physical as the kinetic momentum (its value depends on the gauge convention). For this reason one may prefer formulations based on the kinetic momentum. The canonical momentum, however, is necessary for a Hamiltonian formulation of mechanics in the presence of magnetic fields, and it is the classical momentum that corresponds to the momentum operator $-i\hbar\nabla$ in quantum mechanics.

The extra term in the canonical momentum (48) is somewhat mysterious in this formulation of particle mechanics, but it acquires an explanation in terms of the momentum of the electromagnetic field when the field degrees of freedom are included in the system. This is especially simple in the case of Coulomb gauge. This point is discussed further in Notes 36.

Using the relativistic Lagrangian (43), we find the canonical momentum

$$\mathbf{p} = m\gamma\dot{\mathbf{x}} + \frac{q}{c}\mathbf{A}, \tag{49}$$

of which the first term is the usual definition of momentum in relativistic theory (it is in effect the relativistic "kinetic" momentum). The second term again can be expressed in terms of the momentum of the field.

In curvilinear coordinates the momentum has other interpretations. For example, consider the case of a particle moving in a potential V in 3-dimensional space, for which the Lagrangian in spherical coordinates is

$$L = \frac{m}{2}(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2) - V(r,\theta,\phi). \tag{50}$$

Then the momentum conjugate to the azimuthal angle ϕ is

$$p_\phi = \frac{\partial L}{\partial\dot{\phi}} = mr^2\sin^2\theta\,\dot{\phi} = \hat{\mathbf{z}}\cdot(\mathbf{x}\times\mathbf{p}), \tag{51}$$

which physically is the z-component of the angular momentum $\mathbf{L}$, as indicated.

To see how the generalized momentum changes under a change of coordinates, let us temporarily write q^i (with a superscript, that is, a contravariant index) for the coordinates, and let us consider a change of coordinates $q'^{\,i} = q'^{\,i}(q)$. For simplicity we assume the coordinate transformation is time-independent. Then the generalized velocity transforms according to the chain rule,

$$\dot{q}'^{\,i} = \sum_j \frac{\partial q'^{\,i}}{\partial q^j}\,\dot{q}^j, \tag{52}$$

which is the transformation law for a contravariant vector. Similarly, the generalized momentum transforms as

$$p'_i = \sum_j \frac{\partial q^j}{\partial q'^{\,i}}\,p_j, \tag{53}$$

that is, as a covariant vector. For example, in spherical coordinates, the generalized momenta (p_r, p_θ, p_ϕ) are the covariant components of the momentum vector $\mathbf{p}$ in spherical coordinates.

Taken together, Eqs. (52) and (53) imply that the quantity $\sum_i p_i\, dq^i/dt$, and hence the differential form $\sum_i p_i\, dq^i$, are invariants under general coordinate transformations on configuration space.

13. Regular and Irregular Lagrangians

Expanding out the time derivative in the Euler-Lagrange equations (19), we have

$$\sum_j \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j}\, \ddot{q}_j = -\sum_j \frac{\partial^2 L}{\partial \dot{q}_i \partial q_j}\, \dot{q}_j - \frac{\partial^2 L}{\partial \dot{q}_i \partial t} + \frac{\partial L}{\partial q_i}, \tag{54}$$

which can be solved for the generalized accelerations $\ddot{q}_i$ if

$$\det \frac{\partial^2 L}{\partial \dot{q}_i \partial \dot{q}_j} \neq 0. \tag{55}$$

Lagrangians for which Eq. (55) holds are said to be *regular.* All Lagrangians normally encountered in particle mechanics are regular, and, in particular, it is easily verified that all the Lagrangians cited in this Appendix are regular. Irregular Lagrangians are the norm, however, in relativistic field theory.

For a regular Lagrangian, we can solve the Euler-Lagrange for the generalized accelerations,

$$\ddot{q}_i = F_i(q, \dot{q}, t), \tag{56}$$

where F_i are some functions. That is, the equations of motion are a system of coupled, second-order differential equations on configuration space. These can be converted into a system of coupled, first-order differential equations by introducing the generalized velocities v_i, and writing,

$$\begin{aligned} \dot{q}_i &= v_i, \\ \dot{v}_i &= F_i(q, v, t). \end{aligned} \tag{57}$$

14. Hamilton's Equations

The condition (55) is also the condition that the definition (45) of the generalized momentum can be solved for the velocities, that is, that we can express the velocities as functions of the q's, the p's and the time:

$$\dot{q}_j = \dot{q}_j(q, p, t), \tag{58}$$

where p stands for all the generalized momenta, $(p_1, \ldots, p_n)$. Thus for regular Lagrangians we can introduce the p's as new coordinates, eliminate the v's from Eqs. (57), and express the equations of motion as a system of first-order differential equations involving the q's and the p's.

The equations of motion can be put into an especially symmetrical form in terms of the *Hamiltonian* function, defined by

$$H(q,p,t) = \sum_j p_j \dot{q}_j - L. \tag{59}$$

The right-hand side of this expression is a function of $(q, \dot{q}, t)$, in particular, p_i emerges as a function of $(q, \dot{q}, t)$ due to its definition (45), but we use Eq. (58) to eliminate the $\dot{q}$'s in favor of the p's, and express H as a function of (q, p, t) as shown.

Now we compute $\partial H/\partial q_i$, noting carefully that this derivative means holding p fixed. We have

$$\frac{\partial H}{\partial q_i} = \sum_j p_j \frac{\partial \dot{q}_j}{\partial q_i} - \frac{\partial L}{\partial q_i} - \sum_j \frac{\partial L}{\partial \dot{q}_j} \frac{\partial \dot{q}_j}{\partial q_i}, \tag{60}$$

where $\partial \dot{q}_j/\partial q_i$ means the derivative of functions (58), taken at fixed p. But by Eq. (45) the first and third terms on the right of Eq. (60) cancel, while the second term on the right can be simplified with the help of Eq. (46). The result is $\partial H/\partial q_i = -\dot{p}_i$. Similarly, we compute $\partial H/\partial p_i$, which means holding q fixed. We have

$$\frac{\partial H}{\partial p_i} = \dot{q}_i + \sum_j p_j \frac{\partial \dot{q}_j}{\partial p_i} - \sum_j \frac{\partial L}{\partial \dot{q}_j} \frac{\partial \dot{q}_j}{\partial p_i}, \tag{61}$$

where $\partial \dot{q}_j/\partial p_i$ means the derivative of functions (58), taken at fixed q. But by Eq. (45) the two sums on the right hand side cancel, and we have $\partial H/\partial p_i = \dot{q}_i$.

In summary, we find

$$\boxed{\frac{dq_i}{dt} = \frac{\partial H}{\partial p_i}, \qquad \frac{dp_i}{dt} = -\frac{\partial H}{\partial q_i}.} \tag{62}$$

These are *Hamilton's equations.* They are equivalent to the Euler-Lagrange equations in the case of a regular Lagrangian. Hamilton's equations display a notable symmetry between q and p.

15. Examples of Hamiltonians

Let us take a nonrelativistic, unconstrained system of N particles interacting by means of a potential $V(\mathbf{x}_1, \ldots, \mathbf{x}_N, t)$, such as the electrostatic potential (24). The kinetic energy is

$$T = \frac{1}{2} \sum_{i=1}^{N} m_i |\dot{\mathbf{x}}_i|^2, \tag{63}$$

and the Lagrangian is $L = T - V$. The momentum conjugate to $\mathbf{x}_i$ is

$$\mathbf{p}_i = m_i \dot{\mathbf{x}}_i, \tag{64}$$

which can be inverted to solve for $\dot{\mathbf{x}}_i$ as a function of $\mathbf{p}_i$,

$$\dot{\mathbf{x}}_i = \frac{1}{m_i} \mathbf{p}_i, \tag{65}$$

a special case of Eq. (58). Then according to Eq. (59) the Hamiltonian is

$$H = \sum_{i=1}^{N} \mathbf{p}_i \cdot \dot{\mathbf{x}}_i - L. \tag{66}$$

But the first term is twice the kinetic energy, $2T$, so altogether,

$$H = T + V = \sum_{i=1}^{N} \frac{|\mathbf{p}_i|^2}{2m_i} + V(\mathbf{x}_1, \ldots, \mathbf{x}_N, t), \tag{67}$$

where we have used Eq. (65) to eliminate $\dot{\mathbf{x}}_i$ in favor of $\mathbf{p}_i$. In this example, the Hamiltonian is the sum of the kinetic plus the potential energy, that is, it is the total energy.

Let us take the previous example and add an external magnetic field, producing the Lagrangian (30). Now the canonical momentum is

$$\mathbf{p}_i = m_i \dot{\mathbf{x}}_i + \frac{q_i}{c} \mathbf{A}(\mathbf{x}_i, t), \tag{68}$$

where we drop the "ext" subscript on $\mathbf{A}$. This can be solved for $\dot{\mathbf{x}}_i$,

$$\dot{\mathbf{x}}_i = \frac{1}{m_i} \Big[\mathbf{p}_i - \frac{q_i}{c} \mathbf{A}(\mathbf{x}_i, t) \Big], \tag{69}$$

giving another version of Eq. (58). Now computing H according to Eq. (59) and using Eq. (69) to eliminate $\dot{\mathbf{x}}_i$ in favor of $\mathbf{p}_i$, we find

$$H = \sum_{i=1}^{N} \frac{1}{2m_i} \Big[\mathbf{p}_i - \frac{q_i}{c} \mathbf{A}(\mathbf{x}_i, t) \Big]^2 + V(\mathbf{x}_1, \ldots, \mathbf{x}_N, t). \tag{70}$$

We will use this Hamiltonian frequently in this course. It is important to note that the first term is still the kinetic energy, just expressed in a somewhat complicated form because of the use of the canonical momentum (68).

Even in the presence of a magnetic field, the total energy of the system is still the sum of the kinetic and the potential energies. This corresponds to the physical fact that magnetic forces do not change the energy of a particle, because they are always orthogonal to the velocity vector.

The Hamiltonian is the total energy of the system in other examples, too, such as in relativistic problems. More generally, we may *define* the total energy of the system to be the Hamiltonian.

16. Phase Space

The space upon which (q, p) are coordinates is called *phase space*. Phase space can be thought of as the state space of the classical system, that is, the space in which a single point represents the *dynamical state* of the system. The dynamical state of a classical system is the specification of enough information to determine the future evolution of the system. Specifying a point of configuration space (the q's) is not enough, instead we must also include the velocities (the $\dot{q}$'s), or, alternatively, the momenta (the p's).

A point (q, p) of phase space represents a complete set of initial conditions because Hamilton's equations on phase space are first order in time. We can visualize the flow in phase space through the vector field with components $(\dot{q}, \dot{p})$, which is called a *Hamiltonian vector field.* Figure 5 illustrates the Hamiltonian vector field in the 2-dimensional phase space for the harmonic oscillator,

$$H = \frac{p^2}{2m} + \frac{m\omega^2 q^2}{2}. \tag{71}$$

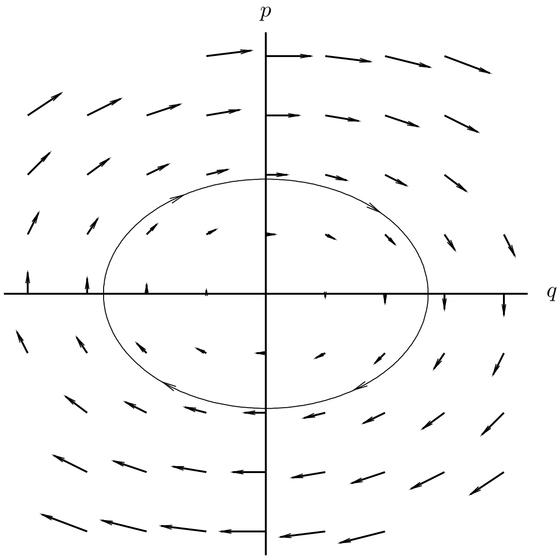

Fig. 5. The Hamiltonian vector field in phase space for the harmonic oscillator. One orbit (an ellipse) is shown.

The orbit passing through a point of phase space is obtained by following the arrows of the Hamiltonian vector field. The orbit is tangent to the vector field at each point. One orbit of the harmonic oscillator (an ellipse) is illustrated in Fig. 5. Theorems about differential equations guarantee a unique solution $q(t)$, $p(t)$ passing through a given initial point (q_0, p_0) at $t = t_0$ when the vector field is smooth. Thus, there is a unique orbit passing through each point of phase space. Orbits cannot cross one another in phase space, because the intersection point would have two orbits passing through it. Orbits in configuration space, however, are free to cross one another, since the equations of motion on configuration space are second order in time.

17. Time Evolution of Observables

A *classical observable* is simply any function of q, p, and possibly t. If the observable depends on t it is said to have an *explicit* time-dependence. An example of an explicitly time-dependent observable is the electrostatic potential (27). In addition, any observable has a time-dependence when seen by a particle (or phase point) moving along its orbit, due to the time-dependence of q and

p. This is called an *implicit* time-dependence. For an observable $A(q,p,t)$, the total time derivative as seen by the phase point moving along its orbit is

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \sum_i \Bigl(\frac{\partial A}{\partial q_i}\,\dot{q}_i + \frac{\partial A}{\partial p_i}\,\dot{p}_i\Bigr) = \frac{\partial A}{\partial t} + \sum_i \Bigl(\frac{\partial A}{\partial q_i}\frac{\partial H}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial H}{\partial q_i}\Bigr), \tag{72}$$

where we have used Hamilton's equations (62) in the final form. The first term $\partial A/\partial t$ in Eq. (72) represents the explicit time-dependence, while the sum represents the implicit time-dependence.

18. Conservation of Energy

Setting $A = H$ in Eq. (72), we find the rate of change of the Hamiltonian itself along an orbit,

$$\frac{dH}{dt} = \frac{\partial H}{\partial t}, \tag{73}$$

since in that case the sum in Eq. (72) vanishes. In particular, if the Hamiltonian has no explicit time-dependence, then $\partial H/\partial t = 0$, and the Hamiltonian is conserved. This is the general law of *conservation of energy.*

The energy (the Hamiltonian) is not conserved if the Hamiltonian has an explicit time-dependence. In practice, such a time-dependence usually arises when a system is interacting with time-dependent, "external" forces, for example, the atom between the capacitor plates that feels the time-dependent electric field (26). Notice in this case that we assume that the external electric field is just a given function of space and time, and we ignore any reaction of the atom back on the charges in the capacitor. In reality, the atom consists of charged particles that produce their own electric field, which has an effect on the external system producing the external electric field. If we include the external charges in our definition of "the system" as well as the back reaction of the atom on these charges, then we have an enlarged system in which energy is conserved.

Similarly, consider a space craft that is obtaining a "gravity boost" by passing close to the planet Jupiter. Because of the large disparity in the masses of the spacecraft and of Jupiter, it is a good approximation to assume that the position of Jupiter is just a given function of time, independent of the position of the spacecraft. Then the spacecraft feels a time-dependent gravitational field due to Jupiter, from which it can extract energy and get expelled from the solar system. In reality, Jupiter does react to the spacecraft, and moves slightly closer to the sun as the spacecraft leaves the solar system. Overall, energy is conserved.

19. The Poisson Bracket

The final sum in Eq. (72) is an example of a Poisson bracket, which is defined in general as follows. Let A and B be any two classical observables, possibly time-dependent. Then the *Poisson bracket* of A and B is another classical observable, given by

$$\{A,B\} = \sum_{i=1}^{n} \Bigl(\frac{\partial A}{\partial q_i}\frac{\partial B}{\partial p_i} - \frac{\partial A}{\partial p_i}\frac{\partial B}{\partial q_i}\Bigr). \tag{74}$$

In terms of the Poisson bracket, Eq. (72) can be written,

$$\frac{dA}{dt} = \frac{\partial A}{\partial t} + \{A, H\}. \tag{75}$$

The Poisson bracket has the following properties. In the following we use capital letters for classical observables, and lower case letters for numbers (real or complex). First, it is linear in both operands,

$$\begin{aligned} \{a_1 A_1 + a_2 A_2, B\} &= a_1\{A_1, B\} + a_2\{A_2, B\}, \\ \{A, b_1 B_1 + b_2 B_2\} &= b_1\{A, B_1\} + b_2\{A, B_2\}. \end{aligned} \tag{76}$$

Second, it is antisymmetric in the two operands,

$$\{A, B\} = -\{B, A\}. \tag{77}$$

Third, it obeys the Leibnitz rule,

$$\{A, BC\} = B\{A, C\} + \{A, B\}C. \tag{78}$$

Closely related to the Leibnitz rule is the chain rule for Poisson brackets, which is the following. If observable B depends on a collection of other observables, say,

$$B = B(C_1, \ldots, C_K), \tag{79}$$

then

$$\{A, B\} = \sum_{k=1}^{K} \{A, C_k\} \frac{\partial B}{\partial C_k}. \tag{80}$$

A similar expression holds when A depends on further observables.

Fourth, it satisfies the Jacobi identity,

$$\{A, \{B, C\}\} + \{B, \{C, A\}\} + \{C, \{A, B\}\} = 0. \tag{81}$$

All these properties are straightforward to prove from the definition.

The Poisson brackets of the phase space coordinates, the q's and p's, among themselves are notable. These are

$$\{q_i, q_j\} = 0, \qquad \{q_i, p_j\} = \delta_{ij}, \qquad \{p_i, p_j\} = 0. \tag{82}$$

20. Conserved Quantities

In general, a *conserved quantity* or *constant of the motion* is a classical observable that is constant as we follow the Hamiltonian flow along an orbit. That is, C is a constant of motion if

$$\frac{dC}{dt} = \frac{\partial C}{\partial t} + \{C, H\} = 0. \tag{83}$$

A *time-independent* constant of the motion is one with no explicit time-dependence, so that $\partial C/\partial t = 0$. It follows that time-independent constants of the motion are classical observables whose Poisson bracket with the Hamiltonian vanishes. As we say, such observable *Poisson commute* with the Hamiltonian.

In particular, the Hamiltonian Poisson commutes with itself, $\{H, H\} = 0$, due to the antisymmetry of the Poisson bracket, property (77). Thus we recover Eq. (73).

21. Liouville's Theorem

Liouville's theorem states that the volume of a region of phase space is constant in time, when the boundary of that volume and all the points inside it are allowed to move down their respective orbits by the same amount of elapsed time. In systems of one degree of freedom, in which phase space is the 2-dimensional q-p plane, Liouville's theorem implies conservation of area, as illustrated in Fig. 6.

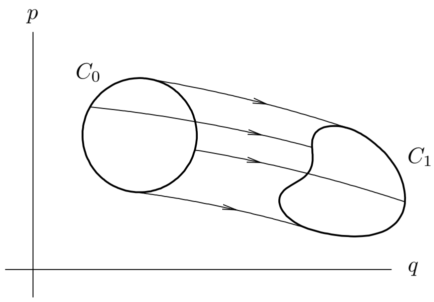

Fig. 6. Liouville's theorem in the phase plane. All points on the boundary of curve C_0 are allow to follow the Hamiltonian flow for the same amount of elapsed time, thereby mapping C_0 into C_1. The area enclosed is constant.

Liouville's theorem is proved by a $2n$-dimensional version of Gauss's law. The essence of the proof is the vanishing of the total divergence of the flow vector in phase space,

$$\sum_{i=1}^{n}\Bigl(\frac{\partial \dot q_i}{\partial q_i} + \frac{\partial \dot p_i}{\partial p_i}\Bigr) = \sum_{i=1}^{n}\Bigl[\frac{\partial}{\partial q_i}\Bigl(\frac{\partial H}{\partial p_i}\Bigr) - \frac{\partial}{\partial p_i}\Bigl(\frac{\partial H}{\partial q_i}\Bigr)\Bigr] = 0. \tag{84}$$

22. Classical Statistical Mechanics

A classical system whose state is only known statistically may be described at a certain time by means of a probability distribution ρ on phase space. In general ρ changes in time, so ρ is a function of q, p and t. The normalization condition is

$$\int d^n q\, d^n p\, \rho(q, p, t) = 1, \tag{85}$$

for all t. Alternatively, we may interpret ρ as giving the density (per unit volume in phase space) of an ensemble of systems, each of which is governed by the same Hamiltonian H. The systems do not interact with one another. In this interpretation, the integral of ρ over all phase space is the number of systems in the ensemble. It differs from the previous interpretation only in the normalization of ρ.

The evolution equation for ρ may be obtained as follows. Consider an infinitesimal volume of phase space, with a certain number of systems of the ensemble in it. Let the volume move with the Hamiltonian flow, as in Fig. 6. The volume of the volume element does not change in this process, according to Liouville's theorem. Nor does the number of systems in the volume element, since the individual systems just evolve according to the (common) Hamiltonian. Therefore the number of systems per unit volume, which is ρ, is constant as we move along orbits. That is,

$$\frac{\partial \rho}{\partial t} + \{\rho, H\} = 0. \tag{86}$$

This is the *Liouville equation.*

Let H be time-independent. It follows from the Liouville equation that the probability density can be constant in time (that is, $\partial\rho/\partial t = 0$) only if $\{\rho, H\} = 0$. The only way this can happen is if ρ is a function of the time-independent constants of the motion. That is, let $\rho = \rho(C_1, \ldots, C_K)$, where the C's are time-independent constants of the motion. Then by the chain rule property (80) of the Poisson bracket we have

$$\{\rho, H\} = \sum_k \frac{\partial \rho}{\partial C_k} \{C_k, H\} = 0. \tag{87}$$

To analyze these statements carefully requires ergodic theory, which we will not go into.

In some cases the only time-independent constant of the motion is the Hamiltonian itself. Then ρ is constant in time only if ρ is a function of the Hamiltonian. Two cases are of interest. One is

$$\rho = A\,\delta(H - E), \tag{88}$$

where A is a normalization constant and E is an energy. The ensemble consists of systems of a known energy E. This is called the *microcanonical* ensemble. It is appropriate for isolated systems, in which the energy is constant.

Another important case is

$$\rho = \frac{1}{Z} e^{-\beta H}, \tag{89}$$

where $\beta = 1/kT$ is the usual thermodynamic parameter (k is the Boltzmann constant, and T the temperature). This is the *canonical* ensemble, appropriate for a system in contact with a heat bath at temperature T. Here $1/Z$ is the normalization of ρ, so that

$$Z(\beta) = \int d^n q\, d^n p\, e^{-\beta H}. \tag{90}$$

The quantity $Z(\beta)$ is the (classical) *partition function.*

23. Hamilton's Principal Function

Let us return to the action functional (11), which is defined on paths belonging to the path space defined by Eq. (13). This path space is parameterized by the endpoints q_0 and q_1 and the endtimes t_0 and t_1. As noted, there may be more than one physical path in this path space, that is, more than one path on which the action is stationary. Let us denote these physical paths by $q_b(t)$, where b is a "branch" index labeling the paths. Generally the branch index is discrete, so we can take $b = 1, 2, \ldots$.

The physical paths cause the action to be stationary. Does the action play any other role in mechanics? In particular, is there any significance to the *value* of the action functional, evaluated on a physical path? This value depends only on the branch index, and on the parameters of the path space itself. We will write $S_b(q_0, t_0; q_1, t_1)$ for this function, so that

$$S_b(q_0, t_0; q_1, t_1) = A[q_b(t)] = \int_{t_0}^{t_1} dt\, L\big(q_b(t), \dot{q}_b(t), t\big). \tag{91}$$

This function is called *Hamilton's principal function.* It is also called "the action," but you must not confuse it with the action functional $A[q(t)]$ or the other quantities called "the action" in classical mechanics. (They all have dimensions of action, but otherwise are different.)

Hamilton found that his principal function satisfies some interesting differential equations, which imply that knowledge of this function is equivalent to knowing the general solution of the classical equations of motion. This is remarkable, because Hamilton's principal function is a single function, while the general solution of the equations of motion involves $2n$ functions, giving the final q, p coordinates as functions of the initial q, p coordinates and the time.

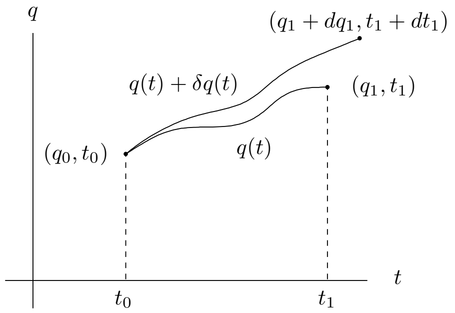

Fig. 7. A path $q(t)$ satisfying fixed boundary conditions $q(t_0) = q_0$ and $q(t_1) = q_1$. Also shown is a modified path $q(t) + \delta q(t)$, satisfying the same endpoint and endtime conditions.

Figure 7 illustrates a physical orbit $q(t)$ in the q-t plane and a nearby physical orbit $q(t) + \delta q(t)$, whose final position q_1 and t_1 have been shifted slightly by amounts dq_1 and dt_1. That is, the modified orbit $q(t) + \delta q(t)$ satisfies

$$q(t_1 + dt_1) + \delta q(t_1 + dt_1) = q_1 + dq_1. \tag{92}$$

This is at the upper limit. As shown in Fig. 7, the modified orbit coincides with the original orbit at the lower limit, that is,

$$q(t_0) + \delta q(t_0) = q_0, \tag{93}$$

so $\delta q(t_0) = 0$.

Figure 7 differs from Fig. 4 in several respects. First, the path $q(t)$ in Fig. 4 is any path satisfying the endpoint and endtime conditions (13), generally a nonphysical path, while $q(t)$ in Fig. 7 is a physical path. That is, $q(t)$ in Fig. 7 is one of the paths $q_b(t)$ mentioned above, but we have suppressed the b index. Next, $\delta q(t)$ in Fig. 4 is a variation about the generally nonphysical path $q(t)$, producing another, generally nonphysical path, $q(t)+\delta q(t)$ in the same path space specified by Eq. (13). In Fig. 7, however, $\delta q(t)$ is the difference between two physical paths, $q(t)$ and $q(t)+\delta q(t)$. Also, the modified path $q(t)+\delta q(t)$ belongs to a different path space than $q(t)$, that is, it satisfies the modified endpoint and endtime conditions (92) at the upper limit.

With these understandings, we can compute the difference in Hamilton's principal function between the two paths of Fig. 7. We suppress the branch index on S as well as on $q(t)$. We have

$$\begin{aligned} dS &= S(q_0, t_0; q_1 + dq_1, t_1 + dt_1) - S(q_0, t_0; q_1, t_1) \\ &= \int_{t_0}^{t_1+dt_1} L\big(q(t) + \delta q(t), \dot{q}(t) + \delta\dot{q}(t), t\big)\, dt - \int_{t_0}^{t_1} L\big(q(t), \dot{q}(t), t\big)\, dt \\ &= L(q_1, \dot{q}_1, t_1)\, dt_1 + \int_{t_0}^{t_1} dt \sum_{i=1}^{n} \Big[\frac{\partial L}{\partial q_i}\delta q_i(t) + \frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i(t)\Big], \end{aligned} \tag{94}$$

where $\dot{q}_1 = \dot{q}(t_1)$ and where we have expanded the right hand side out to first order in small quantities. We integrate the last term by parts as we did in Eq. (16), to obtain

$$\int_{t_0}^{t_1} dt \sum_{i=1}^{n} \Big[\frac{\partial L}{\partial q_i}\delta q_i(t) + \frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i(t)\Big] = \sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}_i}\delta q_i(t)\Big|_{t_0}^{t_1} - \int_{t_0}^{t_1} dt \sum_{i=1}^{n} \Big[\frac{\partial L}{\partial q_i} - \frac{d}{dt}\Big(\frac{\partial L}{\partial \dot{q}_i}\Big)\Big]\delta q(t). \tag{95}$$

This time, however, δq does not vanish at t_1, while the integral on the right does vanish because the physical path $q(t)$ satisfies the Euler-Lagrange equations. The result is simply

$$\sum_{i=1}^{n} \frac{\partial L}{\partial \dot{q}_i}\delta q_i(t_1) = \sum_{i=1}^{n} p_{1i}\, \delta q_i(t_1), \tag{96}$$

where p_1 means the momentum of the path $q(t)$, evaluated at t_1. Putting this back into Eq. (92), we have

$$dS = L_1\, dt_1 + \sum_{i=1}^{n} p_{1i}\, \delta q_i(t_1). \tag{97}$$

Now expanding Eq. (92) out to first order in small quantities, we have

$$\dot{q}(t_1)\, dt_1 + \delta q(t_1) = dq_1, \tag{98}$$

which can be solved for $\delta q(t_1)$. Substituting this into Eq. (97) gives

$$dS = \sum_{i=1}^{n} p_{1i}\, dq_{1i} - H_1\, dt_1, \tag{99}$$

where

$$H_1 = \sum_{i=1}^{n} p_{1i}\, \dot{q}_{1i} - L_1 \tag{100}$$

is the Hamiltonian, evaluated at the t_1 endpoint of the path $q(t)$.

Equation (99) implies

$$\frac{\partial S}{\partial q_{1i}} = p_{1i}, \qquad \frac{\partial S}{\partial t_1} = -H_1. \tag{101}$$

In Fig. 7 we have varied the upper endpoint and endtime of the physical path $q(t)$, while keeping the lower endpoint and endtime fixed. If we also vary the lower endpoint and endtime, we find

$$\frac{\partial S}{\partial q_{0i}} = -p_{0i}, \qquad \frac{\partial S}{\partial t_0} = H_0. \tag{102}$$

24. The Hamilton-Jacobi Equation

Hamilton realized that Eqs. (101) and (102) allow the general solution of the equations of motion to be obtained once his principal function S is known. Since S is a function of the parameters of the path space, $(q_0, t_0; q_1, t_1)$, the q-derivative in Eq. (102) gives the initial momentum as a function of these parameters,

$$p_0 = p_0(q_0, t_0; q_1, t_1), \tag{103}$$

where we just write p_0 for the vector of initial momenta p_{0i}, $i = 1, \ldots, n$. This equation can be solved for q_1, giving us functions,

$$q_1 = Q(q_0, p_0, t_0, t_1), \tag{104}$$

where q_1 is the value of the function Q. The function Q gives the final position q_1 as a function of the initial position and momentum (q_0, p_0) and the initial and final times t_0 and t_1.

Similarly, the q-derivative of Eq. (101) gives the final momentum p_1 as a functions of the parameters of the path space,

$$p_1 = p_1(q_0, t_0; q_1, t_1). \tag{105}$$

But by substituting Eq. (104) into this, we can eliminate q_1 and express p_1 as a function of (q_0, p_0, t_0, t_1):

$$p_1 = P(q_0, p_0, t_0, t_1), \tag{106}$$

in which p_1 is the value of the function P. The function P gives the final momentum p_1 as a function of the initial position and momentum (q_0, p_0) and the initial and final times t_0 and t_1.

Taken together, the functions Q and P constitute the general solution of the equations of motion, that is, these functions specify the final dynamical state (q_1, p_1) given the initial dynamical state (q_0, p_0) and the two times. Given Hamilton's principal function S, the derivation only involves differentiations and algebraic operations (inverting systems of equations). There are no differential equations to be solved.

Thus we would like to find a way of determining Hamilton's principal function S. This function is defined in Eq. (91) as the integral of the Lagrangian along the orbit, but if the orbits are not known we cannot do the integral. Therefore we seek an equation satisfied by S.

The desired equation comes from the time derivative in Eq. (101). In that equation, H_1 means the Hamiltonian evaluated at the final point on the orbit, that is, it is $H(q_1, p_1, t_1)$. At that point, however, p_1 is given by the q-derivative in Eq. (101). Putting the pieces together, we have

$$H\Big(q_1, \frac{\partial S}{\partial q_1}, t_1\Big) + \frac{\partial S}{\partial t_1} = 0. \tag{107}$$

This is the *Hamilton-Jacobi equation* for the function S. It determines the q_1 and t_1 dependence of S. To determine the q_0 and t_0 dependence, we treat Eq. (102) in a similar manner. This gives

$$H\Big(q_0, -\frac{\partial S}{\partial q_0}, t_0\Big) - \frac{\partial S}{\partial t_0} = 0, \tag{108}$$

in effect a second Hamilton-Jacobi equation also satisfied by S. The solution of the classical equations of motion (Hamilton's equations), a system of coupled, ordinary differential equations, is equivalent to the solution of the pair of partial differential equations, Eqs. (107) and (108).

Solving Eqs. (107) and (108) for Hamilton's principal function is not usually a practical way of solving the equations of motion of classical mechanics. There is a modification of this procedure, however, due to Jacobi, that is quite useful in certain applications. For the case of time-independent Hamiltonians, Jacobi's theory leads to another version of the Hamilton-Jacobi equation,

$$H\Big(q, \frac{\partial W}{\partial q}\Big) = E, \tag{109}$$

where E is an energy. This can be regarded as a time-independent version of the Hamilton-Jacobi equation. Here W is a function of q and a collection of parameters which turn out to be constants of motion. We will not pursue this topic here, which is covered in advanced books on classical mechanics. But Hamilton's principal function S is important in semiclassical treatments of quantum mechanics (see Notes 7 and 9), where it appears as the phase of the propagator. If one must evaluate it, the easiest way is usually to use the definition (91). The function W also appears in semiclassical treatments of quantum mechanics, where it is essentially the phase of an energy eigenfunction.

25. Canonical Transformations

Coordinate transformations on phase space are used for many purposes. Classical perturbation theory, for example, is based on the idea of applying successive coordinate transformations that simplify the equations of motion to progressively higher powers of the perturbation parameter.

When a coordinate transformation is carried out, a question is whether the new equations of motion can be written in the standard Hamiltonian form (62).

Let (q, p) be the old coordinates on phase space and let Q and P stand for $2n$ invertible functions of (q, p, t) that represent a new coordinate system on phase space. That is, let

$$\begin{aligned} Q &= Q(q, p, t), \\ P &= P(q, p, t), \end{aligned} \tag{110}$$

where, as indicated, the coordinate transformation is allowed to depend on time. Suppose now that for all old Hamiltonians $H(q, p, t)$ there exists a "new Hamiltonian," call it $K(Q, P, t)$, such that the new equations of motion are given by

$$\dot{Q}_i = \frac{\partial K}{\partial P_i}, \qquad \dot{P}_i = -\frac{\partial K}{\partial Q_i}. \tag{111}$$

Then it can be shown that the new coordinates Q and P satisfy the Poisson bracket relations,

$$\begin{aligned} \{Q_i, Q_j\} &= \sum_k \Big(\frac{\partial Q_i}{\partial q_k} \frac{\partial Q_j}{\partial p_k} - \frac{\partial Q_i}{\partial p_k} \frac{\partial Q_j}{\partial q_k} \Big) = 0, \\ \{Q_i, P_j\} &= \sum_k \Big(\frac{\partial Q_i}{\partial q_k} \frac{\partial P_j}{\partial p_k} - \frac{\partial Q_i}{\partial p_k} \frac{\partial P_j}{\partial q_k} \Big) = \delta_{ij}, \\ \{P_i, P_j\} &= \sum_k \Big(\frac{\partial P_i}{\partial q_k} \frac{\partial P_j}{\partial p_k} - \frac{\partial P_i}{\partial p_k} \frac{\partial P_j}{\partial q_k} \Big) = 0. \end{aligned} \tag{112}$$

In other words, the new coordinates satisfy the same Poisson bracket relations among themselves as the old coordinates. See Eq. (82).

Conversely, suppose the coordinate transformation satisfies Eqs. (112). Then it can be shown that a new Hamiltonian $K(Q, P, t)$ exists such that the new equations of motion are Eqs. (111).

A coordinate transformation that satisfies Eqs. (112) is called a *canonical transformation.* Canonical transformations are conveniently specified in terms of scalar "generating functions," of which Hamilton's principal function S and the function W of Eq. (109) are examples. There is also the problem of determining the new Hamiltonian K, given a canonical transformation and an old Hamiltonian H. If the transformation is time-independent, then the new Hamiltonian is the same as the old one, just transformed to the new coordinates. That is, $K(Q, P) = H(q, p)$ in this case. In the general, time-dependent, case, K is not the same as H, rather there is an extra term involving the generating function.