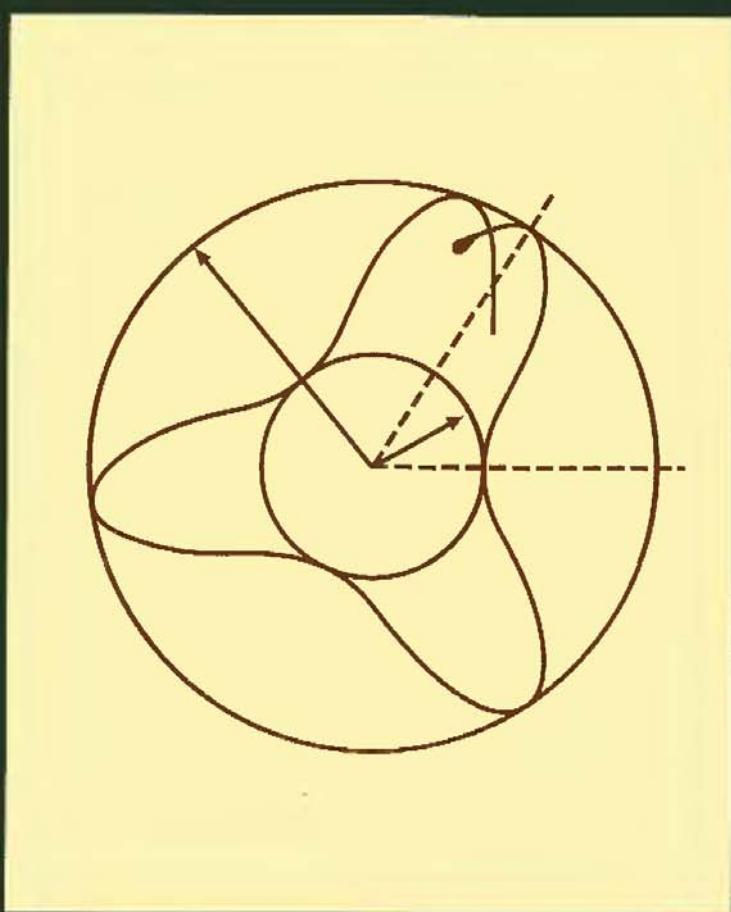


Classical
MECHANICS

THIRD EDITION



Goldstein
Poole & Safko

CLASSICAL MECHANICS

THIRD EDITION

Herbert Goldstein

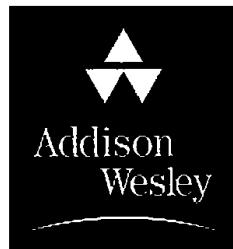
Columbia University

Charles Poole

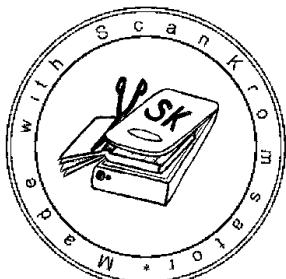
University of South Carolina

John Safko

University of South Carolina



San Francisco Boston New York
Capetown Hong Kong London Madrid Mexico City
Montreal Munich Paris Singapore Sydney Tokyo Toronto



Contents

1 ■ Survey of the Elementary Principles	1
1.1 Mechanics of a Particle	1
1.2 Mechanics of a System of Particles	5
1.3 Constraints	12
1.4 D'Alembert's Principle and Lagrange's Equations	16
1.5 Velocity-Dependent Potentials and the Dissipation Function	22
1.6 Simple Applications of the Lagrangian Formulation	24
2 ■ Variational Principles and Lagrange's Equations	34
2.1 Hamilton's Principle	34
2.2 Some Techniques of the Calculus of Variations	36
2.3 Derivation of Lagrange's Equations from Hamilton's Principle	44
2.4 Extension of Hamilton's Principle to Nonholonomic Systems	45
2.5 Advantages of a Variational Principle Formulation	51
2.6 Conservation Theorems and Symmetry Properties	54
2.7 Energy Function and the Conservation of Energy	60
3 ■ The Central Force Problem	70
3.1 Reduction to the Equivalent One-Body Problem	70
3.2 The Equations of Motion and First Integrals	72
3.3 The Equivalent One-Dimensional Problem, and Classification of Orbits	76
3.4 The Virial Theorem	83
3.5 The Differential Equation for the Orbit, and Integrable Power-Law Potentials	86
3.6 Conditions for Closed Orbits (Bertrand's Theorem)	89
3.7 The Kepler Problem: Inverse-Square Law of Force	92
3.8 The Motion in Time in the Kepler Problem	98
3.9 The Laplace–Runge–Lenz Vector	102
3.10 Scattering in a Central Force Field	106
3.11 Transformation of the Scattering Problem to Laboratory Coordinates	114
3.12 The Three-Body Problem	121

4 ■ The Kinematics of Rigid Body Motion	134
4.1 The Independent Coordinates of a Rigid Body	134
4.2 Orthogonal Transformations	139
4.3 Formal Properties of the Transformation Matrix	144
4.4 The Euler Angles	150
4.5 The Cayley–Klein Parameters and Related Quantities	154
4.6 Euler's Theorem on the Motion of a Rigid Body	155
4.7 Finite Rotations	161
4.8 Infinitesimal Rotations	163
4.9 Rate of Change of a Vector	171
4.10 The Coriolis Effect	174
5 ■ The Rigid Body Equations of Motion	184
5.1 Angular Momentum and Kinetic Energy of Motion about a Point	184
5.2 Tensors	188
5.3 The Inertia Tensor and the Moment of Inertia	191
5.4 The Eigenvalues of the Inertia Tensor and the Principal Axis Transformation	195
5.5 Solving Rigid Body Problems and the Euler Equations of Motion	198
5.6 Torque-free Motion of a Rigid Body	200
5.7 The Heavy Symmetrical Top with One Point Fixed	208
5.8 Precession of the Equinoxes and of Satellite Orbits	223
5.9 Precession of Systems of Charges in a Magnetic Field	230
6 ■ Oscillations	238
6.1 Formulation of the Problem	238
6.2 The Eigenvalue Equation and the Principal Axis Transformation	241
6.3 Frequencies of Free Vibration, and Normal Coordinates	250
6.4 Free Vibrations of a Linear Triatomic Molecule	253
6.5 Forced Vibrations and the Effect of Dissipative Forces	259
6.6 Beyond Small Oscillations: The Damped Driven Pendulum and the Josephson Junction	265
7 ■ The Classical Mechanics of the Special Theory of Relativity	276
7.1 Basic Postulates of the Special Theory	277
7.2 Lorentz Transformations	280
7.3 Velocity Addition and Thomas Precession	282
7.4 Vectors and the Metric Tensor	286

7.5	1-Forms and Tensors	289
7.6	Forces in the Special Theory; Electromagnetism	297
7.7	Relativistic Kinematics of Collisions and Many-Particle Systems	300
7.8	Relativistic Angular Momentum	309
7.9	The Lagrangian Formulation of Relativistic Mechanics	312
7.10	Covariant Lagrangian Formulations	318
7.11	Introduction to the General Theory of Relativity	324
8	■ The Hamilton Equations of Motion	334
8.1	Legendre Transformations and the Hamilton Equations of Motion	334
8.2	Cyclic Coordinates and Conservation Theorems	343
8.3	Routh's Procedure	347
8.4	The Hamiltonian Formulation of Relativistic Mechanics	349
8.5	Derivation of Hamilton's Equations from a Variational Principle	353
8.6	The Principle of Least Action	356
9	■ Canonical Transformations	368
9.1	The Equations of Canonical Transformation	368
9.2	Examples of Canonical Transformations	375
9.3	The Harmonic Oscillator	377
9.4	The Symplectic Approach to Canonical Transformations	381
9.5	Poisson Brackets and Other Canonical Invariants	388
9.6	Equations of Motion, Infinitesimal Canonical Transformations, and Conservation Theorems in the Poisson Bracket Formulation	396
9.7	The Angular Momentum Poisson Bracket Relations	408
9.8	Symmetry Groups of Mechanical Systems	412
9.9	Liouville's Theorem	419
10	■ Hamilton–Jacobi Theory and Action-Angle Variables	430
10.1	The Hamilton–Jacobi Equation for Hamilton's Principal Function	430
10.2	The Harmonic Oscillator Problem as an Example of the Hamilton–Jacobi Method	434
10.3	The Hamilton–Jacobi Equation for Hamilton's Characteristic Function	440
10.4	Separation of Variables in the Hamilton–Jacobi Equation	444
10.5	Ignorable Coordinates and the Kepler Problem	445
10.6	Action-angle Variables in Systems of One Degree of Freedom	452



Contents

10.7 Action-Angle Variables for Completely Separable Systems	457
10.8 The Kepler Problem in Action-angle Variables	466
11 ■ Classical Chaos	483
11.1 Periodic Motion	484
11.2 Perturbations and the Kolmogorov–Arnold–Moser Theorem	487
11.3 Attractors	489
11.4 Chaotic Trajectories and Liapunov Exponents	491
11.5 Poincaré Maps	494
11.6 Hénon–Heiles Hamiltonian	496
11.7 Bifurcations, Driven-damped Harmonic Oscillator, and Parametric Resonance	505
11.8 The Logistic Equation	509
11.9 Fractals and Dimensionality	516
12 ■ Canonical Perturbation Theory	526
12.1 Introduction	526
12.2 Time-dependent Perturbation Theory	527
12.3 Illustrations of Time-dependent Perturbation Theory	533
12.4 Time-independent Perturbation Theory	541
12.5 Adiabatic Invariants	549
13 ■ Introduction to the Lagrangian and Hamiltonian Formulations for Continuous Systems and Fields	558
13.1 The Transition from a Discrete to a Continuous System	558
13.2 The Lagrangian Formulation for Continuous Systems	561
13.3 The Stress-energy Tensor and Conservation Theorems	566
13.4 Hamiltonian Formulation	572
13.5 Relativistic Field Theory	577
13.6 Examples of Relativistic Field Theories	583
13.7 Noether's Theorem	589
Appendix A ■ Euler Angles in Alternate Conventions and Cayley–Klein Parameters	601
Appendix B ■ Groups and Algebras	605
Selected Bibliography	617
Author Index	623
Subject Index	625

Preface to the Third Edition

The first edition of this text appeared in 1950, and it was so well received that it went through a second printing the very next year. Throughout the next three decades it maintained its position as the acknowledged standard text for the introductory Classical Mechanics course in graduate level physics curricula throughout the United States, and in many other countries around the world. Some major institutions also used it for senior level undergraduate Mechanics. Thirty years later, in 1980, a second edition appeared which was “a thorough-going revision of the first edition.” The preface to the second edition contains the following statement: “I have tried to retain, as much as possible, the advantages of the first edition while taking into account the developments of the subject itself, its position in the curriculum, and its applications to other fields.” This is the philosophy which has guided the preparation of this third edition twenty more years later.

The second edition introduced one additional chapter on Perturbation Theory, and changed the ordering of the chapter on Small Oscillations. In addition it added a significant amount of new material which increased the number of pages by about 68%. This third edition adds still one more new chapter on Nonlinear Dynamics or Chaos, but counterbalances this by reducing the amount of material in several of the other chapters, by shortening the space allocated to appendices, by considerably reducing the bibliography, and by omitting the long lists of symbols. Thus the third edition is comparable in size to the second.

In the chapter on relativity we have abandoned the complex Minkowski space in favor of the now standard real metric. Two of the authors prefer the complex metric because of its pedagogical advantages (HG) and because it fits in well with Clifford Algebra formulations of Physics (CPP), but the desire to prepare students who can easily move forward into other areas of theory such as field theory and general relativity dominated over personal preferences. Some modern notation such as 1-forms, mapping and the wedge product is introduced in this chapter.

The chapter on Chaos is a necessary addition because of the current interest in nonlinear dynamics which has begun to play a significant role in applications of classical dynamics. The majority of classical mechanics problems and applications in the real world include nonlinearities, and it is important for the student to have a grasp of the complexities involved, and of the new properties that can emerge. It is also important to realize the role of fractal dimensionality in chaos.

New sections have been added and others combined or eliminated here and there throughout the book, with the omissions to a great extent motivated by the desire not to extend the overall length beyond that of the second edition. A section

Preface to the Third Edition

was added on the Euler and Lagrange exact solutions to the three body problem. In several places phase space plots and Lissajous figures were appended to illustrate solutions. The damped driven pendulum was discussed as an example that explains the workings of Josephson junctions. The symplectic approach was clarified by writing out some of the matrices. The harmonic oscillator was treated with anisotropy, and also in polar coordinates. The last chapter on continua and fields was formulated in the modern notation introduced in the relativity chapter. The significances of the special unitary group in two dimensions $SU(2)$ and the special orthogonal group in three dimensions $SO(3)$ were presented in more up-to-date notation, and an appendix was added on groups and algebras. Special tables were introduced to clarify properties of ellipses, vectors, vector fields and 1-forms, canonical transformations, and the relationships between the spacetime and symplectic approaches.

Several of the new features and approaches in this third edition had been mentioned as possibilities in the preface to the second edition, such as properties of group theory, tensors in non-Euclidean spaces, and “new mathematics” of theoretical physics such as manifolds. The reference to “One area omitted that deserves special attention—nonlinear oscillation and associated stability questions” now constitutes the subject matter of our new Chapter 11 “Classical Chaos.” We debated whether to place this new chapter after Perturbation theory where it fits more logically, or before Perturbation theory where it is more likely to be covered in class, and we chose the latter. The referees who reviewed our manuscript were evenly divided on this question.

The mathematical level of the present edition is about the same as that of the first two editions. Some of the mathematical physics, such as the discussions of hermitean and unitary matrices, was omitted because it pertains much more to quantum mechanics than it does to classical mechanics, and little used notations like dyadics were curtailed. Space devoted to power law potentials, Cayley-Klein parameters, Routh’s procedure, time independent perturbation theory, and the stress-energy tensor was reduced. In some cases reference was made to the second edition for more details. The problems at the end of the chapters were divided into “derivations” and “exercises,” and some new ones were added.

The authors are especially indebted to Michael A. Unseren and Forrest M. Hoffman of the Oak Ridge National laboratory for their 1993 compilation of errata in the second edition that they made available on the Internet. It is hoped that not too many new errors have slipped into this present revision. We wish to thank the students who used this text in courses with us, and made a number of useful suggestions that were incorporated into the manuscript. Professors Thomas Sayetta and the late Mike Schuette made helpful comments on the Chaos chapter, and Professors Joseph Johnson and James Knight helped to clarify our ideas on Lie Algebras. The following professors reviewed the manuscript and made many helpful suggestions for improvements: Yoram Alhassid, Yale University; Dave Ellis, University of Toledo; John Gruber, San Jose State; Thomas Handler, University of Tennessee; Daniel Hong, Lehigh University; Kara Keeter, Idaho State University; Carolyn Lee; Yannick Meurice, University of Iowa; Daniel

Marlow, Princeton University; Julian Noble, University of Virginia; Muhammad Numan, Indiana University of Pennsylvania; Steve Ruden, University of California, Irvine; Jack Semura, Portland State University; Tammy Ann Smecker-Hane, University of California, Irvine; Daniel Stump, Michigan State University; Robert Wald, University of Chicago; Doug Wells, Idaho State University.

It has indeed been an honor for two of us (CPP and JLS) to collaborate as co-authors of this third edition of such a classic book fifty years after its first appearance. We have admired this text since we first studied Classical Mechanics from the first edition in our graduate student days (CPP in 1953 and JLS in 1960), and each of us used the first and second editions in our teaching throughout the years. Professor Goldstein is to be commended for having written and later enhanced such an outstanding contribution to the classic Physics literature.

Above all we register our appreciation and acknowledgement in the words of Psalm 19,1:

Oἱ οὐρανοὶ διηγοῦνται δοξαν Θεοῦ

Flushing, New York
Columbia, South Carolina
Columbia, South Carolina
July, 2000

HERBERT GOLDSTEIN
CHARLES P. POOLE, JR.
JOHN L. SAFKO

CHAPTER

1

Survey of the Elementary Principles

The motion of material bodies formed the subject of some of the earliest research pursued by the pioneers of physics. From their efforts there has evolved a vast field known as analytical mechanics or dynamics, or simply, mechanics. In the present century the term "classical mechanics" has come into wide use to denote this branch of physics in contradistinction to the newer physical theories, especially quantum mechanics. We shall follow this usage, interpreting the name to include the type of mechanics arising out of the special theory of relativity. It is the purpose of this book to develop the structure of classical mechanics and to outline some of its applications of present-day interest in pure physics. Basic to any presentation of mechanics are a number of fundamental physical concepts, such as space, time, simultaneity, mass, and force. For the most part, however, these concepts will not be analyzed critically here; rather, they will be assumed as undefined terms whose meanings are familiar to the reader.

1.1 ■ MECHANICS OF A PARTICLE

Let \mathbf{r} be the radius vector of a particle from some given origin and \mathbf{v} its vector velocity:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt}. \quad (1.1)$$

The *linear momentum* \mathbf{p} of the particle is defined as the product of the particle mass and its velocity:

$$\mathbf{p} = m\mathbf{v}. \quad (1.2)$$

In consequence of interactions with external objects and fields, the particle may experience forces of various types, e.g., gravitational or electrodynamic; the vector sum of these forces exerted on the particle is the total force \mathbf{F} . The mechanics of the particle is contained in *Newton's second law of motion*, which states that there exist frames of reference in which the motion of the particle is described by the differential equation

$$\mathbf{F} = \frac{d\mathbf{p}}{dt} \equiv \dot{\mathbf{p}}, \quad (1.3)$$

Chapter 1 Survey of the Elementary Principles

or

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}). \quad (1.4)$$

In most instances, the mass of the particle is constant and Eq. (1.4) reduces to

$$\mathbf{F} = m \frac{d\mathbf{v}}{dt} = m\mathbf{a}. \quad (1.5)$$

where \mathbf{a} is the vector acceleration of the particle defined by

$$\mathbf{a} = \frac{d^2\mathbf{r}}{dt^2}. \quad (1.6)$$

The equation of motion is thus a differential equation of second order, assuming \mathbf{F} does not depend on higher-order derivatives.

A reference frame in which Eq. (1.3) is valid is called an *inertial* or *Galilean system*. Even within classical mechanics the notion of an inertial system is something of an idealization. In practice, however, it is usually feasible to set up a coordinate system that comes as close to the desired properties as may be required. For many purposes, a reference frame fixed in Earth (the “laboratory system”) is a sufficient approximation to an inertial system, while for some astronomical purposes it may be necessary to construct an inertial system by reference to distant galaxies.

Many of the important conclusions of mechanics can be expressed in the form of conservation theorems, which indicate under what conditions various mechanical quantities are constant in time. Equation (1.3) directly furnishes the first of these, the

Conservation Theorem for the Linear Momentum of a Particle: If the total force, \mathbf{F} , is zero, then $\dot{\mathbf{p}} = 0$ and the linear momentum, \mathbf{p} , is conserved.

The angular momentum of the particle about point O , denoted by \mathbf{L} , is defined as

$$\mathbf{L} = \mathbf{r} \times \mathbf{p}, \quad (1.7)$$

where \mathbf{r} is the radius vector from O to the particle. Notice that the order of the factors is important. We now define the *moment of force* or *torque* about O as

$$\mathbf{N} = \mathbf{r} \times \mathbf{F}. \quad (1.8)$$

The equation analogous to (1.3) for \mathbf{N} is obtained by forming the cross product of \mathbf{r} with Eq. (1.4):

$$\mathbf{r} \times \mathbf{F} = \mathbf{N} = \mathbf{r} \times \frac{d}{dt}(m\mathbf{v}). \quad (1.9)$$

Equation (1.9) can be written in a different form by using the vector identity:

$$\frac{d}{dt}(\mathbf{r} \times m\mathbf{v}) = \mathbf{v} \times m\mathbf{v} + \mathbf{r} \times \frac{d}{dt}(m\mathbf{v}), \quad (1.10)$$

where the first term on the right obviously vanishes. In consequence of this identity, Eq. (1.9) takes the form

$$\mathbf{N} = \frac{d}{dt}(\mathbf{r} \times m\mathbf{v}) = \frac{d\mathbf{L}}{dt} \equiv \dot{\mathbf{L}}. \quad (1.11)$$

Note that both \mathbf{N} and \mathbf{L} depend on the point O , about which the moments are taken.

As was the case for Eq. (1.3), the torque equation, (1.11), also yields an immediate conservation theorem, this time the

Conservation Theorem for the Angular Momentum of a Particle: If the total torque, \mathbf{N} , is zero then $\dot{\mathbf{L}} = 0$, and the angular momentum \mathbf{L} is conserved.

Next consider the work done by the external force \mathbf{F} upon the particle in going from point 1 to point 2. By definition, this work is

$$W_{12} = \int_1^2 \mathbf{F} \cdot d\mathbf{s}. \quad (1.12)$$

For constant mass (as will be assumed from now on unless otherwise specified), the integral in Eq. (1.12) reduces to

$$\int \mathbf{F} \cdot d\mathbf{s} = m \int \frac{d\mathbf{v}}{dt} \cdot \mathbf{v} dt = \frac{m}{2} \int \frac{d}{dt}(v^2) dt,$$

and therefore

$$W_{12} = \frac{m}{2}(v_2^2 - v_1^2). \quad (1.13)$$

The scalar quantity $mv^2/2$ is called the kinetic energy of the particle and is denoted by T , so that the work done is equal to the change in the kinetic energy:

$$W_{12} = T_2 - T_1. \quad (1.14)$$

If the force field is such that the work W_{12} is the same for any physically possible path between points 1 and 2, then the force (and the system) is said to be *conservative*. An alternative description of a conservative system is obtained by imagining the particle being taken from point 1 to point 2 by one possible path and then being returned to point 1 by another path. The independence of W_{12} on the particular path implies that the work done around such a closed circuit is zero, i.e.

$$\oint \mathbf{F} \cdot d\mathbf{s} = 0. \quad (1.15)$$

Chapter 1 Survey of the Elementary Principles

Physically it is clear that a system cannot be conservative if friction or other dissipation forces are present, because $\mathbf{F} \cdot d\mathbf{s}$ due to friction is always positive and the integral cannot vanish.

By a well-known theorem of vector analysis, a necessary and sufficient condition that the work, W_{12} , be independent of the physical path taken by the particle is that \mathbf{F} be the gradient of some scalar function of position:

$$\mathbf{F} = -\nabla V(\mathbf{r}), \quad (1.16)$$

where V is called the *potential*, or *potential energy*. The existence of V can be inferred intuitively by a simple argument. If W_{12} is independent of the path of integration between the end points 1 and 2, it should be possible to express W_{12} as the change in a quantity that depends only upon the positions of the end points. This quantity may be designated by $-V$, so that for a differential path length we have the relation

$$\mathbf{F} \cdot d\mathbf{s} = -dV$$

or

$$F_s = -\frac{\partial V}{\partial s},$$

which is equivalent to Eq. (1.16). Note that in Eq. (1.16) we can add to V any quantity constant in space, without affecting the results. Hence *the zero level of V is arbitrary*.

For a conservative system, the work done by the forces is

$$W_{12} = V_1 - V_2. \quad (1.17)$$

Combining Eq. (1.17) with Eq. (1.14), we have the result

$$T_1 + V_1 = T_2 + V_2, \quad (1.18)$$

which states in symbols the

Energy Conservation Theorem for a Particle: If the forces acting on a particle are conservative, then the total energy of the particle, $T + V$, is conserved.

The force applied to a particle may in some circumstances be given by the gradient of a scalar function that depends explicitly on both the position of the particle and the time. However, the work done on the particle when it travels a distance ds ,

$$\mathbf{F} \cdot d\mathbf{s} = -\frac{\partial V}{\partial s} ds,$$

is then no longer the total change in $-V$ during the displacement, since V also changes explicitly with time as the particle moves. Hence, the work done as the

particle goes from point 1 to point 2 is no longer the difference in the function V between those points. While a total energy $T + V$ may still be defined, it is not conserved during the course of the particle's motion.

1.2 ■ MECHANICS OF A SYSTEM OF PARTICLES

In generalizing the ideas of the previous section to systems of many particles, we must distinguish between the *external forces* acting on the particles due to sources outside the system, and *internal forces* on, say, some particle i due to all other particles in the system. Thus, the equation of motion (Newton's second law) for the i th particle is written as

$$\sum_j \mathbf{F}_{ji} + \mathbf{F}_i^{(e)} = \dot{\mathbf{p}}_i, \quad (1.19)$$

where $\mathbf{F}_i^{(e)}$ stands for an external force, and \mathbf{F}_{ji} is the internal force on the i th particle due to the j th particle (\mathbf{F}_{ii} , naturally, is zero). We shall assume that the \mathbf{F}_{ij} (like the $\mathbf{F}_i^{(e)}$) obey Newton's third law of motion in its original form: that the forces two particles exert on each other are equal and opposite. This assumption (which does not hold for all types of forces) is sometimes referred to as the *weak law of action and reaction*.

Summed over all particles, Eq. (1.19) takes the form

$$\frac{d^2}{dt^2} \sum_i m_i \mathbf{r}_i = \sum_i \mathbf{F}_i^{(e)} + \sum_{\substack{i,j \\ i \neq j}} \mathbf{F}_{ji}. \quad (1.20)$$

The first sum on the right is simply the total external force $\mathbf{F}^{(e)}$, while the second term vanishes, since the law of action and reaction states that each pair $\mathbf{F}_{ij} + \mathbf{F}_{ji}$ is zero. To reduce the left-hand side, we define a vector \mathbf{R} as the average of the radii vectors of the particles, weighted in proportion to their mass:

$$\mathbf{R} - \frac{\sum m_i \mathbf{r}_i}{\sum m_i} = \frac{\sum m_i \mathbf{r}_i}{M}. \quad (1.21)$$

The vector \mathbf{R} defines a point known as the *center of mass*, or more loosely as the center of gravity, of the system (cf. Fig. 1.1). With this definition, (1.20) reduces to

$$M \frac{d^2 \mathbf{R}}{dt^2} = \sum_i \mathbf{F}_i^{(e)} \equiv \mathbf{F}^{(e)}, \quad (1.22)$$

which states that the center of mass moves as if the total external force were acting on the entire mass of the system concentrated at the center of mass. Purely internal forces, if they obey Newton's third law, therefore have no effect on the

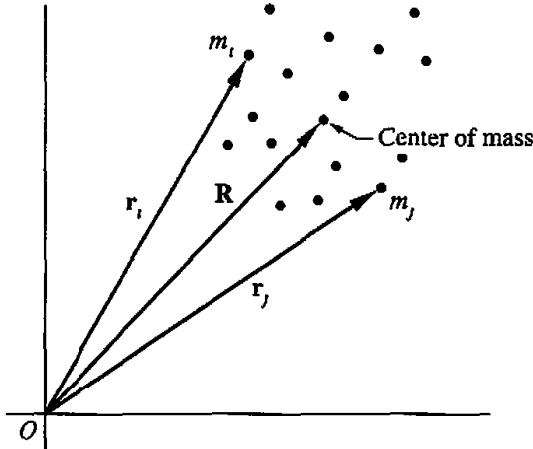


FIGURE 1.1 The center of mass of a system of particles.

motion of the center of mass. An oft-quoted example is the motion of an exploding shell—the center of mass of the fragments traveling as if the shell were still in a single piece (neglecting air resistance). The same principle is involved in jet and rocket propulsion. In order that the motion of the center of mass be unaffected, the ejection of the exhaust gases at high velocity must be counterbalanced by the forward motion of the vehicle at a slower velocity.

By Eq. (1.21) the total linear momentum of the system,

$$\mathbf{P} = \sum_i m_i \frac{d\mathbf{r}_i}{dt} = M \frac{d\mathbf{R}}{dt}, \quad (1.23)$$

is the total mass of the system times the velocity of the center of mass. Consequently, the equation of motion for the center of mass, (1.23), can be restated as the

Conservation Theorem for the Linear Momentum of a System of Particles: *If the total external force is zero, the total linear momentum is conserved.*

We obtain the total angular momentum of the system by forming the cross product $\mathbf{r}_i \times \dot{\mathbf{p}}_i$ and summing over i . If this operation is performed in Eq. (1.19), there results, with the aid of the identity, Eq. (1.10),

$$\sum_i (\mathbf{r}_i \times \dot{\mathbf{p}}_i) = \sum_i \frac{d}{dt} (\mathbf{r}_i \times \mathbf{p}_i) = \dot{\mathbf{L}} = \sum_i \mathbf{r}_i \times \mathbf{F}_i^{(e)} + \sum_{i \neq j} \mathbf{r}_i \times \mathbf{F}_{ji}. \quad (1.24)$$

The last term on the right in (1.24) can be considered a sum of the pairs of the form

$$\mathbf{r}_i \times \mathbf{F}_{ji} + \mathbf{r}_j \times \mathbf{F}_{ij} = (\mathbf{r}_i - \mathbf{r}_j) \times \mathbf{F}_{ji}, \quad (1.25)$$

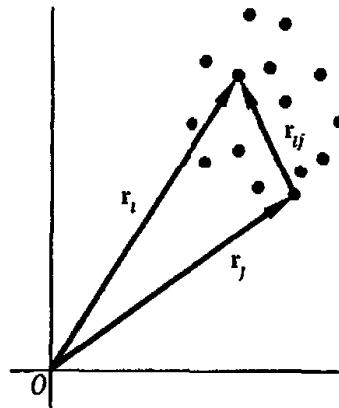


FIGURE 1.2 The vector r_{ij} between the i th and j th particles.

using the equality of action and reaction. But $\mathbf{r}_i - \mathbf{r}_j$ is identical with the vector \mathbf{r}_{ij} from j to i (cf. Fig. 1.2), so that the right-hand side of Eq. (1.25) can be written as

$$\mathbf{r}_{ij} \times \mathbf{F}_{ji}.$$

If the internal forces between two particles, in addition to being equal and opposite, also lie along the line joining the particles—a condition known as the *strong law of action and reaction*—then all of these cross products vanish. The sum over pairs is zero under this assumption and Eq. (1.24) may be written in the form

$$\frac{d\mathbf{L}}{dt} = \mathbf{N}^{(e)}. \quad (1.26)$$

The time derivative of the total angular momentum is thus equal to the moment of the external force about the given point. Corresponding to Eq. (1.26) is the

Conservation Theorem for Total Angular Momentum: \mathbf{L} is constant in time if the applied (external) torque is zero.

(It is perhaps worthwhile to emphasize that this is a *vector* theorem; i.e., L_z will be conserved if $N_z^{(e)}$ is zero, even if $N_x^{(e)}$ and $N_y^{(e)}$ are not zero.)

Note that the conservation of linear momentum in the absence of applied forces assumes that the weak law of action and reaction is valid for the internal forces. The conservation of the total angular momentum of the system in the absence of applied torques requires the validity of the strong law of action and reaction—that the internal forces in addition be *central*. Many of the familiar physical forces, such as that of gravity, satisfy the strong form of the law. But it is possible to find forces for which action and reaction are equal even though the forces are not central (see below). In a system involving moving charges, the forces between the charges predicted by the Biot-Savart law may indeed violate both forms of

Chapter 1 Survey of the Elementary Principles

the action and reaction law.* Equations (1.23) and (1.26), and their corresponding conservation theorems, are not applicable in such cases, at least in the form given here. Usually it is then possible to find some generalization of \mathbf{P} or \mathbf{L} that is conserved. Thus, in an isolated system of moving charges it is the sum of the mechanical angular momentum and the electromagnetic "angular momentum" of the field that is conserved.

Equation (1.23) states that the total linear momentum of the system is the same as if the entire mass were concentrated at the center of mass and moving with it. The analogous theorem for angular momentum is more complicated. With the origin O as reference point, the total angular momentum of the system is

$$\mathbf{L} = \sum_i \mathbf{r}_i \times \mathbf{p}_i.$$

Let \mathbf{R} be the radius vector from O to the center of mass, and let \mathbf{r}'_i be the radius vector from the center of mass to the i th particle. Then we have (cf. Fig. 1.3)

$$\mathbf{r}_i = \mathbf{r}'_i + \mathbf{R} \quad (1.27)$$

and

$$\mathbf{v}_i = \mathbf{v}'_i + \mathbf{v}$$

where

$$\mathbf{v} = \frac{d\mathbf{R}}{dt}$$

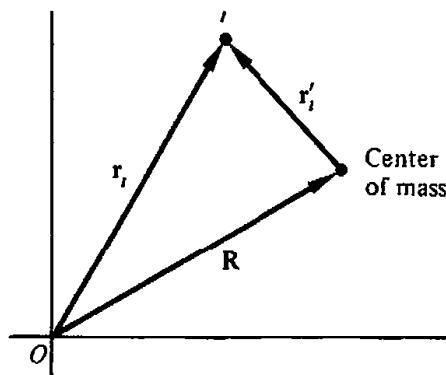


FIGURE 1.3 The vectors involved in the shift of reference point for the angular momentum.

*If two charges are moving uniformly with parallel velocity vectors that are not perpendicular to the line joining the charges, then the net mutual forces are equal and opposite but do not lie along the vector between the charges. Consider, further, two charges moving (instantaneously) so as to "cross the T," i.e., one charge moving directly at the other, which in turn is moving at right angles to the first. Then the second charge exerts a nonvanishing magnetic force on the first, without experiencing any magnetic reaction force at that instant.

is the velocity of the center of mass relative to O , and

$$\mathbf{v}'_i = \frac{d\mathbf{r}'_i}{dt}$$

is the velocity of the i th particle relative to the center of mass of the system. Using Eq. (1.27), the total angular momentum takes on the form

$$\mathbf{L} = \sum_i \mathbf{R} \times m_i \mathbf{v} + \sum_i \mathbf{r}'_i \times m_i \mathbf{v}'_i + \left(\sum_i m_i \mathbf{r}'_i \right) \times \mathbf{v} + \mathbf{R} \times \frac{d}{dt} \sum_i m_i \mathbf{r}'_i.$$

The last two terms in this expression vanish, for both contain the factor $\sum m_i \mathbf{r}'_i$, which, it will be recognized, defines the radius vector of the center of mass in the very coordinate system whose origin is the center of mass and is therefore a null vector. Rewriting the remaining terms, the total angular momentum about O is

$$\mathbf{L} = \mathbf{R} \times M\mathbf{v} + \sum_i \mathbf{r}'_i \times \mathbf{p}'_i. \quad (1.28)$$

In words, Eq. (1.28) says that the total angular momentum about a point O is the angular momentum of motion concentrated at the center of mass, plus the angular momentum of motion about the center of mass. The form of Eq. (1.28) emphasizes that in general \mathbf{L} depends on the origin O , through the vector \mathbf{R} . Only if the center of mass is at rest with respect to O will the angular momentum be independent of the point of reference. In this case, the first term in (1.28) vanishes, and \mathbf{L} always reduces to the angular momentum taken about the center of mass.

Finally, let us consider the energy equation. As in the case of a single particle, we calculate the work done by all forces in moving the system from an initial configuration 1, to a final configuration 2:

$$W_{12} = \sum_i \int_1^2 \mathbf{F}_i \cdot d\mathbf{s}_i = \sum_i \int_1^2 \mathbf{F}_i^{(e)} \cdot d\mathbf{s}_i + \sum_{i \neq j} \int_1^2 \mathbf{F}_{ji} \cdot d\mathbf{s}_i. \quad (1.29)$$

Again, the equations of motion can be used to reduce the integrals to

$$\sum_i \int_1^2 \mathbf{F}_i \cdot d\mathbf{s} = \sum_i \int_1^2 m_i \dot{\mathbf{v}}_i \cdot \mathbf{v}_i dt = \sum_i \int_1^2 d \left(\frac{1}{2} m_i v_i^2 \right).$$

Hence, the work done can still be written as the difference of the final and initial kinetic energies:

$$W_{12} = T_2 - T_1,$$

where T , the total kinetic energy of the system, is

$$T = \frac{1}{2} \sum_i m_i v_i^2. \quad (1.30)$$

Making use of the transformations to center-of-mass coordinates, given in Eq. (1.27), we may also write T as

$$\begin{aligned} T &= \frac{1}{2} \sum_i m_i (\mathbf{v} + \mathbf{v}'_i) \cdot (\mathbf{v} + \mathbf{v}'_i) \\ &= \frac{1}{2} \sum_i m_i v^2 + \frac{1}{2} \sum_i m_i v_i'^2 + \mathbf{v} \cdot \frac{d}{dt} \left(\sum_i m_i \mathbf{r}'_i \right), \end{aligned}$$

and by the reasoning already employed in calculating the angular momentum, the last term vanishes, leaving

$$T = \frac{1}{2} M v^2 + \frac{1}{2} \sum_i m_i v_i'^2 \quad (1.31)$$

The kinetic energy, like the angular momentum, thus also consists of two parts: the kinetic energy obtained if all the mass were concentrated at the center of mass, plus the kinetic energy of motion about the center of mass.

Consider now the right-hand side of Eq. (1.29). In the special case that the external forces are derivable in terms of the gradient of a potential, the first term can be written as

$$\sum_i \int_1^2 \mathbf{F}_i^{(e)} \cdot d\mathbf{s}_i = - \sum_i \int_1^2 \nabla_i V_i \cdot d\mathbf{s}_i = - \sum_i V_i \Big|_1^2,$$

where the subscript i on the del operator indicates that the derivatives are with respect to the components of \mathbf{r}_i . If the internal forces are also conservative, then the mutual forces between the i th and j th particles, \mathbf{F}_{ij} , and \mathbf{F}_{ji} , can be obtained from a potential function V_{ij} . To satisfy the strong law of action and reaction, V_{ij} can be a function only of the distance between the particles:

$$V_{ij} = V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|). \quad (1.32)$$

The two forces are then automatically equal and opposite,

$$\mathbf{F}_{ji} = -\nabla_i V_{ij} = +\nabla_j V_{ij} = -\mathbf{F}_{ij}, \quad (1.33)$$

and lie along the line joining the two particles,

$$\nabla V_{ij}(|\mathbf{r}_i - \mathbf{r}_j|) = (\mathbf{r}_i - \mathbf{r}_j) f, \quad (1.34)$$

where f is some scalar function. If V_{ij} were also a function of the difference of some other pair of vectors associated with the particles, such as their velocities or (to step into the domain of modern physics) their intrinsic “spin” angular momenta, then the forces would still be equal and opposite, but would not necessarily lie along the direction between the particles.

When the forces are all conservative, the second term in Eq. (1.29) can be rewritten as a sum over *pairs* of particles, the terms for each pair being of the form

$$-\int_1^2 (\nabla_i V_{ij} \cdot d\mathbf{s}_i + \nabla_j V_{ij} \cdot d\mathbf{s}_j).$$

If the difference vector $\mathbf{r}_i - \mathbf{r}_j$ is denoted by \mathbf{r}_{ij} , and if ∇_{ij} stands for the gradient with respect to \mathbf{r}_{ij} , then

$$\nabla_i V_{ij} = \nabla_{ij} V_{ij} = -\nabla_j V_{ij},$$

and

$$d\mathbf{s}_i - d\mathbf{s}_j = d\mathbf{r}_i - d\mathbf{r}_j = d\mathbf{r}_{ij},$$

so that the term for the ij pair has the form

$$-\int \nabla_{ij} V_{ij} \cdot d\mathbf{r}_{ij}.$$

The total work arising from internal forces then reduces to

$$-\frac{1}{2} \sum_{i,j} \int_1^2 \nabla_{ij} V_{ij} \cdot d\mathbf{r}_{ij} = -\frac{1}{2} \sum_{i,j} V_{ij} \Big|_1^2. \quad (1.35)$$

The factor $\frac{1}{2}$ appears in Eq. (1.35) because in summing over *both* i and j each member of a given pair is included twice, first in the i summation and then in the j summation.

From these considerations, it is clear that if the external and internal forces are both derivable from potentials it is possible to define a *total potential energy*, V , of the system,

$$V = \sum_i V_i + \frac{1}{2} \sum_{i,j} V_{ij}. \quad (1.36)$$

such that the total energy $T + V$ is conserved, the analog of the conservation theorem (1.18) for a single particle.

The second term on the right in Eq. (1.36) will be called the internal potential energy of the system. In general, it need not be zero and, more important, it may vary as the system changes with time. Only for the particular class of systems known as *rigid bodies* will the internal potential always be constant. Formally, a rigid body can be defined as a system of particles in which the distances r_{ij} are fixed and cannot vary with time. In such case, the vectors $d\mathbf{r}_{ij}$ can only be perpendicular to the corresponding \mathbf{r}_{ij} , and therefore to the \mathbf{F}_{ij} . Therefore, in a rigid body the *internal forces do no work*, and the internal potential must remain

constant. Since the total potential is in any case uncertain to within an additive constant, an unvarying internal potential can be completely disregarded in discussing the motion of the system.

1.3 ■ CONSTRAINTS

From the previous sections one might obtain the impression that all problems in mechanics have been reduced to solving the set of differential equations (1.19):

$$m_i \ddot{\mathbf{r}}_i = \mathbf{F}_i^{(e)} - \sum_j \mathbf{F}_{ji}.$$

One merely substitutes the various forces acting upon the particles of the system, turns the mathematical crank, and grinds out the answers! Even from a purely physical standpoint, however, this view is oversimplified. For example, it may be necessary to take into account the *constraints* that limit the motion of the system. We have already met one type of system involving constraints, namely rigid bodies, where the constraints on the motions of the particles keep the distances r_{ij} unchanged. Other examples of constrained systems can easily be furnished. The beads of an abacus are constrained to one-dimensional motion by the supporting wires. Gas molecules within a container are constrained by the walls of the vessel to move only *inside* the container. A particle placed on the surface of a solid sphere is subject to the constraint that it can move only on the surface or in the region exterior to the sphere.

Constraints may be classified in various ways, and we shall use the following system. If the conditions of constraint can be expressed as equations connecting the coordinates of the particles (and possibly the time) having the form

$$f(\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \dots, t) = 0, \quad (1.37)$$

then the constraints are said to be *holonomic*. Perhaps the simplest example of holonomic constraints is the rigid body, where the constraints are expressed by equations of the form

$$(\mathbf{r}_i - \mathbf{r}_j)^2 - c_{ij}^2 = 0.$$

A particle constrained to move along any curve or on a given surface is another obvious example of a holonomic constraint, with the equations defining the curve or surface acting as the equations of a constraint.

Constraints not expressible in this fashion are called nonholonomic. The walls of a gas container constitute a nonholonomic constraint. The constraint involved in the example of a particle placed on the surface of a sphere is also nonholonomic, for it can be expressed as an inequality

$$r^2 - a^2 \geq 0$$

(where a is the radius of the sphere), which is not in the form of (1.37). Thus, in a gravitational field a particle placed on the top of the sphere will slide down the surface part of the way but will eventually fall off.

Constraints are further classified according to whether the equations of constraint contain the time as an explicit variable (rheonomous) or are not explicitly dependent on time (scleronomous). A bead sliding on a rigid curved wire fixed in space is obviously subject to a scleronomous constraint; if the wire is moving in some prescribed fashion, the constraint is rheonomous. Note that if the wire moves, say, as a reaction to the bead's motion, then the time dependence of the constraint enters in the equation of the constraint only through the coordinates of the curved wire (which are now part of the system coordinates). The overall constraint is then scleronomous.

Constraints introduce two types of difficulties in the solution of mechanical problems. First, the coordinates r_i are no longer all independent, since they are connected by the equations of constraint; hence the equations of motion (1.19) are not all independent. Second, the forces of constraint, e.g., the force that the wire exerts on the bead (or the wall on the gas particle), is not furnished a priori. They are among the unknowns of the problem and must be obtained from the solution we seek. Indeed, imposing constraints on the system is simply another method of stating that there are forces present in the problem that cannot be specified directly but are known rather in terms of their effect on the motion of the system.

In the case of holonomic constraints, the first difficulty is solved by the introduction of *generalized coordinates*. So far we have been thinking implicitly in terms of Cartesian coordinates. A system of N particles, free from constraints, has $3N$ independent coordinates or *degrees of freedom*. If there exist holonomic constraints, expressed in k equations in the form (1.37), then we may use these equations to eliminate k of the $3N$ coordinates, and we are left with $3N - k$ independent coordinates, and the system is said to have $3N - k$ degrees of freedom. This elimination of the dependent coordinates can be expressed in another way, by the introduction of new, $3N - k$, independent variables $q_1, q_2, \dots, q_{3N-k}$ in terms of which the old coordinates $\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N$ are expressed by equations of the form

$$\begin{aligned} \mathbf{r} &= \mathbf{r}_1(q_1, q_2, \dots, q_{3N-k}, t) \\ &\vdots \\ \mathbf{r}_N &= \mathbf{r}_N(q_1, q_2, \dots, q_{3N-k}, t) \end{aligned} \tag{1.38}$$

containing the constraints in them implicitly. These are *transformation* equations from the set of (\mathbf{r}_i) variables to the (q_i) set, or alternatively Eqs. (1.38) can be considered as parametric representations of the (\mathbf{r}_i) variables. It is always assumed that we can also transform back from the (q_i) to the (\mathbf{r}_i) set, i.e., that Eqs. (1.38) combined with the k equations of constraint can be inverted to obtain any q_i as a function of the (\mathbf{r}_i) variable and time.

Usually the generalized coordinates, q_i , unlike the Cartesian coordinates, will not divide into convenient groups of three that can be associated together to form vectors. Thus, in the case of a particle constrained to move *on* the surface of a sphere, the two angles expressing position on the sphere, say latitude and longitude, are obvious possible generalized coordinates. Or, in the example of a double pendulum moving in a plane (two particles connected by an inextensible light rod and suspended by a similar rod fastened to one of the particles), satisfactory generalized coordinates are the two angles θ_1, θ_2 . (Cf. Fig. 1.4.) Generalized coordinates, in the sense of coordinates other than Cartesian, are often useful in systems without constraints. Thus, in the problem of a particle moving in an external central force field ($V = V(r)$), there is no constraint involved, but it is clearly more convenient to use spherical polar coordinates than Cartesian coordinates. Do not, however, think of generalized coordinates in terms of conventional orthogonal position coordinates. All sorts of quantities may be impressed to serve as generalized coordinates. Thus, the amplitudes in a Fourier expansion of \mathbf{r} , may be used as generalized coordinates, or we may find it convenient to employ quantities with the dimensions of energy or angular momentum.

If the constraint is nonholonomic, the equations expressing the constraint cannot be used to eliminate the dependent coordinates. An oft-quoted example of a nonholonomic constraint is that of an object rolling on a rough surface without slipping. The coordinates used to describe the system will generally involve angular coordinates to specify the orientation of the body, plus a set of coordinates describing the location of the point of contact on the surface. The constraint of "rolling" connects these two sets of coordinates; they are not independent. A change in the position of the point of contact inevitably means a change in its orientation. Yet we cannot reduce the number of coordinates, for the "rolling" condition is not expressible as a equation between the coordinates, in the manner of (1.37). Rather, it is a condition on the *velocities* (i.e., the point of contact is stationary), a differential condition that can be given in an integrated form only *after* the problem is solved.

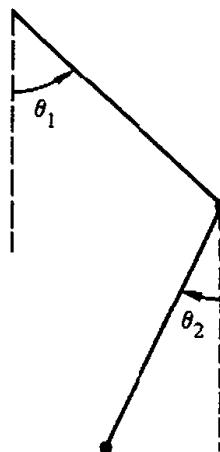


FIGURE 1.4 Double pendulum.

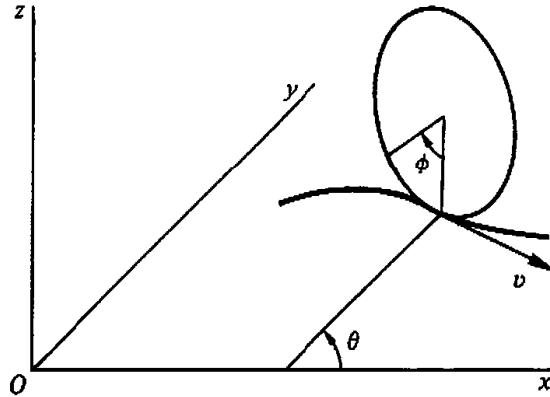


FIGURE 1.5 Vertical disk rolling on a horizontal plane.

A simple case will illustrate the point. Consider a disk rolling on the horizontal xy plane constrained to move so that the plane of the disk is always vertical. The coordinates used to describe the motion might be the x , y coordinates of the center of the disk, an angle of rotation ϕ about the axis of the disk, and an angle θ between the axis of the disk and say, the x axis (cf. Fig 1.5). As a result of the constraint the velocity of the center of the disk, v , has a magnitude proportional to $\dot{\phi}$,

$$v = a\dot{\phi},$$

where a is the radius of the disk, and its direction is perpendicular to the axis of the disk:

$$\dot{x} = v \sin \theta$$

$$\dot{y} = -v \cos \theta.$$

Combining these conditions, we have two *differential* equations of constraint:

$$\begin{aligned} dx - a \sin \theta d\phi &= 0, \\ dy + a \cos \theta d\phi &= 0. \end{aligned} \tag{1.39}$$

Neither of Eqs. (1.39) can be integrated without in fact solving the problem; i.e., we cannot find an integrating factor $f(x, y, \theta, \phi)$ that will turn either of the equations into perfect differentials (cf. Derivation 4).* Hence, the constraints cannot be reduced to the form of Eq. (1.37) and are therefore nonholonomic. Physically we can see that there can be no direct functional relation between ϕ and the other coordinates x , y , and θ by noting that at any point on its path the disk can be

*In principle, an integrating factor can always be found for a first-order differential equation of constraint in systems involving only two coordinates and such constraints are therefore holonomic. A familiar example is the two-dimensional motion of a circle rolling on an inclined plane.

made to roll around in a circle tangent to the path and of arbitrary radius. At the end of the process, x , y , and θ have been returned to their original values, but ϕ has changed by an amount depending on the radius of the circle.

Nonintegrable *differential* constraints of the form of Eqs. (1.39) are of course not the only type of nonholonomic constraints. The constraint conditions may involve higher-order derivatives, or may appear in the form of inequalities, as we have seen.

Partly because the dependent coordinates can be eliminated, problems involving holonomic constraints are always amenable to a formal solution. But there is no general way to attack nonholonomic examples. True, if the constraint is nonintegrable, the differential equations of constraint can be introduced into the problem along with the differential equations of motion, and the dependent equations eliminated, in effect, by the method of Lagrange multipliers.

We shall return to this method at a later point. However, the more vicious cases of nonholonomic constraint must be tackled individually, and consequently in the development of the more formal aspects of classical mechanics, it is almost invariably assumed that any constraint, if present, is holonomic. This restriction does not greatly limit the applicability of the theory, despite the fact that many of the constraints encountered in everyday life are nonholonomic. The reason is that the entire concept of constraints imposed in the system through the medium of wires or surfaces or walls is particularly appropriate only in macroscopic or large-scale problems. But today physicists are more interested in atomic and nuclear problems. On this scale all objects, both in and out of the system, consist alike of molecules, atoms, or smaller particles, exerting definite forces, and the notion of constraint becomes artificial and rarely appears. Constraints are then used only as mathematical idealizations to the actual physical case or as classical approximations to a quantum-mechanical property, e.g., rigid body rotations for "spin." Such constraints are always holonomic and fit smoothly into the framework of the theory.

To surmount the second difficulty, namely, that the forces of constraint are unknown *a priori*, we should like to so formulate the mechanics that the forces of constraint disappear. We need then deal only with the known applied forces. A hint as to the procedure to be followed is provided by the fact that in a particular system with constraints i.e. a rigid body, the work done by internal forces (which are here the forces of constraint) vanishes. We shall follow up this clue in the ensuing sections and generalize the ideas contained in it.

1.4 ■ D'ALEMBERT'S PRINCIPLE AND LAGRANGE'S EQUATIONS

A virtual (infinitesimal) displacement of a system refers to a change in the configuration of the system as the result of any arbitrary infinitesimal change of the coordinates $\delta \mathbf{r}_t$, *consistent with the forces and constraints imposed on the system at the given instant t*. The displacement is called virtual to distinguish it from an actual displacement of the system occurring in a time interval dt , during which

the forces and constraints may be changing. Suppose the system is in equilibrium; i.e., the total force on each particle vanishes, $\mathbf{F}_i = 0$. Then clearly the dot product $\mathbf{F}_i \cdot \delta\mathbf{r}_i$, which is the virtual work of the force \mathbf{F}_i in the displacement $\delta\mathbf{r}_i$, also vanishes. The sum of these vanishing products over all particles must likewise be zero:

$$\sum_i \mathbf{F}_i \cdot \delta\mathbf{r}_i = 0. \quad (1.40)$$

As yet nothing has been said that has any new physical content. Decompose \mathbf{F}_i into the applied force, $\mathbf{F}_i^{(a)}$, and the force of constraint, \mathbf{f}_i ,

$$\mathbf{F}_i = \mathbf{F}_i^{(a)} + \mathbf{f}_i. \quad (1.41)$$

so that Eq. (1.40) becomes

$$\sum_i \mathbf{F}_i^{(a)} \cdot \delta\mathbf{r}_i + \sum_i \mathbf{f}_i \cdot \delta\mathbf{r}_i = 0 \quad (1.42)$$

We now restrict ourselves to systems for which *the net virtual work of the forces of constraint is zero*. We have seen that this condition holds true for rigid bodies and it is valid for a large number of other constraints. Thus, if a particle is constrained to move on a surface, the force of constraint is perpendicular to the surface, while the virtual displacement must be tangent to it, and hence the virtual work vanishes. This is no longer true if sliding friction forces are present, and we must exclude such systems from our formulation. The restriction is not unduly hampering, since the friction is essentially a macroscopic phenomenon. On the other hand, the forces of rolling friction do not violate this condition, since the forces act on a point that is momentarily at rest and can do no work in an infinitesimal displacement consistent with the rolling constraint. Note that if a particle is constrained to a surface that is itself moving in time, the force of constraint is instantaneously perpendicular to the surface and the work during a virtual displacement is still zero even though the work during an actual displacement in the time dt does not necessarily vanish.

We therefore have as the condition for equilibrium of a system that the virtual work of the *applied forces* vanishes:

$$\sum_i \mathbf{F}_i^{(a)} \cdot \delta\mathbf{r}_i = 0. \quad (1.43)$$

Equation (1.43) is often called the *principle of virtual work*. Note that the coefficients of $\delta\mathbf{r}_i$ can no longer be set equal to zero; i.e., in general $\mathbf{F}_i^{(a)} \neq 0$, since the $\delta\mathbf{r}_i$ are not completely independent but are connected by the constraints. In order to equate the coefficients to zero, we must transform the principle into a form involving the virtual displacements of the q_i , which are independent. Equation (1.43) satisfies our needs in that it does not contain the \mathbf{f}_i , but it deals only with statics; we want a condition involving the general motion of the system.

Chapter 1 Survey of the Elementary Principles

To obtain such a principle, we use a device first thought of by James Bernoulli and developed by D'Alembert. The equation of motion,

$$\mathbf{F}_i = \dot{\mathbf{p}}_i,$$

can be written as

$$\mathbf{F}_i - \dot{\mathbf{p}}_i = 0,$$

which states that the particles in the system will be in equilibrium under a force equal to the actual force plus a “reversed effective force” $-\dot{\mathbf{p}}_i$. Instead of (1.40), we can immediately write

$$\sum_i (\mathbf{F}_i - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0, \quad (1.44)$$

and, making the same resolution into applied forces and forces of constraint, there results

$$\sum_i (\mathbf{F}_i^{(a)} - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i + \sum_i \mathbf{f}_i \cdot \delta \mathbf{r}_i = 0.$$

We again restrict ourselves to systems for which the virtual work of the forces of constraint vanishes and therefore obtain

$$\sum_i (\mathbf{F}_i^{(a)} - \dot{\mathbf{p}}_i) \cdot \delta \mathbf{r}_i = 0, \quad (1.45)$$

which is often called *D'Alembert's principle*. We have achieved our aim, in that the forces of constraint no longer appear, and the superscript ^(a) can now be dropped without ambiguity. It is still not in a useful form to furnish equations of motion for the system. We must now transform the principle into an expression involving virtual displacements of the generalized coordinates, which are then independent of each other (for holonomic constraints), so that the coefficients of the δq_j can be set separately equal to zero.

The translation from \mathbf{r}_i to q_j language starts from the transformation equations (1.38),

$$\mathbf{r}_i = \mathbf{r}_i(q_1, q_2, \dots, q_n, t) \quad (1.45')$$

(assuming n independent coordinates), and is carried out by means of the usual “chain rules” of the calculus of partial differentiation. Thus, \mathbf{v}_i is expressed in terms of the \dot{q}_k by the formula

$$\mathbf{v}_i \equiv \frac{d\mathbf{r}_i}{dt} = \sum_k \frac{\partial \mathbf{r}_i}{\partial q_k} \dot{q}_k + \frac{\partial \mathbf{r}_i}{\partial t}. \quad (1.46)$$

Similarly, the arbitrary virtual displacement $\delta\mathbf{r}_i$ can be connected with the virtual displacements δq_j by

$$\delta\mathbf{r}_i = \sum_j \frac{\partial\mathbf{r}_i}{\partial q_j} \delta q_j \quad (1.47)$$

Note that no variation of time, δt , is involved here, since a virtual displacement by definition considers only displacements of the coordinates. (Only then is the virtual displacement perpendicular to the force of constraint if the constraint itself is changing in time.)

In terms of the generalized coordinates, the virtual work of the \mathbf{F}_i becomes

$$\begin{aligned} \sum_i \mathbf{F}_i \cdot \delta\mathbf{r}_i &= \sum_{i,j} \mathbf{F}_i \cdot \frac{\partial\mathbf{r}_i}{\partial q_j} \delta q_j \\ &= \sum_j Q_j \delta q_j, \end{aligned} \quad (1.48)$$

where the Q_j are called the components of the *generalized force*, defined as

$$Q_j = \sum_i \mathbf{F}_i \cdot \frac{\partial\mathbf{r}_i}{\partial q_j}. \quad (1.49)$$

Note that just as the q 's need not have the dimensions of length, so the Q 's do not necessarily have the dimensions of force, but $Q_j \delta q_j$ must always have the dimensions of work. For example, Q_j might be a torque N_j and dq_j a differential angle $d\theta_j$, which makes $N_j d\theta_j$ a differential of work.

We turn now to the other term involved in Eq. (1.45), which may be written as

$$\sum_i \dot{\mathbf{p}}_i \cdot \delta\mathbf{r}_i = \sum_i m_i \ddot{\mathbf{r}}_i \cdot \delta\mathbf{r}_i.$$

Expressing $\delta\mathbf{r}_i$ by (1.47), this becomes

$$\sum_{i,j} m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial\mathbf{r}_i}{\partial q_j} \delta q_j.$$

Consider now the relation

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial\mathbf{r}_i}{\partial q_j} = \sum_i \left[\frac{d}{dt} \left(m_i \dot{\mathbf{r}}_i \cdot \frac{\partial\mathbf{r}_i}{\partial q_j} \right) - m_i \dot{\mathbf{r}}_i \cdot \frac{d}{dt} \left(\frac{\partial\mathbf{r}_i}{\partial q_j} \right) \right]. \quad (1.50)$$

In the last term of Eq. (1.50) we can interchange the differentiation with respect to t and q_j , for, in analogy to (1.46).

$$\begin{aligned} \frac{d}{dt} \left(\frac{\partial \mathbf{r}_i}{\partial q_j} \right) &= \frac{\partial \dot{\mathbf{r}}_i}{\partial q_j} = \sum_k \frac{\partial^2 \mathbf{r}_i}{\partial q_j \partial q_k} \dot{q}_k + \frac{\partial^2 \mathbf{r}_i}{\partial q_j \partial t} \\ &\quad - \frac{\partial \mathbf{v}_i}{\partial q_j}, \end{aligned}$$

by Eq. (1.46). Further, we also see from Eq. (1.46) that

$$\frac{\partial \mathbf{v}_i}{\partial \dot{q}_j} = \frac{\partial \mathbf{r}_i}{\partial q_j}. \quad (1.51)$$

Substitution of these changes in (1.50) leads to the result that

$$\sum_i m_i \ddot{\mathbf{r}}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = \sum_i \left[\frac{d}{dt} \left(m_i \mathbf{v}_i \cdot \frac{\partial \mathbf{v}_i}{\partial \dot{q}_j} \right) - m_i \mathbf{v}_i \cdot \frac{\partial \mathbf{v}_i}{\partial q_j} \right],$$

and the second term on the left-hand side of Eq. (1.45) can be expanded into

$$\sum_j \left\{ \frac{d}{dt} \left[\frac{\partial}{\partial \dot{q}_j} \left(\sum_i \frac{1}{2} m_i v_i^2 \right) \right] - \frac{\partial}{\partial q_j} \left(\sum_i \frac{1}{2} m_i v_i^2 \right) - Q_j \right\} \delta q_j.$$

Identifying $\sum_i \frac{1}{2} m_i v_i^2$ with the system kinetic energy T , D'Alembert's principle (cf. Eq. (1.45)) becomes

$$\sum_j \left\{ \left[\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial T}{\partial q_j} \right] - Q_j \right\} \delta q_j = 0. \quad (1.52)$$

Note that in a system of Cartesian coordinates the partial derivative of T with respect to q_j vanishes. Thus, speaking in the language of differential geometry, this term arises from the curvature of the coordinates q_j . In polar coordinates, e.g., it is in the partial derivative of T with respect to an angle coordinate that the centripetal acceleration term appears.

Thus far, no restriction has been made on the nature of the constraints other than that they be workless in a virtual displacement. The variables q_j can be any set of coordinates used to describe the motion of the system. If, however, the constraints are holonomic, then it is possible to find sets of independent coordinates q_j that contain the constraint conditions implicitly in the transformation equations (1.38). Any virtual displacement δq_j is then independent of δq_k , and therefore the only way for (1.52) to hold is for the individual coefficients to vanish:

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial T}{\partial q_j} - Q_j. \quad (1.53)$$

There are n such equations in all.

When the forces are derivable from a scalar potential function V ,

$$\mathbf{F}_i = -\nabla_i V.$$

Then the generalized forces can be written as

$$Q_j = \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = - \sum_i \nabla_i V \cdot \frac{\partial \mathbf{r}_i}{\partial q_j},$$

which is exactly the same expression for the partial derivative of a function $-V(\mathbf{r}_1, \mathbf{r}_2, \dots, \mathbf{r}_N, t)$ with respect to q_j :

$$Q_j = -\frac{\partial V}{\partial q_j}. \quad (1.54)$$

Equations (1.53) can then be rewritten as

$$\frac{d}{dt} \left(\frac{\partial T}{\partial \dot{q}_j} \right) - \frac{\partial(T - V)}{\partial q_j} = 0. \quad (1.55)$$

The equations of motion in the form (1.55) are not necessarily restricted to conservative systems, only if V is not an explicit function of time is the system conservative (cf. p. 4). As here defined, the potential V does not depend on the generalized velocities. Hence, we can include a term in V in the partial derivative with respect to \dot{q}_j :

$$\frac{d}{dt} \left(\frac{\partial(T - V)}{\partial \dot{q}_j} \right) - \frac{\partial(T - V)}{\partial q_j} = 0.$$

Or, defining a new function, the *Lagrangian* L , as

$$L = T - V, \quad (1.56)$$

the Eqs. (1.53) become

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} = 0, \quad (1.57)$$

expressions referred to as "Lagrange's equations."

Note that for a particular set of equations of motion there is no unique choice of Lagrangian such that Eqs. (1.57) lead to the equations of motion in the given generalized coordinates. Thus, in Derivations 8 and 10 it is shown that if $L(q, \dot{q}, t)$ is an approximate Lagrangian and $F(q, t)$ is *any* differentiable function of the generalized coordinates and time, then

$$L'(q, \dot{q}, t) = L(q, \dot{q}, t) + \frac{dF}{dt} \quad (1.57')$$

is a Lagrangian also resulting in the same equations of motion. It is also often possible to find alternative Lagrangians beside those constructed by this prescription (see Exercise 20). While Eq. (1.56) is always a suitable way to construct a Lagrangian for a conservative system, it does not provide the *only* Lagrangian suitable for the given system.

1.5 ■ VELOCITY-DEPENDENT POTENTIALS AND THE DISSIPATION FUNCTION

Lagrange's equations can be put in the form (1.57) even if there is no potential function, V , in the usual sense, providing the generalized forces are obtained from a function $U(q_j, \dot{q}_j)$ by the prescription

$$Q_j = -\frac{\partial U}{\partial q_j} + \frac{d}{dt} \left(\frac{\partial U}{\partial \dot{q}_j} \right). \quad (1.58)$$

In such case, Eqs. (1.57) still follow from Eqs. (1.53) with the Lagrangian given by

$$L = T - U. \quad (1.59)$$

Here U may be called a "generalized potential," or "velocity-dependent potential." The possibility of using such a "potential" is not academic; it applies to one very important type of force field, namely, the electromagnetic forces on moving charges. Considering its importance, a digression on this subject is well worthwhile.

Consider an electric charge, q , of mass m moving at a velocity, \mathbf{v} , in an otherwise charge-free region containing both an electric field, \mathbf{E} , and a magnetic field, \mathbf{B} , which may depend upon time and position. The charge experiences a force, called the Lorentz force, given by

$$\mathbf{F} = q[\mathbf{E} + (\mathbf{v} \times \mathbf{B})]. \quad (1.60)$$

Both $\mathbf{E}(t, x, y, z)$ and $\mathbf{B}(t, x, y, z)$ are continuous functions of time and position derivable from a scalar potential $\phi(t, x, y, z)$ and a vector potential $\mathbf{A}(t, x, y, z)$ by

$$\mathbf{E} = -\nabla\phi - \frac{\partial \mathbf{A}}{\partial t} \quad (1.61a)$$

and

$$\mathbf{B} = \nabla \times \mathbf{A}. \quad (1.61b)$$

The force on the charge can be derived from the following velocity-dependent potential energy

$$U = q\phi - q\mathbf{A} \cdot \mathbf{v}, \quad (1.62)$$

so the Lagrangian, $L = T - U$, is

$$L = \frac{1}{2}m\mathbf{v}^2 - q\phi + q\mathbf{A} \cdot \mathbf{v}. \quad (1.63)$$

Considering just the x -component of Lagrange's equations gives

$$m\ddot{x} = q \left(v_x \frac{\partial A_x}{\partial x} + v_y \frac{\partial A_y}{\partial x} + v_z \frac{\partial A_z}{\partial x} \right) - q \left(\frac{\partial \phi}{\partial x} + \frac{dA_x}{dt} \right). \quad (1.64)$$

The total time derivative of A_x is related to the particle time derivative through

$$\begin{aligned} \frac{dA_x}{dt} &= \frac{\partial A_x}{\partial t} + \mathbf{v} \cdot \nabla A_x \\ &= \frac{\partial A_x}{\partial t} + v_x \frac{\partial A_x}{\partial x} + v_y \frac{\partial A_x}{\partial y} + v_z \frac{\partial A_x}{\partial z}. \end{aligned} \quad (1.65)$$

Equation (1.61b) gives

$$(\mathbf{v} \times \mathbf{B})_x = v_y \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right) + v_z \left(\frac{\partial A_z}{\partial x} - \frac{\partial A_x}{\partial z} \right).$$

Combining these expressions gives the equation of motion in the x -direction

$$m\ddot{x} = q [E_x + (\mathbf{v} \times \mathbf{B})_x]. \quad (1.66)$$

On a component-by-component comparison, Eqs. (1.66) and (1.60) are identical, showing that the Lorentz force equation is derivable from Eqs. (1.61) and (1.62).

Note that if not all the forces acting on the system are derivable from a potential, then Lagrange's equations can always be written in the form

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_j} \right) - \frac{\partial L}{\partial q_j} = Q_j,$$

where L contains the potential of the conservative forces as before, and Q_j represents the forces *not* arising from a potential. Such a situation often occurs when frictional forces are present. It frequently happens that the frictional force is proportional to the velocity of the particle, so that its x -component has the form

$$F_{fx} = -k_x v_x.$$

Frictional forces of this type may be derived in terms of a function \mathcal{F} , known as *Rayleigh's dissipation function*, and defined as

$$\mathcal{F} = \frac{1}{2} \sum_i \left(k_x v_{ix}^2 + k_y v_{iy}^2 + k_z v_{iz}^2 \right), \quad (1.67)$$

where the summation is over the particles of the system. From this definition it is clear that

$$F_{fx} = -\frac{\partial \mathcal{F}}{\partial v_x}.$$

or, symbolically,

$$\mathbf{F}_f = -\nabla_v \mathcal{F}. \quad (1.68)$$

We can also give a physical interpretation to the dissipation function. The work done by the system *against* friction is

$$dW_f = -\mathbf{F}_f \cdot d\mathbf{r} = -\mathbf{F}_f \cdot \mathbf{v} dt = (k_x v_x^2 + k_y v_y^2 + k_z v_z^2) dt.$$

Hence, $2\mathcal{F}$ is the rate of energy dissipation due to friction. The component of the generalized force resulting from the force of friction is then given by

$$\begin{aligned} Q_J &= \sum_i \mathbf{F}_{f_i} \cdot \frac{\partial \mathbf{r}_i}{\partial q_J} = - \sum_i \nabla_v \mathcal{F} \cdot \frac{\partial \mathbf{r}_i}{\partial q_i} \\ &= - \sum_i \nabla_v \mathcal{F} \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_J}, \quad \text{by (1.51),} \\ &= - \frac{\partial \mathcal{F}}{\partial \dot{q}_J}. \end{aligned} \quad (1.69)$$

An example is Stokes' law, whereby a sphere of radius a moving at a speed v , in a medium of viscosity η experiences the frictional drag force $\mathbf{F}_f = 6\pi\eta av$. The Lagrange equations with dissipation become

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_J} \right) - \frac{\partial L}{\partial q_J} + \frac{\partial \mathcal{F}}{\partial \dot{q}_J} = 0, \quad (1.70)$$

so that two scalar functions, L and \mathcal{F} , must be specified to obtain the equations of motion.

1.6 ■ SIMPLE APPLICATIONS OF THE LAGRANGIAN FORMULATION

The previous sections show that for systems where we can define a Lagrangian, i.e., holonomic systems with applied forces derivable from an ordinary or generalized potential and workless constraints, we have a very convenient way of setting up the equations of motion. We were led to the Lagrangian formulation by the desire to eliminate the forces of constraint from the equations of motion, and in achieving this goal we have obtained many other benefits. In setting up the original form of the equations of motion, Eqs. (1.19), it is necessary to work with many *vector* forces and accelerations. With the Lagrangian method we only deal with two *scalar* functions, T and V , which greatly simplifies the problem.

A straightforward routine procedure can now be established for all problems of mechanics to which the Lagrangian formulation is applicable. We have only to write T and V in generalized coordinates, form L from them, and substitute in (1.57) to obtain the equations of motion. The needed transformation of T and V from Cartesian coordinates to generalized coordinates is obtained by applying the

transformation equations (1.38) and (1.45'). Thus, T is given in general by

$$T = \sum_i \frac{1}{2} m_i v_i^2 = \sum_i \frac{1}{2} m_i \left(\sum_j \frac{\partial \mathbf{r}_i}{\partial q_j} \dot{q}_j + \frac{\partial \mathbf{r}_i}{\partial t} \right)^2.$$

It is clear that on carrying out the expansion, the expression for T in generalized coordinates will have the form

$$T = M_0 + \sum_j M_j \dot{q}_j + \frac{1}{2} \sum_{j,k} M_{jk} \dot{q}_j \dot{q}_k, \quad (1.71)$$

where M_0 , M_j , M_{jk} are definite functions of the \mathbf{r} 's and t and hence of the q 's and t . In fact, a comparison shows that

$$\begin{aligned} M_0 &= \sum_i \frac{1}{2} m_i \left(\frac{\partial \mathbf{r}_i}{\partial t} \right)^2, \\ M_j &= \sum_i m_i \frac{\partial \mathbf{r}_i}{\partial t} \cdot \frac{\partial \mathbf{r}_i}{\partial q_j}, \end{aligned} \quad (1.72)$$

and

$$M_{jk} = \sum_i m_i \frac{\partial \mathbf{r}_i}{\partial q_j} \cdot \frac{\partial \mathbf{r}_i}{\partial q_k}.$$

Thus, the kinetic energy of a system can always be written as the sum of three homogeneous functions of the generalized velocities,

$$T = T_0 + T_1 + T_2, \quad (1.73)$$

where T_0 is independent of the generalized velocities, T_1 is linear in the velocities, and T_2 is quadratic in the velocities. If the transformation equations do not contain the time explicitly, as may occur when the constraints are independent of time (scleronomous), then only the last term in Eq. (1.71) is nonvanishing, and T is always a homogeneous quadratic form in the generalized velocities.

Let us now consider simple examples of this procedure:

1. Single particle in space
 - (a) Cartesian coordinates
 - (b) Plane polar coordinates
2. Atwood's machine
3. Time-dependent constraint—bead sliding on rotating wire

1. (a) *Motion of one particle: using Cartesian coordinates.* The generalized forces needed in Eq. (1.53) are obviously F_x , F_y , and F_z . Then

$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2),$$

$$\frac{\partial T}{\partial x} = \frac{\partial T}{\partial y} = \frac{\partial T}{\partial z} = 0,$$

$$\frac{\partial T}{\partial \dot{x}} = m\dot{x}, \quad \frac{\partial T}{\partial \dot{y}} = m\dot{y}, \quad \frac{\partial T}{\partial \dot{z}} = m\dot{z},$$

and the equations of motion are

$$\frac{d}{dt}(m\dot{x}) = F_x, \quad \frac{d}{dt}(m\dot{y}) = F_y, \quad \frac{d}{dt}(m\dot{z}) = F_z. \quad (1.74)$$

We are thus led back to the original Newton's equations of motion.

(b) *Motion of one particle: using plane polar coordinates.* Here we must express T in terms of \dot{r} and $\dot{\theta}$. The equations of transformation, i.e., Eqs. (1.38), in this case are simply

$$x = r \cos \theta$$

$$y = r \sin \theta.$$

By analogy to (1.46), the velocities are given by

$$\dot{x} = \dot{r} \cos \theta - r\dot{\theta} \sin \theta,$$

$$\dot{y} = \dot{r} \sin \theta + r\dot{\theta} \cos \theta.$$

The kinetic energy $T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2)$ then reduces formally to

$$T = \frac{1}{2}m[\dot{r}^2 + (r\dot{\theta})^2]. \quad (1.75)$$

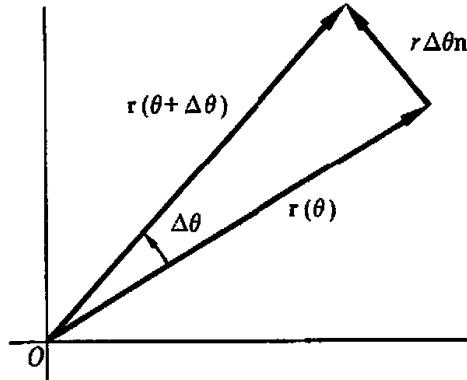
An alternative derivation of Eq. (1.75) is obtained by recognizing that the plane polar components of the velocity are \dot{r} along \mathbf{r} , and $r\dot{\theta}$ along the direction perpendicular to \mathbf{r} , denoted by the unit vector $\hat{\mathbf{n}}$. Hence, the square of the velocity expressed in polar coordinates is simply $\dot{r}^2 + (r\dot{\theta})^2$. With the aid of the expression

$$d\mathbf{r} = \hat{\mathbf{r}} dr + r\hat{\theta} d\theta + \hat{\mathbf{k}} dz$$

for the differential position vector, $d\mathbf{r}$, in cylindrical coordinates, where $\hat{\mathbf{r}}$ and $\hat{\theta}$ are unit vectors in the \mathbf{r} and θ -directions, respectively, the components of the generalized force can be obtained from the definition, Eq. (1.49),

$$Q_r = \mathbf{F} \cdot \frac{\partial \mathbf{r}}{\partial r} = \mathbf{F} \cdot \hat{\mathbf{r}} = F_r,$$

$$Q_\theta = \mathbf{F} \cdot \frac{\partial \mathbf{r}}{\partial \theta} = \mathbf{F} \cdot r\hat{\theta} = rF_\theta,$$

FIGURE 1.6 Derivative of r with respect to θ .

since the derivative of \mathbf{r} with respect to θ is, by the definition of a derivative, a vector in the direction of $\hat{\theta}$ (cf. Fig. 1.6). There are two generalized coordinates, and therefore two Lagrange equations. The derivatives occurring in the r equation are

$$\frac{\partial T}{\partial r} = mr\dot{\theta}^2, \quad \frac{\partial T}{\partial \dot{r}} = m\dot{r}, \quad \frac{d}{dt} \left(\frac{\partial T}{\partial \dot{r}} \right) = m\ddot{r},$$

and the equation itself is

$$m\ddot{r} - mr\dot{\theta}^2 = F_r,$$

the second term being the centripetal acceleration term. For the θ equation, we have the derivatives

$$\frac{\partial T}{\partial \theta} = 0, \quad \frac{\partial T}{\partial \dot{\theta}} = mr^2\dot{\theta}, \quad \frac{d}{dt} \left(mr^2\dot{\theta} \right) = mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta},$$

so that the equation becomes

$$\frac{d}{dt} \left(mr^2\dot{\theta} \right) = mr^2\ddot{\theta} + 2mr\dot{r}\dot{\theta} = rF_\theta.$$

Note that the left side of the equation is just the time derivative of the angular momentum, and the right side is exactly the applied torque, so that we have simply rederived the torque equation (1.26), where $L = mr^2\dot{\theta}$ and $N^{(\ell)} = rF_\theta$.

2. *Atwood's machine*—(See Fig. 1.7) an example of a conservative system with holonomic, scleronomous constraint (the pulley is assumed frictionless and massless). Clearly there is only one independent coordinate x , the position of the other weight being determined by the constraint that the length of the rope between them is l . The potential energy is

$$V = -M_1gx - M_2g(l - x),$$

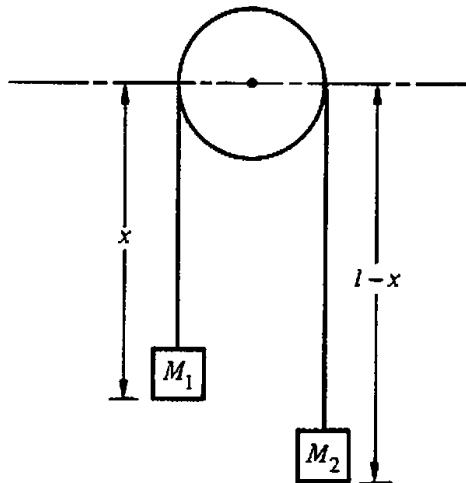


FIGURE 1.7 Atwood's machine.

while the kinetic energy is

$$T = \frac{1}{2} (M_1 + M_2) \dot{x}^2.$$

Combining the two, the Lagrangian has the form

$$L = T - V = \frac{1}{2} (M_1 + M_2) \dot{x}^2 + M_1 g x + M_2 g (l - x).$$

There is only one equation of motion, involving the derivatives

$$\begin{aligned}\frac{\delta L}{\delta x} &= (M_1 - M_2) g, \\ \frac{\delta L}{\delta \dot{x}} &= (M_1 + M_2) \dot{x},\end{aligned}$$

so that we have

$$(M_1 + M_2) \ddot{x} = (M_1 - M_2) g,$$

or

$$\ddot{x} = \frac{M_1 - M_2}{M_1 + M_2} g,$$

which is the familiar result obtained by more elementary means. This trivial problem emphasizes that the forces of constraint—here the tension in the rope—appear nowhere in the Lagrangian formulation. By the same token, neither can the tension in the rope be found directly by the Lagrangian method.

3. *A bead (or ring) sliding on a uniformly rotating wire in a force-free space.* The wire is straight, and is rotated uniformly about some fixed axis perpendicular to the wire. This example has been chosen as a simple illustration of a constraint

being time dependent, with the rotation axis along z and the wire in the xy plane. The transformation equations explicitly contain the time.

$$x = r \cos \omega t. \quad (\omega = \text{angular velocity of rotation})$$

$$y = r \sin \omega t. \quad (r = \text{distance along wire from rotation axis})$$

While we could then find T (here the same as L) by the same procedure used to obtain (1.71), it is simpler to take over (1.75) directly, expressing the constraint by the relation $\dot{\theta} = \omega$:

$$T = \frac{1}{2}m(\dot{r}^2 + r^2\omega^2).$$

Note that T is not a homogeneous quadratic function of the generalized velocities, since there is now an additional term not involving \dot{r} . The equation of motion is then

$$m\ddot{r} = mr\omega^2 = 0$$

or

$$\ddot{r} = r\omega^2,$$

which is the familiar simple harmonic oscillator equation with a change of sign. The solution $r = e^{\omega t}$ shows that the bead moves exponentially outward because of the centripetal acceleration. Again, the method cannot furnish the force of constraint that keeps the bead on the wire. Equation (1.26) with the angular momentum, $\mathbf{L} = mr^2\omega^2e^{\omega t}$, provides the force $\mathbf{F} = \mathbf{N}/r$, which produces the constraint force, $F = mr\omega^2e^{\omega t}$, acting perpendicular to the wire and the axis of rotation.

DERIVATIONS

1. Show that for a single particle with constant mass the equation of motion implies the following differential equation for the kinetic energy:

$$\frac{dT}{dt} = \mathbf{F} \cdot \mathbf{v},$$

while if the mass varies with time the corresponding equation is

$$\frac{d(mT)}{dt} = \mathbf{F} \cdot \mathbf{p}.$$

2. Prove that the magnitude R of the position vector for the center of mass from an arbitrary origin is given by the equation

$$M^2 R^2 = M \sum_i m_i r_i^2 - \frac{1}{2} \sum_{i,j} m_i m_j r_{ij}^2.$$

3. Suppose a system of two particles is known to obey the equations of motion, Eqs. (1.22) and (1.26). From the equations of the motion of the individual particles show that the internal forces between particles satisfy both the weak and the strong laws of action and reaction. The argument may be generalized to a system with arbitrary number of particles, thus proving the converse of the arguments leading to Eqs. (1.22) and (1.26).
4. The equations of constraint for the rolling disk, Eqs. (1.39), are special cases of general linear differential equations of constraint of the form

$$\sum_{i=1}^n g_i(x_1, \dots, x_n) dx_i = 0.$$

A constraint condition of this type is holonomic only if an integrating function $f(x_1, \dots, x_n)$ can be found that turns it into an exact differential. Clearly the function must be such that

$$\frac{\partial(fg_i)}{\partial x_j} = \frac{\partial(fg_j)}{\partial x_i}$$

for all $i \neq j$. Show that no such integrating factor can be found for either of Eqs. (1.39).

5. Two wheels of radius a are mounted on the ends of a common axle of length b such that the wheels rotate independently. The whole combination rolls without slipping on a plane. Show that there are two nonholonomic equations of constraint,

$$\begin{aligned} \cos \theta dx + \sin \theta dy &= 0 \\ \sin \theta dx - \cos \theta dv &= \frac{1}{2}a(d\phi + d\phi'), \end{aligned}$$

(where θ , ϕ , and ϕ' have meanings similar to those in the problem of a single vertical disk, and (x, y) are the coordinates of a point on the axle midway between the two wheels) and one holonomic equation of constraint,

$$\theta = C - \frac{a}{b}(\phi - \phi').$$

where C is a constant.

6. A particle moves in the xy plane under the constraint that its velocity vector is always directed towards a point on the x axis whose abscissa is some given function of time $f(t)$. Show that for $f(t)$ differentiable, but otherwise arbitrary, the constraint is nonholonomic.
7. Show that Lagrange's equations in the form of Eqs. (1.53) can also be written as

$$\frac{\partial \dot{T}}{\partial \dot{q}_j} - 2 \frac{\partial T}{\partial q_j} = Q_j.$$

These are sometimes known as the *Nielsen* form of the Lagrange equations.

8. If L is a Lagrangian for a system of n degrees of freedom satisfying Lagrange's equations, show by direct substitution that

$$L' = L + \frac{dF(q_1, \dots, q_n, t)}{dt}$$

also satisfies Lagrange's equations where F is any arbitrary, but differentiable, function of its arguments.

9. The electromagnetic field is invariant under a gauge transformation of the scalar and vector potential given by

$$\mathbf{A} \rightarrow \mathbf{A} + \nabla\psi(\mathbf{r}, t),$$

$$\phi \rightarrow \phi - \frac{1}{c} \frac{\partial \psi}{\partial t},$$

where ψ is arbitrary (but differentiable). What effect does this gauge transformation have on the Lagrangian of a particle moving in the electromagnetic field? Is the motion affected?

10. Let q_1, \dots, q_n be a set of independent generalized coordinates for a system of n degrees of freedom, with a Lagrangian $L(q, \dot{q}, t)$. Suppose we transform to another set of independent coordinates s_1, \dots, s_n by means of transformation equations

$$q_i = q_i(s_1, \dots, s_n, t), \quad i = 1, \dots, n.$$

(Such a transformation is called a *point transformation*.) Show that if the Lagrangian function is expressed as a function of s_j , \dot{s}_j , and t through the equations of transformation, then L satisfies Lagrange's equations with respect to the s coordinates:

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{s}_j} \right) - \frac{\partial L}{\partial s_j} = 0.$$

In other words, the form of Lagrange's equations is invariant under a point transformation.

EXERCISES

11. Consider a uniform thin disk that rolls without slipping on a horizontal plane. A horizontal force is applied to the center of the disk and in a direction parallel to the plane of the disk.
- Derive Lagrange's equations and find the generalized force.
 - Discuss the motion if the force is not applied parallel to the plane of the disk.
12. The *escape velocity* of a particle on Earth is the minimum velocity required at Earth's surface in order that the particle can escape from Earth's gravitational field. Neglecting the resistance of the atmosphere, the system is conservative. From the conservation theorem for potential plus kinetic energy show that the escape velocity for Earth, ignoring the presence of the Moon, is 11.2 km/s.
13. Rockets are propelled by the momentum reaction of the exhaust gases expelled from the tail. Since these gases arise from the reaction of the fuels carried in the rocket, the mass of the rocket is not constant, but decreases as the fuel is expended. Show that the equation of motion for a rocket projected vertically upward in a uniform gravitational

field, neglecting atmospheric friction, is

$$m \frac{dv}{dt} = -v' \frac{dm}{dt} - mg,$$

where m is the mass of the rocket and v' is the velocity of the escaping gases relative to the rocket. Integrate this equation to obtain v as a function of m , assuming a constant time rate of loss of mass. Show, for a rocket starting initially from rest, with v' equal to 2.1 m/s and a mass loss per second equal to 1/60th of the initial mass, that in order to reach the escape velocity the ratio of the weight of the fuel to the weight of the empty rocket must be almost 300!

14. Two points of mass m are joined by a rigid weightless rod of length l , the center of which is constrained to move on a circle of radius a . Express the kinetic energy in generalized coordinates.
15. A point particle moves in space under the influence of a force derivable from a generalized potential of the form

$$U(\mathbf{r}, \mathbf{v}) = V(r) + \boldsymbol{\sigma} \cdot \mathbf{L},$$

where \mathbf{r} is the radius vector from a fixed point, \mathbf{L} is the angular momentum about that point, and $\boldsymbol{\sigma}$ is a fixed vector in space.

- (a) Find the components of the force on the particle in both Cartesian and spherical polar coordinates, on the basis of Eq. (1.58).
- (b) Show that the components in the two coordinate systems are related to each other as in Eq. (1.49).
- (c) Obtain the equations of motion in spherical polar coordinates.
16. A particle moves in a plane under the influence of a force, acting toward a center of force, whose magnitude is

$$F = \frac{1}{r^2} \left(1 - \frac{\dot{r}^2 - 2\dot{r}r}{c^2} \right),$$

where r is the distance of the particle to the center of force. Find the generalized potential that will result in such a force, and from that the Lagrangian for the motion in a plane. (The expression for F represents the force between two charges in Weber's electrodynamics.)

17. A nucleus, originally at rest, decays radioactively by emitting an electron of momentum 1.73 MeV/c, and at right angles to the direction of the electron a neutrino with momentum 1.00 MeV/c. (The MeV, million electron volt, is a unit of energy used in modern physics, equal to 1.60×10^{-13} J. Correspondingly, MeV/c is a unit of linear momentum equal to 5.34×10^{-22} kg·m/s.) In what direction does the nucleus recoil? What is its momentum in MeV/c? If the mass of the residual nucleus is 3.90×10^{-25} kg what is its kinetic energy, in electron volts?
18. A Lagrangian for a particular physical system can be written as

$$L' = \frac{m}{2} (a\dot{x}^2 + 2b\dot{x}\dot{y} + c\dot{y}^2) - \frac{K}{2} (ax^2 + 2bxy + cy^2),$$

where a , b , and c are arbitrary constants but subject to the condition that $b^2 - ac \neq 0$.

What are the equations of motion? Examine particularly the two cases $a = 0 = c$ and $b = 0, c = -a$. What is the physical system described by the above Lagrangian? Show that the usual Lagrangian for this system as defined by Eq. (1.57') is related to L' by a point transformation (cf. Derivation 10). What is the significance of the condition on the value of $b^2 - ac$?

19. Obtain the Lagrange equations of motion for a spherical pendulum, i.e., a mass point suspended by a rigid weightless rod.
20. A particle of mass m moves in one dimension such that it has the Lagrangian

$$L = \frac{m^2 \dot{x}^4}{12} + m x^2 V(x) - V_2(x),$$

where V is some differentiable function of x . Find the equation of motion for $x(t)$ and describe the physical nature of the system on the basis of this equation

21. Two mass points of mass m_1 and m_2 are connected by a string passing through a hole in a smooth table so that m_1 rests on the table surface and m_2 hangs suspended. Assuming m_2 moves only in a vertical line, what are the generalized coordinates for the system? Write the Lagrange equations for the system and, if possible, discuss the physical significance any of them might have. Reduce the problem to a single second-order differential equation and obtain a first integral of the equation. What is its physical significance? (Consider the motion only until m_1 reaches the hole.)
22. Obtain the Lagrangian and equations of motion for the double pendulum illustrated in Fig. 1.4, where the lengths of the pendula are l_1 and l_2 with corresponding masses m_1 and m_2 .
23. Obtain the equation of motion for a particle falling vertically under the influence of gravity when frictional forces obtainable from a dissipation function $\frac{1}{2}k v^2$ are present. Integrate the equation to obtain the velocity as a function of time and show that the maximum possible velocity for a fall from rest is $v = mg/k$.
24. A spring of rest length L_a (no tension) is connected to a support at one end and has a mass M attached at the other. Neglect the mass of the spring, the dimension of the mass M , and assume that the motion is confined to a vertical plane. Also, assume that the spring only stretches without bending but it can swing in the plane.
 - (a) Using the angular displacement of the mass from the vertical and the length that the string has stretched from its rest length (hanging with the mass m), find Lagrange's equations.
 - (b) Solve these equations for small stretching and angular displacements.
 - (c) Solve the equations in part (a) to the next order in both stretching and angular displacement. This part is amenable to hand calculations. Using some reasonable assumptions about the spring constant, the mass, and the rest length, discuss the motion. Is a resonance likely under the assumptions stated in the problem?
 - (d) (For analytic computer programs.) Consider the spring to have a total mass $m \ll M$. Neglecting the bending of the spring, set up Lagrange's equations correctly to first order in m and the angular and linear displacements.
 - (e) (For numerical computer analysis.) Make sets of reasonable assumptions of the constants in part (a) and make a single plot of the two coordinates as functions of time.

CHAPTER

2

Variational Principles and Lagrange's Equations

2.1 ■ HAMILTON'S PRINCIPLE

The derivation of Lagrange's equations presented in Chapter 1 started from a consideration of the instantaneous state of the system and small virtual displacements about the instantaneous state, i.e., from a "differential principle" such as D'Alembert's principle. It is also possible to obtain Lagrange's equations from a principle that considers the entire motion of the system between times t_1 and t_2 , and small virtual variations of this motion from the actual motion. A principle of this nature is known as an "integral principle."

Before presenting the integral principle, the meaning attached to the phrase "motion of the system between times t_1 and t_2 " must first be stated in more precise language. The instantaneous configuration of a system is described by the values of the n generalized coordinates q_1, \dots, q_n , and corresponds to a particular point in a Cartesian hyperspace where the q 's form the n coordinate axes. This n -dimensional space is therefore known as configuration space. As time goes on, the state of the system changes and the system point moves in configuration space tracing out a curve, described as "the path of motion of the system." The "motion of the system," as used above, then refers to the motion of the system point along this path in *configuration space*. Time can be considered formally as a parameter of the curve; to each point on the path there is associated one or more values of the time. Note that configuration space has no necessary connection with the physical three-dimensional space, just as the generalized coordinates are not necessarily position coordinates. The path of motion in configuration space has no resemblance to the path in space of any actual particle; each point on the path represents the *entire* system configuration at some given instant of time.

The integral *Hamilton's principle* describes the motion of those mechanical systems for which all forces (except the forces of constraint) are derivable from a generalized scalar potential that may be a function of the coordinates, velocities, and time. Such systems will be denoted as *monogenic*. Where the potential is an explicit function of position coordinates only, then a monogenic system is also conservative (cf. Section 1.2).

For monogenic systems, Hamilton's principle can be stated as

The motion of the system from time t_1 to time t_2 is such that the line integral (called the action or the action integral),

$$I = \int_{t_1}^{t_2} L dt, \quad (2.1)$$

where $L = T - V$, has a stationary value for the actual path of the motion.

That is, out of all possible paths by which the system point could travel from its position at time t_1 to its position at time t_2 , it will actually travel along that path for which the value of the integral (2.1) is stationary. By the term "stationary value" for a line integral, we mean that the integral along the given path has the same value to within first-order infinitesimals as that along all neighboring paths (i.e., those that differ from it by infinitesimal displacements). (Cf. Fig. 2.1.) The notion of a stationary value for a line integral thus corresponds in ordinary function theory to the vanishing of the first derivative.

We can summarize Hamilton's principle by saying that the motion is such that the *variation* of the line integral I for fixed t_1 and t_2 is zero:

$$\delta I = \delta \int_{t_1}^{t_2} L(q_1, \dots, q_n, \dot{q}_1, \dots, \dot{q}_n, t) dt = 0. \quad (2.2)$$

Where the system constraints are holonomic, Hamilton's principle, Eq. (2.2), is both a necessary and sufficient condition for Lagrange's equations, Eqs. (1.57). Thus, it can be shown that Hamilton's principle follows directly from Lagrange's equations. Instead, however, we shall prove the converse, namely, that Lagrange's equations follow from Hamilton's principle, as being the more important theorem. That Hamilton's principle is a sufficient condition for deriving the equations of motion enables us to construct the mechanics of monogenic systems from Hamilton's principle as the basic postulate rather than Newton's laws of motion. Such a formulation has advantages; e.g., since the integral I is obviously invariant to the system of generalized coordinates used to express L , the equations of motion must always have the Lagrangian form no matter how the generalized coordinates

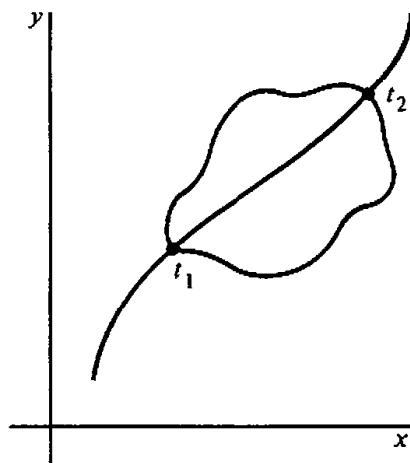


FIGURE 2.1 Path of the system point in configuration space.

are transformed. More important, the formulation in terms of a variational principle is the route that is generally followed when we try to describe apparently nonmechanical systems in the mathematical clothes of classical mechanics, as in the theory of fields.

2.2 ■ SOME TECHNIQUES OF THE CALCULUS OF VARIATIONS

Before demonstrating that Lagrange's equations do follow from (2.2), we must first examine the methods of the calculus of variations, for a chief problem of this calculus is to find the curve for which some given line integral has a stationary value.

Consider first the problem in an essentially one-dimensional form: We have a function $f(y, \dot{y}, x)$ defined on a path $y = y(x)$ between two values x_1 and x_2 , where \dot{y} is the derivative of y with respect to x . We wish to find a particular path $y(x)$ such that the line integral J of the function f between x_1 and x_2 ,

$$\begin{aligned}\dot{y} &\equiv \frac{dy}{dx}, \\ J &= \int_{x_1}^{x_2} f(y, \dot{y}, x) dx,\end{aligned}\quad (2.3)$$

has a stationary value relative to paths differing infinitesimally from the correct function $y(x)$. The variable x here plays the role of the parameter t , and we consider only such varied paths for which $y(x_1) = y_1$, $y(x_2) = y_2$. (Cf. Fig. 2.2.) Note that Fig. 2.2 does *not* represent configuration space. In the one-dimensional configuration space, both the correct and varied paths are the segment of the straight line connecting y_1 and y_2 ; the paths differ only in the functional relation between y and x . The problem is one-dimensional, y is a function of x not a coordinate.

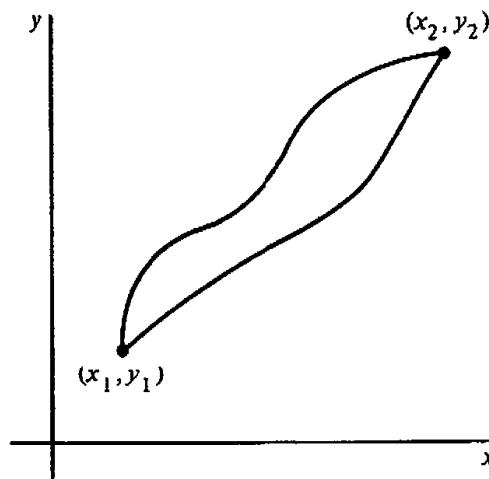


FIGURE 2.2 Varied paths of the function of $y(x)$ in the one-dimensional extremum problem.

We put the problem in a form that enables us to use the familiar apparatus of the differential calculus for finding the stationary points of a function. Since J must have a stationary value for the correct path relative to *any* neighboring path, the variation must be zero relative to *some* particular set of neighboring paths labeled by an infinitesimal parameter α . Such a set of paths might be denoted by $y(x, \alpha)$, with $y(x, 0)$ representing the correct path. For example, if we select any function $\eta(x)$ that vanishes at $x = x_1$ and $x = x_2$, then a possible set of varied paths is given by

$$y(x, \alpha) = y(x, 0) + \alpha\eta(x). \quad (2.4)$$

For simplicity, it is assumed that both the correct path $y(x)$ and the auxiliary function $\eta(x)$ are well-behaved functions—continuous and nonsingular between x_1 and x_2 , with continuous first and second derivatives in the same interval. For any such parametric family of curves, J in Eq. (2.3) is also a function of α :

$$J(\alpha) = \int_{x_1}^{x_2} f(y(x, \alpha), \dot{y}(x, \alpha), x) dx. \quad (2.5)$$

and the condition for obtaining a stationary point is the familiar one that

$$\left(\frac{dJ}{d\alpha} \right)_{\alpha=0} = 0. \quad (2.6)$$

By the usual methods of differentiating under the integral sign, we find that

$$\frac{dJ}{d\alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} \frac{\partial y}{\partial \alpha} + \frac{\partial f}{\partial \dot{y}} \frac{\partial \dot{y}}{\partial \alpha} \right) dx. \quad (2.7)$$

Consider the second of these integrals.

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial \dot{y}} \frac{\partial \dot{y}}{\partial \alpha} dx = \int_{x_1}^{x_2} \frac{\partial f}{\partial \dot{y}} \frac{\partial^2 y}{\partial x \partial \alpha} dx.$$

Integrating by parts, the integral becomes

$$\int_{x_1}^{x_2} \frac{\partial f}{\partial \dot{y}} \frac{\partial^2 y}{\partial x \partial \alpha} dx = \frac{\partial f}{\partial \dot{y}} \frac{\partial y}{\partial \alpha} \Big|_{x_1}^{x_2} - \int_{x_1}^{x_2} \frac{d}{dx} \left(\frac{\partial f}{\partial \dot{y}} \right) \frac{\partial y}{\partial \alpha} dx. \quad (2.8)$$

The conditions on all the varied curves are that they pass through the points (x_1, y_1) , (x_2, y_2) , and hence the partial derivative of y with respect to α at x_1 and x_2 must vanish. Therefore, the first term of (2.8) vanishes and Eq. (2.7) reduces to

$$\frac{dJ}{d\alpha} = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial \dot{y}} \right) \frac{\partial y}{\partial \alpha} dx.$$

The condition for a stationary value, Eq. (2.6), is therefore equivalent to the equation

$$\int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial \dot{y}} \right) \left(\frac{\partial y}{\partial \alpha} \right)_0 dx = 0. \quad (2.9)$$

Now, the partial derivative of y with respect to α occurring in Eq. (2.9) is a function of x that is arbitrary except for continuity and end point conditions. For example, for the particular parametric family of varied paths given by Eq. (2.4), it is the arbitrary function $\eta(x)$. We can therefore apply to Eq (2.9) the so-called “fundamental lemma” of the calculus of variations, which says if

$$\int_{x_1}^{x_2} M(x)\eta(x) dx = 0 \quad (2.10)$$

for all arbitrary functions $\eta(x)$ continuous through the second derivative, then $M(x)$ must identically vanish in the interval (x_1, x_2) . While a formal mathematical proof of the lemma can be found in texts on the calculus of variations, the validity of the lemma is easily seen intuitively. We can imagine constructing a function η that is positive in the immediate vicinity of any chosen point in the interval and zero everywhere else. Equation (2.10) can then hold only if $M(x)$ vanishes at that (arbitrarily) chosen point, which shows M must be zero throughout the interval. From Eq. (2.9) and the fundamental lemma, it therefore follows that J can have a stationary value only if

$$\frac{\partial f}{\partial y} - \frac{d}{dx} \left(\frac{\partial f}{\partial \dot{y}} \right) = 0. \quad (2.11)$$

The differential quantity,

$$\left(\frac{\partial y}{\partial \alpha} \right)_0 d\alpha \equiv \delta y, \quad (2.12)$$

represents the infinitesimal departure of the varied path from the correct path $y(x)$ at the point x and thus corresponds to the virtual displacement introduced in Chapter 1 (hence the notation δy). Similarly, the infinitesimal variation of J about the correct path can be designated

$$\left(\frac{dJ}{d\alpha} \right)_0 d\alpha \equiv \delta J. \quad (2.13)$$

The assertion that J is stationary for the correct path can thus be written

$$\delta J = \int_{x_1}^{x_2} \left(\frac{\partial f}{\partial y} - \frac{d}{dx} \frac{\partial f}{\partial \dot{y}} \right) \delta y dx = 0,$$

requiring that $y(x)$ satisfy the differential equation (2.11). The δ -notation, introduced through Eqs. (2.12) and (2.13), may be used as a convenient shorthand for treating the variation of integrals, remembering always that it stands for the manipulation of parametric families of varied paths such as Eq. (2.4).

Some simple examples of the application of Eq. (2.11) (which clearly resembles a Lagrange equation) may now be considered:

1. *Shortest distance between two points in a plane.* An element of length in a plane is

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

and the total length of any curve going between points 1 and 2 is

$$I = \int_1^2 ds = \int_{x_1}^{x_2} \sqrt{1 + \left(\frac{dy}{dx}\right)^2} dx.$$

The condition that the curve be the shortest path is that I be a minimum. This is an example of the extremum problem as expressed by Eq. (2.3), with

$$f = \sqrt{1 + \dot{y}^2}.$$

Substituting in (2.11) with

$$\frac{\partial f}{\partial y} = 0, \quad \frac{\partial f}{\partial \dot{y}} = \frac{\dot{y}}{\sqrt{1 + \dot{y}^2}},$$

we have

$$\frac{d}{dx} \left(\frac{\dot{y}}{\sqrt{1 + \dot{y}^2}} \right) = 0$$

or

$$\frac{\dot{y}}{\sqrt{1 + \dot{y}^2}} = c,$$

where c is constant. This solution can be valid only if

$$\dot{y} = a,$$

where a is a constant related to c by

$$a = \frac{c}{\sqrt{1 - c^2}}.$$

But this is clearly the equation of a straight line,

$$y = ax + b,$$

where b is another constant of integration. Strictly speaking, the straight line has only been proved to be an extremum path, but for this problem it is obviously also a minimum. The constants of integration, a and b , are determined by the condition that the curve pass through the two end points, (x_1, y_1) , (x_2, y_2) .

In a similar fashion we can obtain the shortest distance between two points on a sphere, by setting up the arc length on the surface of the sphere in terms of the angle coordinates of position on the sphere. In general, curves that give the shortest distance between two points on a given surface are called the *geodesics* of the surface.

2. *Minimum surface of revolution.* Suppose we form a surface of revolution by taking some curve passing between two fixed end points (x_1, y_1) and (x_2, y_2) defining the xy plane, and revolving it about the y axis (cf. Fig. 2.3a). The problem then is to find that curve for which the surface area is a minimum. The area of a strip of the surface is $2\pi x \, ds = 2\pi x \sqrt{1 + \dot{y}^2} \, dx$, and the total area is

$$2\pi \int_1^2 x \sqrt{1 + \dot{y}^2} \, dx.$$

The extremum of this integral is again given by (2.11) where

$$f = x \sqrt{1 + \dot{y}^2}$$

and

$$\frac{\partial f}{\partial y} = 0, \quad \frac{\partial f}{\partial \dot{y}} = \frac{x \dot{y}}{\sqrt{1 + \dot{y}^2}}.$$

Equation (2.11) becomes in this case

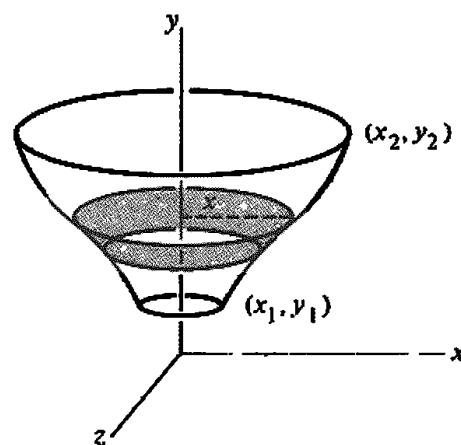


FIGURE 2.3a Minimum surface of revolution. Note that this figure is drawn for y_1 and y_2 having the same sign relative to the rotation axis. This is not assumed in the general solution.

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

or

$$\frac{x\dot{y}}{\sqrt{1+\dot{y}^2}} = a,$$

where a is some constant of integration clearly smaller than the minimum value of x . Squaring the above equation and factoring terms, we have

$$\dot{y}^2(x^2 - a^2) = a^2,$$

or solving,

$$\frac{dy}{dx} = \frac{a}{\sqrt{x^2 - a^2}}.$$

The general solution of this differential equation, in light of the nature of a , is

$$y = a \int \frac{dx}{\sqrt{x^2 - a^2}} + b = a \operatorname{arcosh} \frac{x}{a} + b$$

or

$$x = a \cosh \frac{y - b}{a},$$

which is the equation of a catenary. Again the two constants of integration, a and b , are determined in principle by the requirements that the curve pass through the two given end points, as shown in Fig. 2.3b.

Curves satisfying the preceding equation all scale as x/a and y/a with one independent parameter b/a . This suggests that when the solutions are examined in detail they turn out to be a great deal more complicated than these considera-

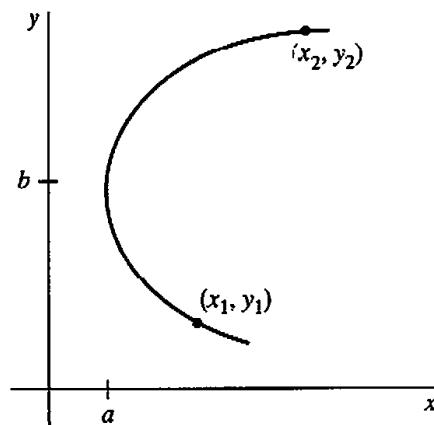


FIGURE 2.3b General catenary solution for minimum surface of revolution.

tions suggest. For some pairs of end points, unique constants of integration a and b can be found. But for other end points, it is possible to draw two different catenary curves through the end points, while for additional cases no possible values can be found for a and b . Further, recall that Eq. (2.11) represents a condition for finding curves $y(x)$ continuous through the second derivative that render the integral stationary. The catenary solutions therefore do not always represent minimum values, but may represent "inflection points" where the length of the curve is stationary but not minimum.

For certain combinations of end points (an example is x_1 and x_2 both positive and both much smaller than $y_2 - y_1$), the absolute minimum in the surface of revolution is provided (cf. Exercise 8) by a curve composed of straight line segments—from the first end point parallel to the x axis until the y axis is reached, then along the y axis until the point $(0, y_2)$ and then out in a straight line to the second end point corresponding to the area $\pi(x_1^2 + x_2^2)$. This curve results when $a = 0$, forcing either $x = 0$ or $y = \text{constant}$. Since this curve has discontinuous first derivatives, we should not expect to find it as a solution to Eq. (2.11).

This example is valuable in emphasizing the restrictions that surround the derivation and the meaning of the stationary condition. Exercises 7 and 8 examine the conditions for the pathological behavior for a symmetric example. More information can be found in many texts on the calculus of variations.

3. *The brachistochrone problem.* (See Fig. 2.4a.) This well-known problem is to find the curve joining two points, along which a particle falling from rest under the influence of gravity travels from the higher to the lower point in the least time.

If v is the speed along the curve, then the time required to fall an arc length ds is ds/v , and the problem is to find a minimum of the integral

$$t_{12} = \int^2 \frac{ds}{v}.$$

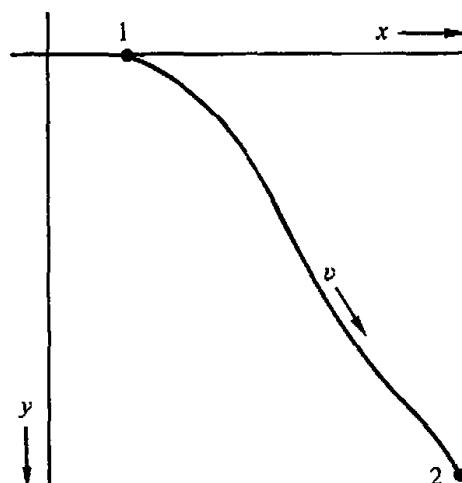


FIGURE 2.4a The brachistochrone problem.

If y is measured down from the initial point of release, the conservation theorem for the energy of the particle can be written as

$$\frac{1}{2}mv^2 = mgy$$

or

$$v = \sqrt{2gy}.$$

Then the expression for t_{12} becomes

$$t_{12} = \int_1^2 \frac{\sqrt{1 + \dot{y}^2}}{\sqrt{2gy}} dx,$$

and f is identified as

$$f = \sqrt{\frac{1 + \dot{y}^2}{2gy}}.$$

The integration of Eq. (2.11) with this form for f is straightforward and is left as an exercise.

The solution in terms of its one parameter, a , given by

$$\frac{y}{a} = 1 - \cos \left[\frac{x + \sqrt{y(2a - y)}}{a} \right],$$

is sketched in Fig. 2.4b for the first cycle ($0 \leq x \leq 2\pi a$) and the beginning of the second cycle. Three cases of solutions are indicated. A power-series expansion of the solution for the limit $y \ll a$ gives

$$y = \frac{x^2}{2}a.$$

The brachistochrone problem is famous in the history of mathematics, for it was the analysis of this problem by John Bernoulli that led to the formal foundation of the calculus of variations.

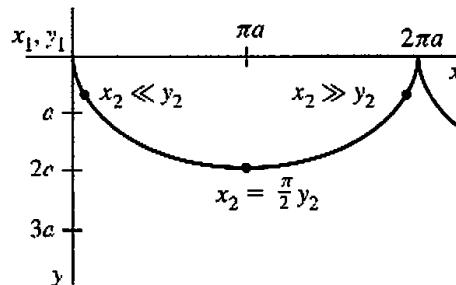


FIGURE 2.4b Catenary solution to the brachistochrone problem showing positions on the curve for the three cases $x_2 \ll y_2$, $x_2 = \frac{\pi}{2}y_2$, and $x_2 \gg y_2$

2.3 ■ DERIVATION OF LAGRANGE'S EQUATIONS FROM HAMILTON'S PRINCIPLE

The fundamental problem of the calculus of variations is easily generalized to the case where f is a function of many independent variables y_i , and their derivatives \dot{y}_i . (Of course, all these quantities are considered as functions of the parametric variable x .) Then a variation of the integral J ,

$$\delta J = \delta \int_1^2 f(y_1(x); y_2(x), \dots, \dot{y}_1(x); \dot{y}_2(x), \dots, x) dx, \quad (2.14)$$

is obtained, as before, by considering J as a function of parameter α that labels a possible set of curves $y_i(x, \alpha)$. Thus, we may introduce α by setting

$$\begin{aligned} y_1(x, \alpha) &= y_1(x, 0) + \alpha \eta_1(x), \\ y_2(x, \alpha) &= y_2(x, 0) + \alpha \eta_2(x), \\ &\vdots \qquad \vdots \qquad \vdots \end{aligned} \quad (2.15)$$

where $y_1(x, 0)$, $y_2(x, 0)$, etc., are the solutions of the extremum problem (to be obtained) and η_1 , η_2 , etc., are independent functions of x that vanish at the end points and that are continuous through the second derivative, but otherwise are completely arbitrary.

The calculation proceeds as before. The variation of J is given in terms of

$$\frac{\partial J}{\partial \alpha} d\alpha = \int_1^2 \sum_i \left(\frac{\partial f}{\partial y_i} \frac{\partial y_i}{\partial \alpha} d\alpha + \frac{\partial f}{\partial \dot{y}_i} \frac{\partial \dot{y}_i}{\partial \alpha} d\alpha \right) dx. \quad (2.16)$$

Again we integrate by parts the integral involved in the second sum of Eq. (2.16):

$$\int_1^2 \frac{\partial f}{\partial \dot{y}_i} \frac{\partial^2 y_i}{\partial \alpha \partial x} dx = \left. \frac{\partial f}{\partial \dot{y}_i} \frac{\partial y_i}{\partial \alpha} \right|_1^2 - \int_1^2 \frac{\partial y_i}{\partial \alpha} \frac{d}{dx} \left(\frac{\partial f}{\partial \dot{y}_i} \right) dx,$$

where the first term vanishes because all curves pass through the fixed end points. Substituting in (2.16), δJ becomes

$$\delta J = \int_1^2 \sum_i \left(\frac{\partial f}{\partial y_i} - \frac{d}{dx} \frac{\partial f}{\partial \dot{y}_i} \right) \delta y_i dx, \quad (2.17)$$

where, in analogy with (2.12), the variation δy_i is

$$\delta y_i = \left(\frac{\partial y_i}{\partial \alpha} \right)_0 d\alpha.$$

Since the y variables are independent, the variations δy_i are independent (e.g., the functions $\eta_i(x)$ will be independent of each other). Hence, by an obvious extension of the fundamental lemma (cf. Eq. (2.10)), the condition that δJ is zero

requires that the coefficients of the δy_i separately vanish:

$$\frac{\partial f}{\partial y_i} - \frac{d}{dx} \frac{\partial f}{\partial \dot{y}_i} = 0, \quad i = 1, 2, \dots, n. \quad (2.18)$$

Equations (2.18) represent the appropriate generalization of (2.11) to several variables and are known as the *Euler-Lagrange differential equations*. Their solutions represent curves for which the variation of an integral of the form given in (2.14) vanishes. Further generalizations of the fundamental variational problem are easily possible. Thus, we can take f as a function of higher derivatives \ddot{y} , \dot{y} , etc., leading to equations different from (2.18). Or we can extend it to cases where there are several parameters x_j and the integral is then multiple, with f also involving as variables derivatives of y_i with respect to each of the parameters x_j . Finally, it is possible to consider variations in which the end points are *not* held fixed.

For present purposes, what we have derived here suffices, for the integral in Hamilton's principle,

$$I = \int_1^2 L(q_i, \dot{q}_i, t) dt, \quad (2.19)$$

has just the form stipulated in (2.14) with the transformation

$$x \rightarrow t$$

$$y_i \rightarrow q_i$$

$$f(y_i, \dot{y}_i, x) \rightarrow L(q_i, \dot{q}_i, t).$$

In deriving Eqs. (2.18), we assumed that the y_i variables are independent. The corresponding condition in connection with Hamilton's principle is that the generalized coordinates q_i be independent, which requires that the constraints be holonomic. The Euler-Lagrange equations corresponding to the integral I then become the Lagrange equations of motion,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_i} - \frac{\partial L}{\partial q_i} = 0, \quad i = 1, 2, \dots, n,$$

and we have accomplished our original aim, to show that Lagrange's equations follow from Hamilton's principle—for monogenic systems with holonomic constraints.

2.4 ■ EXTENSION OF HAMILTON'S PRINCIPLE TO NONHOLONOMIC SYSTEMS

It is possible to extend Hamilton's principle, at least in a formal sense, to cover certain types of nonholonomic systems. In deriving Lagrange's equations from

either Hamilton's or D'Alembert's principle, the requirement of holonomic constraints does not appear until the last step, when the variations q_i are considered as independent of each other. With nonholonomic systems the generalized coordinates are not independent of each other, and it is not possible to reduce them further by means of equations of constraint of the form $f(q_1, q_2, \dots, q_n, t) = 0$. Hence, it is no longer true that the q_i 's are all independent.

Another difference that must be considered in treating the variational principle is the manner in which the varied paths are constructed. In the discussion of Section 2.2, we pointed out that δy (or δq) represents a virtual displacement from a point on the actual path to some point on the neighboring varied path. But, with independent coordinates it is the final varied path that is significant, not how it is constructed. When the coordinates are not independent, but subject to constraint relations, it becomes important whether the varied path is or is not constructed by displacements consistent with the constraints. Virtual displacements, in particular, may or may not satisfy the constraints.

It appears that a reasonably straightforward treatment of nonholonomic systems by a variational principle is possible only when the equations of constraint can be put in the form

$$f_\alpha(q_1, \dots, q_n; \dot{q}_1, \dots, \dot{q}_n) = 0, \quad (2.20)$$

when this can be done the constraints are called semi-holonomic. The index α indicates that there may be more than one such equation. We will assume there are m equations in all, i.e., $\alpha = 1, 2, \dots, m$. Equation (2.20) commonly appears in the restricted form

$$\sum_k a_{ik} dq_k + a_{it} dt = 0. \quad (2.20')$$

We might expect that the varied paths, or equivalently, the displacements constructing the varied path, should satisfy the constraints of Eq. (2.20). However, it has been proven that no such varied path can be constructed unless Eqs. (2.20) are integrable, in which case the constraints are actually holonomic. A variational principle leading to the correct equations of motion can nonetheless be obtained when the varied paths are constructed from the actual motion by virtual displacements.

The procedure for eliminating these extra virtual displacements is the method of *Lagrange undetermined multipliers*. If Eqs. (2.20) hold, then it is also true that

$$\sum_{\alpha=1}^m \lambda_\alpha f_\alpha = 0, \quad (2.21)$$

where the λ_α , $\alpha = 1, 2, \dots, m$, are some undetermined quantities, functions in general of the coordinates and of the time t . In addition, Hamilton's principle,

$$\delta \int_{t_1}^{t_2} L dt = 0, \quad (2.2)$$

is assumed to hold for this semiholonomic system. Following the development of Section 2.3, Hamilton's principle then implies that

$$\int_1^2 dt \sum_k \left(\frac{\partial L}{\partial q_k} - \frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} \right) \delta q_k = 0. \quad (2.22)$$

The variation cannot be taken as before since the q_k are not independent; however, combining (2.21) with (2.2) gives

$$\delta \int_{t_1}^{t_2} \left(L + \sum_{\alpha=1}^m \lambda_{\alpha} f_{\alpha} \right) dt = 0 \quad (2.23)$$

The variation can now be performed with the $n \delta q_i$ and $m \lambda_{\alpha}$ for $m+n$ independent variables. For the simplifying assumption that $\lambda_{\alpha} = \lambda_{\alpha}(t)$, the resulting equations from δq_i become*

$$\frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_k} \right) - \frac{\partial L}{\partial q_k} = Q_k, \quad (2.24)$$

where

$$Q_k = \sum_{\alpha=1}^m \left\{ \lambda_{\alpha} \left[\frac{\partial f_{\alpha}}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial f_{\alpha}}{\partial \dot{q}_k} \right) \right] - \frac{d \lambda_{\alpha}}{dt} \frac{\partial f_{\alpha}}{\partial \dot{q}_k} \right\}, \quad (2.25)$$

while the $\delta \lambda_{\alpha}$ give the equations of constraint (2.20). Equations (2.24) and (2.20) together constitute $n+m$ equations for $n+m$ unknowns. The system can now be interpreted as an $m+n$ holonomic system with generalized forces Q_k . The generalization to $\lambda_{\alpha} = \lambda_{\alpha}(q_1, \dots, q_n; \dot{q}_1, \dots, \dot{q}_n; t)$ is straightforward.

As an example, let us consider a particle whose Lagrangian is

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - V(x, y, z) \quad (2.26)$$

subject to the constraint

$$f(\dot{x}, \dot{y}, y) = \dot{x}\dot{y} + ky = 0 \quad (2.27)$$

with k a constant. The resulting equations of motion are

$$m\ddot{x} + \lambda\ddot{y} + \dot{\lambda}\dot{y} + \frac{\partial V}{\partial x} = 0, \quad (2.28)$$

$$m\ddot{y} + \lambda\ddot{x} - k\lambda + \dot{\lambda}\dot{x} + \frac{\partial V}{\partial y} = 0, \quad (2.29)$$

$$m\ddot{z} + \frac{\partial V}{\partial z} = 0, \quad (2.30)$$

*J. Ray, *Amer. J. Phys.* **34** (406–8), 1996.

and the equation of constraint, (2.20), becomes

$$\dot{y}\dot{x} + ky = 0.$$

In this process we have obtained more information than was originally sought. Not only do we get the q_k 's we set out to find, but we also get $m\lambda_l$'s. What is the physical significance of the λ_l 's? Suppose we remove the constraints on the system, but instead apply external forces Q'_k in such a manner as to keep the motion of the system unchanged. The equations of motion likewise remain the same. Clearly these extra applied forces must be equal to the forces of constraint, for they are the forces applied to the system so as to satisfy the condition of constraint. Under the influence of these forces Q'_k , the equations of motion are

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = Q'_k. \quad (2.31)$$

But these must be identical with Eqs. (2.24). Hence, we can identify (2.25) with Q'_k , the generalized forces of constraint. In this type of problem we really do not eliminate the forces of constraint from the formulation. They are supplied as part of the answer.

Although it is not obvious, the version of Hamilton's principle adopted here for semiholonomic systems also requires that the constraints do no work in virtual displacements. This can be most easily seen by rewriting Hamilton's principle in the form

$$\delta \int_{t_1}^{t_2} L dt = \delta \int_{t_1}^{t_2} T dt - \delta \int_{t_1}^{t_2} U dt = 0. \quad (2.32)$$

If the variation of the integral over the generalized potential is carried out by the procedures of Section 2.3, the principle takes the form

$$\delta \int_{t_1}^{t_2} T dt = \int_{t_1}^{t_2} \sum_k \left[\frac{\partial U}{\partial q_k} - \frac{d}{dt} \left(\frac{\partial U}{\partial \dot{q}_k} \right) \right] \delta q_k dt; \quad (2.33)$$

or, by Eq. (1.58),

$$\delta \int_{t_1}^{t_2} T dt = - \int_{t_1}^{t_2} \sum_k Q_k \delta q_k dt. \quad (2.34)$$

In this dress, Hamilton's principle says that the difference in the time integral of the kinetic energy between two neighboring paths is equal to the negative of the time integral of the work done in the virtual displacements between the paths. The work involved is that done only by the forces derivable from the generalized potential. The same Hamilton's principle holds for both holonomic and semiholonomic systems, it must be required that the additional forces of semiholonomic constraints do no work in the displacements δq_k . This restriction parallels the earlier condition that the virtual work of the forces of holonomic constraint also be

zero (cf. Section 1.4). In practice, the restriction presents little handicap to the applications, as many problems in which the semiholonomic formalism is used relate to rolling without slipping, where the constraints are obviously workless.

Note that Eq. (2.20) is not the most general type of nonholonomic constraint; e.g., it does not include equations of constraint in the form of inequalities. On the other hand, it does include holonomic constraints. A holonomic equation of constraint,

$$f(q_1, q_2, q_3, \dots, q_n, t) = 0, \quad (2.35)$$

is equivalent to (2.20) with no dependence on \dot{q}_k . Thus, the Lagrange multiplier method can be used also for holonomic constraints when (1) it is inconvenient to reduce all the q 's to independent coordinates or (2) we might wish to obtain the forces of constraint.

As another example of the method, let us consider the following somewhat trivial illustration—a hoop rolling, without slipping, down an inclined plane. In this example, the constraint of “rolling” is actually holonomic, but this fact will be immaterial to our discussion. On the other hand, the holonomic constraint that the hoop be on the inclined plane will be contained implicitly in our choice of generalized coordinates.

The two generalized coordinates are x, θ , as in Fig. 2.5, and the equation of rolling constraint is

$$r d\theta = dx.$$

The kinetic energy can be resolved into kinetic energy of motion of the center of mass plus the kinetic energy of motion about the center of mass:

$$T = \frac{1}{2}M\dot{x}^2 + \frac{1}{2}Mr^2\dot{\theta}^2.$$

The potential energy is

$$V = Mg(l - x) \sin \phi,$$

where l is the length of the inclined plane and the Lagrangian is

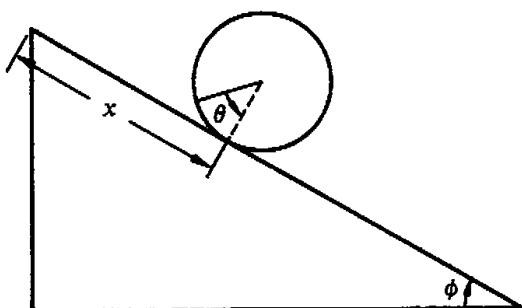


FIGURE 2.5 A hoop rolling down an inclined plane.

$$\begin{aligned} L &= T - V \\ &= \frac{M\dot{x}^2}{2} + \frac{Mr^2\dot{\theta}^2}{2} - Mg(l-x)\sin\phi. \end{aligned} \quad (2.36)$$

Since there is one equation of constraint, only one Lagrange multiplier λ is needed. The coefficients appearing in the constraint equation are:

$$a_\theta = r,$$

$$a_x = -1.$$

The two Lagrange equations therefore are

$$M\ddot{x} - Mg\sin\phi + \lambda = 0, \quad (2.37)$$

$$Mr^2\ddot{\theta} - \lambda r = 0, \quad (2.38)$$

which along with the equation of constraint,

$$r\dot{\theta} = \dot{x}, \quad (2.39)$$

constitutes three equations for three unknowns, θ , x , λ .

Differentiating (2.39) with respect to time, we have

$$r\ddot{\theta} = \ddot{x}.$$

Hence, from (2.38)

$$M\ddot{x} = \lambda,$$

and (2.37) becomes

$$\ddot{x} = \frac{g\sin\phi}{2}.$$

along with

$$\lambda = \frac{Mg\sin\phi}{2}$$

and

$$\ddot{\theta} = \frac{g\sin\phi}{2r}.$$

Thus, the hoop rolls down the incline with only one-half the acceleration it would have slipping down a frictionless plane, and the friction force of constraint is $\lambda = Mg\sin\phi/2$.

2.5 ■ ADVANTAGES OF A VARIATIONAL PRINCIPLE FORMULATION

Although we can extend the original formulation of Hamilton's principle (2.2) to include some nonholonomic constraints, in practice this formulation of mechanics is most useful when a Lagrangian of independent coordinates can be set up for the system. The variational principle formulation has been justly described as "elegant," for in the compact Hamilton's principle is contained all of the mechanics of holonomic systems with forces derivable from potentials. The principle has the further merit that it involves only physical quantities that can be defined without reference to a particular set of generalized coordinates, namely, the kinetic and potential energies. The formulation is therefore automatically invariant with respect to the choice of coordinates for the system.

From the variational Hamilton's principle, it is also obvious why the Lagrangian is always uncertain to a total time derivative of any function of the coordinates and time, as mentioned at the end of Section 1.4. The time integral of such a total derivative between points 1 and 2 depends only on the values of the arbitrary function at the end points. As the variation at the end points is zero, the addition of the arbitrary time derivative to the Lagrangian does not affect the variational behavior of the integral.

Another advantage is that the Lagrangian formulation can be easily extended to describe systems that are not normally considered in dynamics—such as the elastic field, the electromagnetic field, and field properties of elementary particles. Some of these generalizations will be considered later, but as three simple examples of its application outside the usual framework of mechanics, let us consider the cases of an RL circuit, an LC circuit, and coupled circuits.

We consider the physical system of a battery of voltage V in series with an inductance L and a resistance of value R and choose the electric charge q as the dynamical variable. The inductor acts as the kinetic energy term since the inductive effect depends upon the time rate of change of the charge. The resistor provides a dissipative term and the potential energy is qV . The dynamic terms in Lagrange's equation with dissipation (1.70) are

$$T = \frac{1}{2}L\dot{q}^2, \quad \mathcal{F} = \frac{1}{2}R\dot{q}^2,$$

and potential energy $= qV$. The equation of motion is

$$V = L\ddot{q} + R\dot{q} = LI + RI. \quad (2.40)$$

where $I = \dot{q}$ is the electric current. A solution for a battery connected to the circuit at time $t = 0$ is

$$I = I_0(1 - e^{-Rt/L}),$$

where $I_0 = V/R$ is the final steady-state current flow.

The mechanical analog for this is a sphere of radius a and effective mass m' falling in a viscous fluid of constant density and viscosity η under the force of

gravity. The effective mass is the difference between the actual mass and the mass of the displaced fluid, and the direction of motion is along the y axis. For this system,

$$T = \frac{1}{2}m'\dot{y}^2, \quad \mathcal{F} = 3\pi\eta a\dot{y}^2,$$

and potential energy $= m'gy$, where the frictional drag force, $F_f = 6\pi\eta ay$, called Stokes' law, was given at the end of Section 1.5.

The equation of motion is given by Lagrange's equations (1.70) as

$$m'g = m'\ddot{y} + 6\pi\eta a\dot{y}.$$

Using $v = \dot{y}$, the solution (if the motion starts from rest at $t = 0$), is

$$v = v_0(1 - e^{-t/\tau})$$

where $\tau = m'/(6\pi\eta a)$ is a measure of the time it takes for the sphere to reach $1/e$ of its terminal speed of $v_0 = m'g/6\pi\eta a$.

Another example from electrical circuits is an inductance, L , in series with a capacitance, C . The capacitor acts as a source of potential energy given by q^2/C where q is the electric charge. The Lagrangian produces the equation of motion,

$$L\ddot{q} + \frac{q}{C} = 0, \quad (2.41)$$

which has the solution

$$q = q_0 \cos \omega_0 t,$$

where q_0 is the charge stored in the capacitor at $t = 0$, and the assumption is that no charge is flowing at $t = 0$. The quantity

$$\omega_0 = \frac{1}{\sqrt{LC}}$$

is the resonant frequency of the system.

The mechanical analog of this system is the simple harmonic oscillator described by the Lagrangian $L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$, which gives an equation of motion,

$$m\ddot{x} + kx = 0,$$

whose solution for the same boundary conditions is

$$x = x_0 \cos \omega_0 t \quad \text{with} \quad \omega_0 = \sqrt{k/m}.$$

These two examples show that an inductance is an inertial term, the electrical analog of mass. Resistance is the analog of Stokes' law type of frictional drag, and the capacitance term $1/C$ represents a Hooke's law spring constant. With this

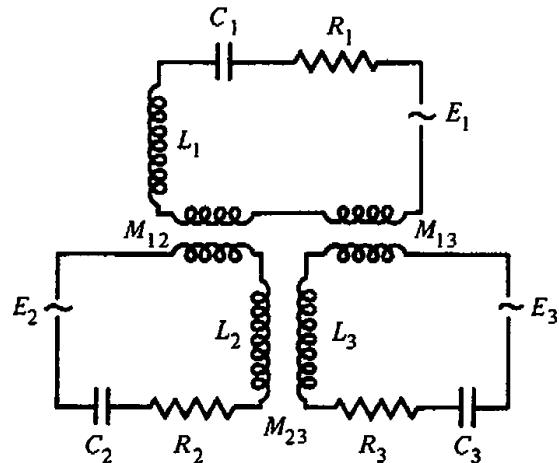


FIGURE 2.6 A system of coupled circuits to which the Lagrangian formulation can be applied.

background, a system of coupled electrical circuits of the type shown in Fig. 2.6 has a Lagrangian of the form

$$L = \frac{1}{2} \sum_j L_j \dot{q}_j^2 + \frac{1}{2} \sum_{\substack{j,k \\ j \neq k}} M_{jk} \dot{q}_j \dot{q}_k - \sum_j \frac{q_j^2}{2C_j} + \sum_j e_j(t) q_j,$$

and a dissipation function

$$\mathcal{F} = \frac{1}{2} \sum_j R_j \dot{q}_j^2.$$

where the mutual inductance terms, $M_{jk} \dot{q}_j \dot{q}_k$, are added to take into account the coupling between inductors. The Lagrange equations are

$$L_j \frac{d^2 q_j}{dt^2} + \sum_k M_{jk} \frac{d^2 q_k}{dt^2} + R_j \frac{dq_j}{dt} + \frac{q_j}{C_j} = E_j(t). \quad (2.42)$$

where the $E_j(t)$ terms are the external emf's.

This description of two different physical systems by Lagrangians of the same form means that all the results and techniques devised for investigating one of the systems can be taken over immediately and applied to the other. In this particular case, the study of the behavior of electrical circuits has been pursued intensely and some special techniques have been developed; these can be directly applied to the corresponding mechanical systems. Much work has been done in formulating equivalent electrical problems for mechanical or acoustical systems, and vice versa. Terms hitherto reserved for electrical circuits (reactance, susceptance, etc.) are now commonly found in treatises on the theory of vibrations of mechanical systems.

Additionally, one type of generalization of mechanics is due to a subtler form of equivalence. We have seen that the Lagrangian and Hamilton's principle together form a compact invariant way of obtaining the mechanical equations of motion. This possibility is not reserved for mechanics only; in almost every field of physics variational principles can be used to express the "equations of motion," whether they be Newton's equations, Maxwell's equations, or the Schrödinger equation. Consequently, when a variational principle is used as the basis of the formulation, all such fields will exhibit, at least to some degree, a *structural analogy*. When the results of experiments show the need for altering the physical content in the theory of one field, this degree of analogy has often indicated how similar alterations may be carried out in other fields. Thus, the experiments performed early in this century showed the need for quantization of both electromagnetic radiation and elementary particles. The methods of quantization, however, were first developed for particle mechanics, starting essentially from the Lagrangian formulation of classical mechanics. By describing the electromagnetic field by a Lagrangian and corresponding Hamilton's variational principle, it is possible to carry over the methods of particle quantization to construct a quantum electrodynamics (cf. Sections 13.5 and 13.6).

2.6 ■ CONSERVATION THEOREMS AND SYMMETRY PROPERTIES

Thus far, we have been concerned primarily with obtaining the equations of motion, but little has been said about how to solve them for a particular problem once they are obtained. In general, this is a question of mathematics. A system of n degrees of freedom will have n differential equations that are second order in time. The solution of each equation will require two integrations resulting, all told, in $2n$ constants of integration. In a specific problem these constants will be determined by the initial conditions, i.e., the initial values of the nq_j 's and the $n\dot{q}_j$'s. Sometimes the equations of motion will be integrable in terms of known functions, but not always. In fact, the majority of problems are not completely integrable. However, even when complete solutions cannot be obtained, it is often possible to extract a large amount of information about the physical nature of the system motion. Indeed, such information may be of greater interest to the physicist than the complete solution for the generalized coordinates as a function of time. It is important, therefore, to see how much can be stated about the motion of a given system without requiring a complete integration of the problem.*

In many problems a number of first integrals of the equations of motion can be obtained immediately; by this we mean relations of the type

$$f(q_1, q_2, \dots, \dot{q}_1, \dot{q}_2, \dots, t) = \text{constant}. \quad (2.43)$$

*In this and succeeding sections it will be assumed, unless otherwise specified, the system is such that its motion is completely described by a Hamilton's principle of the form (2.2).

which are first-order differential equations. These first integrals are of interest because they tell us something physically about the system. They include, in fact, the conservation laws obtained in Chapter 1.

Let us consider as an example a system of mass points under the influence of forces derived from potentials dependent on position only. Then

$$\begin{aligned}\frac{\partial L}{\partial \dot{x}_i} &\equiv \frac{\partial T}{\partial \dot{x}_i} - \frac{\partial V}{\partial \dot{x}_i} = \frac{\partial T}{\partial \dot{x}_i} = \frac{\partial}{\partial \dot{x}_i} \sum \frac{1}{2} m_i (\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2) \\ &= m_i \dot{x}_i = p_{ix}.\end{aligned}$$

which is the x component of the linear momentum associated with the i th particle. This result suggests an obvious extension to the concept of momentum. The generalized momentum associated with the coordinate q_j shall be defined as

$$p_j = \frac{\partial L}{\partial \dot{q}_j}. \quad (2.44)$$

The terms *canonical momentum* and *conjugate momentum* are often also used for p_j . Notice that if q_j is not a Cartesian coordinate, p_j does not necessarily have the dimensions of a linear momentum. Further, if there is a velocity-dependent potential, then even with a Cartesian coordinate q_j , the associated *generalized* momentum will not be identical with the usual *mechanical* momentum. Thus, in the case of a group of particles in an electromagnetic field, the Lagrangian is (cf. 1.63)

$$L = \sum_i \frac{1}{2} m_i \dot{r}_i^2 - \sum_i q_i \phi(x_i) + \sum_i q_i \mathbf{A}(x_i) \cdot \dot{\mathbf{r}}_i$$

(q_i here denotes charge) and the generalized momentum conjugate to x_i is

$$p_{ix} = \frac{\partial L}{\partial \dot{x}_i} = m_i \dot{x}_i + q_i A_x, \quad (2.45)$$

i.e., mechanical momentum plus an additional term.

If the Lagrangian of a system does not contain a given coordinate q_j (although it may contain the corresponding velocity \dot{q}_j), then the coordinate is said to be *cyclic* or *ignorable*. This definition is not universal, but it is the customary one and will be used here. The Lagrange equation of motion,

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0,$$

reduces, for a cyclic coordinate, to

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_j} = 0$$

or

$$\frac{dp_j}{dt} = 0,$$

which mean that

$$p_j = \text{constant.} \quad (2.46)$$

Hence, we can state as a general conservation theorem that *the generalized momentum conjugate to a cyclic coordinate is conserved.*

Note that the derivation of Eq. (2.46) assumes that q_j is a generalized coordinate; one that is linearly independent of all the other coordinates. When equations of constraint exist, all the coordinates are not linearly independent. For example, the angular coordinate θ is not present in the Lagrangian of a hoop rolling without slipping in a horizontal plane that was previously discussed, but the angle appears in the constraint equations $rd\theta = dx$. As a result, the angular momentum, $p_\theta = mr^2\dot{\theta}$, is not a constant of the motion.

Equation (2.46) constitutes a first integral of the form (2.43) for the equations of motion. It can be used formally to eliminate the cyclic coordinate from the problem, which can then be solved entirely in terms of the remaining generalized coordinates. Briefly, the procedure, originated by Routh, consists in modifying the Lagrangian so that it is no longer a function of the generalized velocity corresponding to the cyclic coordinate, but instead involves only its conjugate momentum. The advantage in so doing is that p_j can then be considered one of the constants of integration, and the remaining integrations involve only the non-cyclic coordinates. We shall defer a detailed discussion of Routh's method until the Hamiltonian formulation (to which it is closely related) is treated.

Note that the conditions for the conservation of generalized momenta are more general than the two momentum conservation theorems previously derived. For example, they furnish a conservation theorem for a case in which the law of action and reaction is violated, namely, when electromagnetic forces are present. Suppose we have a single particle in a field in which neither ϕ nor \mathbf{A} depends on x . Then x nowhere appears in L and is therefore cyclic. The corresponding canonical momentum p_x must therefore be conserved. From (1.63) this momentum now has the form

$$p_x = m\dot{x} + qA_x = \text{constant.} \quad (2.47)$$

In this case, it is not the mechanical linear momentum $m\dot{x}$ that is conserved but rather its sum with qA_x .* Nevertheless, it should still be true that the conservation theorems of Chapter 1 are contained within the general rule for cyclic coordinates; with proper restrictions (2.46) should reduce to the theorems of Section 1.2.

*It can be shown from classical electrodynamics that under these conditions, i.e., neither \mathbf{A} nor ϕ depending on x , that qA_x is exactly the x -component of the electromagnetic linear momentum of the field associated with the charge q .

We first consider a generalized coordinate q_j , for which a change dq_j represents a translation of the system as a whole in some given direction. An example would be one of the Cartesian coordinates of the center of mass of the system. Then clearly q_j cannot appear in T , for velocities are not affected by a shift in the origin, and therefore the partial derivative of T with respect to q_j must be zero. Further, we will assume conservative systems for which V is not a function of the velocities, so as to eliminate such complications as electromagnetic forces. The Lagrange equation of motion for a coordinate so defined then reduces to

$$\frac{d}{dt} \frac{\partial T}{\partial \dot{q}_j} \equiv \dot{p}_j = -\frac{\partial V}{\partial q_j} \equiv Q_j. \quad (2.48)$$

We will now show that (2.48) is the equation of motion for the total linear momentum, i.e., that Q_j represents the component of the total force along the direction of translation of q_j , and p_j is the component of the total linear momentum along this direction. In general, the generalized force Q_j is given by Eq. (1.49):

$$Q_j = \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j}.$$

Since dq_j corresponds to a translation of the system along some axis, the vectors $\mathbf{r}_i(q_j)$ and $\mathbf{r}_i(q_j + dq_j)$ are related as shown in Fig. 2.7. By the definition of a derivative, we have

$$\frac{\partial \mathbf{r}_i}{\partial q_j} = \lim_{dq_j \rightarrow 0} \frac{\mathbf{r}_i(q_j + dq_j) - \mathbf{r}_i(q_j)}{dq_j} = \frac{dq_j}{dq_j} \mathbf{n} = \mathbf{n}, \quad (2.49)$$

where \mathbf{n} is the unit vector along the direction of the translation. Hence,

$$Q_j = \sum_i \mathbf{F}_i \cdot \mathbf{n} = \mathbf{n} \cdot \mathbf{F},$$

which (as was stated) is the component of the total force in the direction of \mathbf{n} . To prove the other half of the statement, note that with the kinetic energy in the form

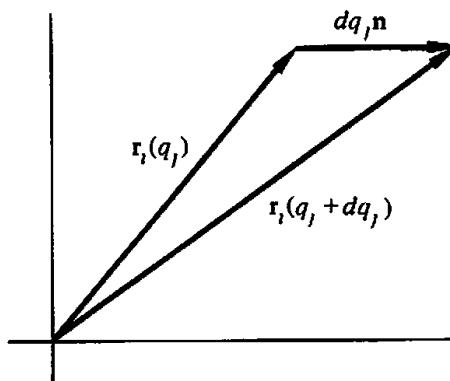


FIGURE 2.7 Change in a position vector under translation of the system.

$$T = \frac{1}{2} \sum_i m_i \dot{\mathbf{r}}_i^2,$$

the conjugate momentum is

$$\begin{aligned} p_J &= \frac{\partial T}{\partial \dot{q}_J} = \sum_i m_i \dot{\mathbf{r}}_i \cdot \frac{\partial \dot{\mathbf{r}}_i}{\partial \dot{q}_J} \\ &= \sum_i m_i \mathbf{v}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_J}, \end{aligned}$$

using Eq. (1.51). Then from Eq. (2.49)

$$p_J = \mathbf{n} \cdot \sum_i m_i \mathbf{v}_i,$$

which again, as predicted, is the component of the total system linear momentum along \mathbf{n} .

Suppose now that the translation coordinate q_J that we have been discussing is cyclic. Then q_J cannot appear in V and therefore

$$-\frac{\partial V}{\partial q_J} \equiv Q_J = 0.$$

But this is simply the familiar conservation theorem for linear momentum—that if a given component of the total applied force vanishes, the corresponding component of the linear momentum is conserved.

In a similar fashion, it can be shown that if a cyclic coordinate q_J is such that dq_J corresponds to a rotation of the system of particles around some axis, then the conservation of its conjugate momentum corresponds to conservation of an angular momentum. By the same argument used above, T cannot contain q_J , for a rotation of the coordinate system cannot affect the magnitude of the velocities. Hence, the partial derivative of T with respect to q_J must again be zero, and since V is independent of q_J , we once more get Eq. (2.48). But now we wish to show that with q_J a rotation coordinate the generalized force is the component of the total applied torque about the axis of rotation, and p_J is the component of the total angular momentum along the same axis.

The generalized force Q_J is again given by

$$Q_J = \sum_i \mathbf{F}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_J},$$

only the derivative now has a different meaning. Here the change in q_J must correspond to an infinitesimal rotation of the vector \mathbf{r}_i , keeping the magnitude of the vector constant. From Fig. 2.8, the magnitude of the derivative can easily be obtained:

$$|d\mathbf{r}_i| = r_i \sin \theta dq_J$$

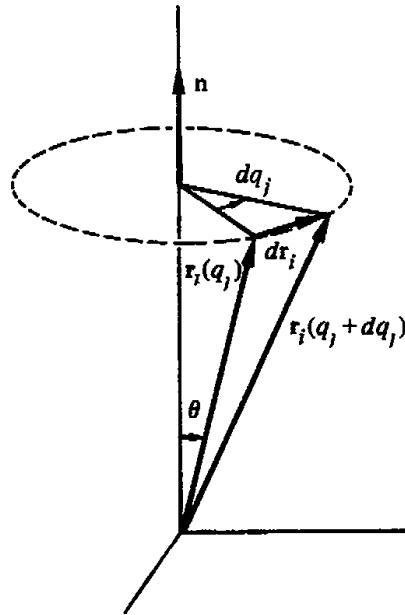


FIGURE 2.8 Change of a position vector under rotation of the system.

and

$$\left| \frac{\partial \mathbf{r}_i}{\partial q_j} \right| = r_i \sin \theta,$$

and its direction is perpendicular to both \mathbf{r}_i and \mathbf{n} . Clearly, the derivative can be written in vector form as

$$\frac{\partial \mathbf{r}_i}{\partial q_j} = \mathbf{n} \times \mathbf{r}_i. \quad (2.50)$$

With this result, the generalized force becomes

$$\begin{aligned} Q_j &= \sum_i \mathbf{F}_i \cdot \mathbf{n} \times \mathbf{r}_i \\ &= \sum_i \mathbf{n} \cdot \mathbf{r}_i \times \mathbf{F}_i, \end{aligned}$$

reducing to

$$Q_j = \mathbf{n} \cdot \sum_i \mathbf{N}_i = \mathbf{n} \cdot \mathbf{N},$$

which proves the first part. A similar manipulation of p_j with the aid of Eq. (2.50) provides proof of the second part of the statement:

$$p_j = \frac{\partial T}{\partial \dot{q}_j} = \sum_i m_i \mathbf{v}_i \cdot \frac{\partial \mathbf{r}_i}{\partial q_j} = \sum_i \mathbf{n} \cdot \mathbf{r}_i \times m_i \mathbf{v}_i = \mathbf{n} \cdot \sum_i \mathbf{L}_i = \mathbf{n} \cdot \mathbf{L}.$$

Summarizing these results, we see that if the rotation coordinate q_j is cyclic, then Q_j , which is the component of the applied torque along \mathbf{n} , vanishes, and the component of \mathbf{L} along \mathbf{n} is constant. Here we have recovered the angular momentum conservation theorem out of the general conservation theorem relating to cyclic coordinates.

The significance of cyclic translation or rotation coordinates in relation to the properties of the system deserves some comment at this point. If a generalized coordinate corresponding to a displacement is cyclic, it means that a translation of the system, as if rigid, has no effect on the problem. In other words, if the system is *invariant* under translation along a given direction, the corresponding linear momentum is conserved. Similarly, the fact that a generalized rotation coordinate is cyclic (and therefore the conjugate angular momentum conserved) indicates that the system is invariant under rotation about the given axis. Thus, the momentum conservation theorems are closely connected with the *symmetry properties* of the system. If the system is spherically symmetric, we can say without further ado that all components of angular momentum are conserved. Or, if the system is symmetric only about the z axis, then only L_z will be conserved, and so on for the other axes. These symmetry considerations can often be used with relatively complicated problems to determine by inspection whether certain constants of the motion exist. (cf. Noether's theorem—Sec. 13.7.)

Suppose, for example, the system consists of a set of mass points moving in a potential field generated by fixed sources uniformly distributed on an infinite plane, say, the $z = 0$ plane. (The sources might be a mass distribution if the forces were gravitational, or a charge distribution for electrostatic forces.) Then the symmetry of the problem is such that the Lagrangian is invariant under a translation of the system of particles in the x - or y -directions (but not in the z -direction) and also under a rotation about the z axis. It immediately follows that the x - and y -components of the total linear momentum, P_x and P_y , are constants of the motion along with L_z , the z -component of the total angular momentum. However, if the sources were restricted only to the half plane, $x \geq 0$, then the symmetry for translation along the x axis and for rotation about the z axis would be destroyed. In that case, P_x and L_z could not be conserved, but P_y would remain a constant of the motion. We will encounter the connections between the constants of motion and the symmetry properties of the system several times in the following chapters.

2.7 ■ ENERGY FUNCTION AND THE CONSERVATION OF ENERGY

Another conservation theorem we should expect to obtain in the Lagrangian formulation is the conservation of total energy for systems where the forces are derivable from potentials dependent only upon position. Indeed, it is possible to demonstrate a conservation theorem for which conservation of total energy represents only a special case. Consider a general Lagrangian, which will be a function of the coordinates q_j and the velocities \dot{q}_j , and may also depend explicitly on the time. (The explicit time dependence may arise from the time variation of external

potentials, or from time-dependent constraints.) Then the total time derivative of L is

$$\frac{dL}{dt} = \sum_J \frac{\partial L}{\partial q_J} \frac{dq_J}{dt} + \sum_J \frac{\partial L}{\partial \dot{q}_J} \frac{d\dot{q}_J}{dt} + \frac{\partial L}{\partial t}. \quad (2.51)$$

From Lagrange's equations,

$$\frac{\partial L}{\partial q_J} = \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_J} \right),$$

and (2.51) can be rewritten as

$$\frac{dL}{dt} = \sum_J \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_J} \right) \dot{q}_J + \sum_J \frac{\partial L}{\partial \dot{q}_J} \frac{d\dot{q}_J}{dt} + \frac{\partial L}{\partial t}$$

or

$$\frac{dL}{dt} = \sum_J \frac{d}{dt} \left(\dot{q}_J \frac{\partial L}{\partial \dot{q}_J} \right) + \frac{\partial L}{\partial t}.$$

It therefore follows that

$$\frac{d}{dt} \left(\sum_J \dot{q}_J \frac{\partial L}{\partial \dot{q}_J} - L \right) + \frac{\partial L}{\partial t} = 0. \quad (2.52)$$

The quantity in parentheses is oftentimes called the *energy function** and will be denoted by h :

$$h(q_1, \dots, q_n; \dot{q}_1, \dots, \dot{q}_n; t) = \sum_J \dot{q}_J \frac{\partial L}{\partial \dot{q}_J} - L, \quad (2.53)$$

and Eq. (2.52) can be looked on as giving the total time derivative of h :

$$\frac{dh}{dt} = -\frac{\partial L}{\partial t}. \quad (2.54)$$

If the Lagrangian is not an explicit function of time, i.e., if t does not appear in L explicitly but only implicitly through the time variation of q and \dot{q} , then Eq. (2.54) says that h is conserved. It is one of the first integrals of the motion and is sometimes referred to as Jacobi's integral.[†]

*The energy function h is identical in value with the Hamiltonian H (See Chapter 8). It is given a different name and symbol here to emphasize that h is considered a function of n independent variables q_J and their time derivatives \dot{q}_J (along with the time), whereas the Hamiltonian will be treated as a function of $2n$ independent variables, q_J, p_J (and possibly the time).

[†]This designation is most often confined to a first integral in the restricted three-body problem. However, the integral there is merely a special case of the energy function h , and there is some historical precedent to apply the name Jacobi integral to the more general situation.

Under certain circumstances, the function h is the total energy of the system. To determine what these circumstances are, we recall that the total kinetic energy of a system can always be written as

$$T = T_0 + T_1 + T_2, \quad (1.73)$$

where T_0 is a function of the generalized coordinates only, $T_1(q, \dot{q})$ is linear in the generalized velocities, and $T_2(q, \dot{q})$ is a quadratic function of the \dot{q} 's. For a very wide range of systems and sets of generalized coordinates, the Lagrangian can be similarly decomposed as regards its functional behavior in the \dot{q} variables:

$$L(q, \dot{q}, t) = L_0(q, t) + L_1(q, \dot{q}, t) + L_2(q, \dot{q}, t). \quad (2.55)$$

Here L_2 is a homogeneous function of the second degree (not merely quadratic) in \dot{q} , while L_1 is homogeneous of the first degree in \dot{q} . There is no reason intrinsic to mechanics that requires the Lagrangian to conform to Eq. (2.55), but in fact it does for most problems of interest. The Lagrangian clearly has this form when the forces are derivable from a potential not involving the velocities. Even with the velocity-dependent potentials, we note that the Lagrangian for a charged particle in an electromagnetic field, Eq. (1.63), satisfies Eq. (2.55). Now, recall that Euler's theorem states that if f is a homogeneous function of degree n in the variables x_i , then

$$\sum_i x_i \frac{\partial f}{\partial x_i} = nf. \quad (2.56)$$

Applied to the function h , Eq. (2.53), for the Lagrangians of the form (2.55), this theorem implies that

$$h = 2L_2 + L_1 - L = L_2 - L_0. \quad (2.57)$$

If the transformation equations defining the generalized coordinates, Eqs. (1.38), do not involve the time explicitly, then by Eqs. (1.73) $T = T_2$. If, further, the potential does not depend on the generalized velocities, then $L_2 = T$ and $L_0 = -V$, so that

$$h = T + V = E, \quad (2.58)$$

and the energy function is indeed the total energy. Under these circumstances, if V does not involve the time explicitly, neither will L . Thus, by Eq. (2.54), h (which is here the total energy), will be conserved.

Note that the conditions for conservation of h are in principle quite distinct from those that identify h as the total energy. We can have a set of generalized coordinates such that in a particular problem h is conserved but is not the total energy. On the other hand, h can be the total energy, in the form $T + V$, but not be conserved. Also note that whereas the Lagrangian is uniquely fixed for each

system by the prescription

$$L = T - U$$

independent of the choice of generalized coordinates, the energy function h depends in magnitude and functional form on the specific set of generalized coordinates. For one and the same system, various energy functions h of different physical content can be generated depending on how the generalized coordinates are chosen.

The most common case that occurs in classical mechanics is one in which the kinetic energy terms are all of the form $m\dot{q}_i^2/2$ or $p_i^2/2m$ and the potential energy depends only upon the coordinates. For these conditions, the energy function is both conserved and is also the total energy.

Finally, note that where the system is not conservative, but there are frictional forces derivable from a dissipation function \mathcal{F} , it can be easily shown that \mathcal{F} is related to the decay rate of h . When the equations of motion are given by Eq. (1.70), including dissipation, then Eq. (2.52) has the form

$$\frac{dh}{dt} + \frac{\partial L}{\partial t} = \sum_j \frac{\partial \mathcal{F}}{\partial \dot{q}_j} \dot{q}_j.$$

By the definition of \mathcal{F} , Eq. (1.67), it is a homogeneous function of the \dot{q} 's of degree 2. Hence, applying Euler's theorem again, we have

$$\frac{dh}{dt} = -2\mathcal{F} - \frac{\partial L}{\partial t}. \quad (2.59)$$

If L is not an explicit function of time, and the system is such that h is the same as the energy, then Eq. (2.59) says that $2\mathcal{F}$ is the rate of energy dissipation,

$$\frac{dE}{dt} = -2\mathcal{F}, \quad (2.60)$$

a statement proved above (cf. Sec. 1.5) in less general circumstances.

DERIVATIONS

1. Complete the solution of the brachistochrone problem begun in Section 2.2 and show that the desired curve is a cycloid with a cusp at the initial point at which the particle is released. Show also that if the particle is projected with an initial kinetic energy $\frac{1}{2}mv_0^2$ that the brachistochrone is still a cycloid passing through the two points with a cusp at a height z above the initial point given by $v_0^2 = 2gz$.
2. Show that if the potential in the Lagrangian contains velocity-dependent terms, the canonical momentum corresponding to a coordinate of rotation θ of the entire system

is no longer the mechanical angular momentum L_θ but is given by

$$p_\theta = L_\theta - \sum_i \mathbf{n} \cdot \mathbf{r}_i \times \nabla_{\mathbf{v}_i} U,$$

where $\nabla_{\mathbf{v}}$ is the gradient operator in which the derivatives are with respect to the velocity components and \mathbf{n} is a unit vector in the direction of rotation. If the forces are electromagnetic in character, the canonical momentum is therefore

$$p_\theta = L_\theta + \sum_i \mathbf{n} \cdot \mathbf{r}_i \times \frac{q_i}{c} \mathbf{A}_i.$$

3. Prove that the shortest distance between two points in space is a straight line.
4. Show that the geodesics of a spherical surface are great circles, i.e., circles whose centers lie at the center of the sphere.

EXERCISES

5. A particle is subjected to the potential $V(x) = -Fx$, where F is a constant. The particle travels from $x = 0$ to $x = a$ in a time interval t_0 . Assume the motion of the particle can be expressed in the form $x(t) = A + Bt + Ct^2$. Find the values of A , B , and C such that the action is a minimum.
6. Find the Euler–Lagrange equation describing the brachistochrone curve for a particle moving *inside* a spherical Earth of uniform mass density. Obtain a first integral for this differential equation by analogy to the Jacobi integral h . With the help of this integral, show that the desired curve is a hypocycloid (the curve described by a point on a circle rolling on the inside of a larger circle). Obtain an expression for the time of travel along the brachistochrone between two points on Earth's surface. How long would it take to go from New York to Los Angeles (assumed to be 4800 km apart on the surface) along a brachistochrone tunnel (assuming no friction) and how far below the surface would the deepest point of the tunnel be?
7. In Example 2 of Section 2.1 we considered the problem of the minimum surface of revolution. Examine the symmetric case $x_1 = x_2$, $y_2 = -y_1 > 0$, and express the condition for the parameter a as a transcendental equation in terms of the dimensionless quantities $k = x_2/a$, and $\alpha = y_2/x_2$. Show that for α greater than a certain value α_0 two values of k are possible, for $\alpha = \alpha_0$ only one value of k is possible, while if $\alpha < \alpha_0$ no real value of k (or a) can be found, so that no catenary solution exists in this region. Find the value of α_0 , numerically if necessary.
8. The broken-segment solution described in the text (cf. p. 42), in which the area of revolution is only that of the end circles of radius y_1 and y_2 , respectively, is known as the *Goldschmidt solution*. For the symmetric situation discussed in Exercise 7, obtain an expression for the ratio of the area generated by the catenary solutions to that given by the Goldschmidt solution. Your result should be a function only of the parameters k and α . Show that for sufficiently large values of α at least one of the catenaries gives an area below that of the Goldschmidt solution. On the other hand, show that if $\alpha = \alpha_0$, the Goldschmidt solution gives a lower area than the catenary.

9. A chain or rope of indefinite length passes freely over pulleys at heights y_1 and y_2 above the plane surface of Earth, with a horizontal distance $x_2 - x_1$ between them. If the chain or rope has a uniform linear mass density, show that the problem of finding the curve assumed between the pulleys is identical with that of the problem of minimum surface of revolution. (The transition to the Goldschmidt solution as the heights y_1 and y_2 are changed makes for a striking lecture demonstration. See Exercise 8.)
10. Suppose it is known experimentally that a particle fell a given distance y_0 in a time $t_0 = \sqrt{2y_0/g}$, but the times of fall for distances other than y_0 is not known. Suppose further that the Lagrangian for the problem is known, but that instead of solving the equation of motion for y as a function of t , it is guessed that the functional form is

$$y = at + bt^2.$$

If the constants a and b are adjusted always so that the time to fall y_0 is correctly given by t_0 , show directly that the integral

$$\int_0^{t_0} L dt$$

is an extremum for real values of the coefficients only when $a = 0$ and $b = g/2$.

11. When two billiard balls collide, the instantaneous forces between them are very large but act only in an infinitesimal time Δt , in such a manner that the quantity

$$\int_{\Delta t} F dt$$

remains finite. Such forces are described as *impulsive* forces, and the integral over Δt is known as the *impulse* of the force. Show that if impulsive forces are present Lagrange's equations may be transformed into

$$\left(\frac{\partial L}{\partial \dot{q}_J} \right)_f - \left(\frac{\partial L}{\partial \dot{q}_J} \right)_i = S_J,$$

where the subscripts i and f refer to the state of the system before and after the impulse, S_J is the impulse of the generalized impulsive force corresponding to q_J , and L is the Lagrangian including all the nonimpulsive forces.

12. The term *generalized mechanics* has come to designate a variety of classical mechanics in which the Lagrangian contains time derivatives of q_i higher than the first. Problems for which $x = f(x, \dot{x}, \ddot{x}, t)$ have been referred to as "jerky" mechanics. Such equations of motion have interesting applications in chaos theory (cf. Chapter 11). By applying the methods of the calculus of variations, show that if there is a Lagrangian of the form $L(q_i, \dot{q}_i, \ddot{q}_i, t)$, and Hamilton's principle holds with the zero variation of both q_i and \dot{q}_i at the end points, then the corresponding Euler-Lagrange equations are

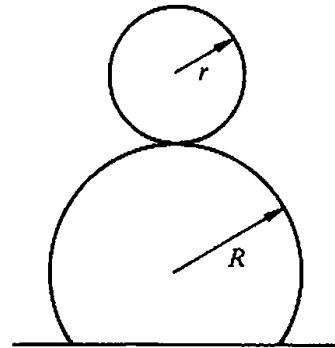
$$\frac{d^2}{dt^2} \left(\frac{\partial L}{\partial \ddot{q}_i} \right) - \frac{d}{dt} \left(\frac{\partial L}{\partial \dot{q}_i} \right) + \frac{\partial L}{\partial q_i} = 0, \quad i = 1, 2, \dots, n.$$

Apply this result to the Lagrangian

$$L = -\frac{m}{2}q\ddot{q} - \frac{k}{2}q^2.$$

Do you recognize the equations of motion?

13. A heavy particle is placed at the top of a vertical hoop. Calculate the reaction of the hoop on the particle by means of the Lagrange's undetermined multipliers and Lagrange's equations. Find the height at which the particle falls off.
14. A uniform hoop of mass m and radius r rolls without slipping on a fixed cylinder of radius R as shown in the figure. The only external force is that of gravity. If the smaller cylinder starts rolling from rest on top of the bigger cylinder, use the method of Lagrange multipliers to find the point at which the hoop falls off the cylinder.



15. A form of the Wheatstone impedance bridge has, in addition to the usual four resistances, an inductance in one arm and a capacitance in the opposite arm. Set up L and \mathcal{F} for the unbalanced bridge, with the charges in the elements as coordinates. Using the Kirchhoff junction conditions as constraints on the currents, obtain the Lagrange equations of motion, and show that eliminating the λ 's reduces these to the usual network equations.
16. In certain situations, particularly one-dimensional systems, it is possible to incorporate frictional effects without introducing the dissipation function. As an example, find the equations of motion for the Lagrangian

$$L = e^{\gamma t} \left(\frac{m\dot{q}^2}{2} - \frac{kq^2}{2} \right).$$

How would you describe the system? Are there any constants of motion? Suppose a point transformation is made of the form

$$s = e^{\gamma t} q.$$

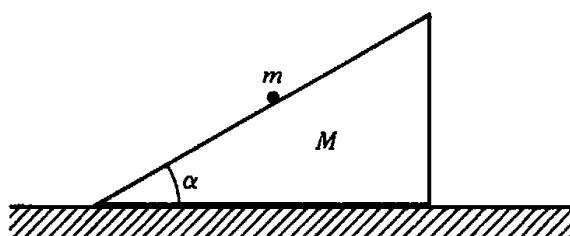
What is the effective Lagrangian in terms of s ? Find the equation of motion for s . What do these results say about the conserved quantities for the system?

17. It sometimes occurs that the generalized coordinates appear separately in the kinetic energy and the potential energy in such a manner that T and V may be written in the form

$$T = \sum_i f_i(q_i) \dot{q}_i^2 \quad \text{and} \quad V = \sum_i V_i(q_i)$$

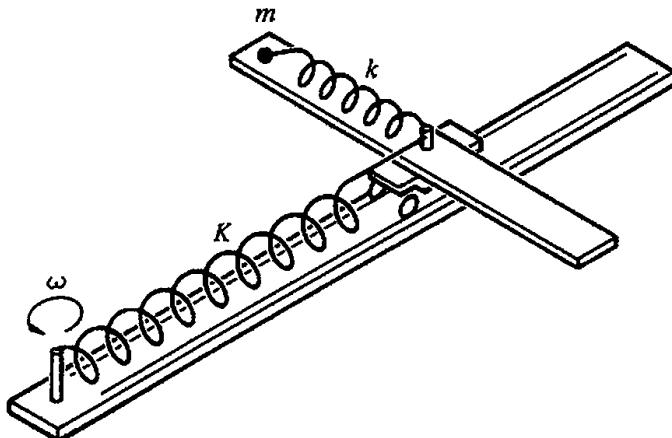
Show that Lagrange's equations then separate, and that the problem can always be reduced to quadratures.

18. A point mass is constrained to move on a massless hoop of radius a fixed in a vertical plane that rotates about its vertical symmetry axis with constant angular speed ω . Obtain the Lagrange equations of motion assuming the only external forces arise from gravity. What are the constants of motion? Show that if ω is greater than a critical value ω_0 , there can be a solution in which the particle remains stationary on the hoop at a point other than at the bottom, but that if $\omega < \omega_0$, the only stationary point for the particle is at the bottom of the hoop. What is the value of ω_0 ?
19. A particle moves without friction in a conservative field of force produced by various mass distributions. In each instance, the force generated by a volume element of the distribution is derived from a potential that is proportional to the mass of the volume element and is a function only of the scalar distance from the volume element. For the following fixed, homogeneous mass distributions, state the conserved quantities in the motion of the particle:
 - (a) The mass is uniformly distributed in the plane $z = 0$.
 - (b) The mass is uniformly distributed in the half-plane $z = 0, y > 0$.
 - (c) The mass is uniformly distributed in a circular cylinder of infinite length, with axis along the z axis.
 - (d) The mass is uniformly distributed in a circular cylinder of finite length, with axis along the z axis.
 - (e) The mass is uniformly distributed in a right cylinder of elliptical cross section and infinite length, with axis along the z axis.
 - (f) The mass is uniformly distributed in a dumbbell whose axis is oriented along the z axis.
 - (g) The mass is in the form of a uniform wire wound in the geometry of an infinite helical solenoid, with axis along the z axis
20. A particle of mass m slides without friction on a wedge of angle α and mass M that can move without friction on a smooth horizontal surface, as shown in the figure. Treating the constraint of the particle on the wedge by the method of Lagrange multipliers, find the equations of motion for the particle and wedge. Also obtain an expression for the forces of constraint. Calculate the work done in time t by the forces of constraint acting on the particle and on the wedge. What are the constants of motion for the system? Contrast the results you have found with the situation when the wedge is fixed. [Suggestion: For the particle you may either use a Cartesian coordinate system with y vertical, or one with y normal to the wedge or, even more instructively, do it in both systems.]



21. A carriage runs along rails on a rigid beam, as shown in the figure below. The carriage is attached to one end of a spring of equilibrium length r_0 and force constant k , whose other end is fixed on the beam. On the carriage, another set of rails is perpendicular to the first along which a particle of mass m moves, held by a spring fixed on the beam, of force constant K and zero equilibrium length. Beam, rails, springs, and carriage are assumed to have zero mass. The whole system is forced to move in a plane about the point of attachment of the first spring, with a constant angular speed ω . The length of the second spring is at all times considered small compared to r_0 .

- (a) What is the energy of the system? Is it conserved?
- (b) Using generalized coordinates in the laboratory system, what is the Jacobi integral for the system? Is it conserved?
- (c) In terms of the generalized coordinates relative to a system rotating with the angular speed ω , what is the Lagrangian? What is the Jacobi integral? Is it conserved? Discuss the relationship between the two Jacobi integrals.



22. Suppose a particle moves in space subject to a conservative potential $V(\mathbf{r})$ but is constrained to always move on a surface whose equation is $\sigma(\mathbf{r}, t) = 0$. (The explicit dependence on t indicates that the surface may be moving.) The instantaneous force of constraint is taken as always perpendicular to the surface. Show analytically that the energy of the particle is not conserved if the surface moves in time. What physically is the reason for nonconservation of the energy under this circumstance?
23. Consider two particles of masses m_1 and m_2 . Let m_1 be confined to move on a circle of radius a in the $z = 0$ plane, centered at $x = y = 0$. Let m_1 be confined to move on a circle of radius b in the $z = c$ plane, centered at $x = y = 0$. A light (massless) spring of spring constant k is attached between the two particles.
- (a) Find the Lagrangian for the system.
 - (b) Solve the problem using Lagrange multipliers and give a physical interpretation for each multiplier.
24. The one-dimensional harmonic oscillator has the Lagrangian $L = m\dot{x}^2/2 - kx^2/2$. Suppose you did not know the solution to the motion, but realized that the motion must be periodic and therefore could be described by a Fourier series of the form

$$x(t) = \sum_{j=0} a_j \cos j\omega t,$$

(taking $t = 0$ at a turning point) where ω is the (unknown) angular frequency of the motion. This representation for $x(t)$ defines a many-parameter path for the system point in configuration space. Consider the action integral I for two points, t_1 and t_2 separated by the period $T = 2\pi/\omega$. Show that with this form for the system path, I is an extremum for nonvanishing x only if $a_j = 0$, for $j \neq 1$, and only if $\omega^2 = k/m$.

25. A disk of radius R rolls without slipping inside the stationary parabola $y = ax^2$. Find the equations of constraint. What condition allows the disk to roll so that it touches the parabola at one and only one point independent of its position?
26. A particle of mass m is suspended by a massless spring of length L . It hangs, without initial motion, in a gravitational field of strength g . It is struck by an impulsive horizontal blow, which introduces an angular velocity ω . If ω is sufficiently small, it is obvious that the mass moves as a simple pendulum. If ω is sufficiently large, the mass will rotate about the support. Use a Lagrange multiplier to determine the conditions under which the string becomes slack at some point in the motion.

CHAPTER

3

The Central Force Problem

In this chapter we shall discuss the problem of two bodies moving under the influence of a mutual central force as an application of the Lagrangian formulation. Not all the problems of central force motion are integrable in terms of well-known functions. However, we shall attempt to explore the problem as thoroughly as is possible with the tools already developed. In the last section of this chapter we consider some of the complications that follow by the presence of a third body.

3.1 ■ REDUCTION TO THE EQUIVALENT ONE-BODY PROBLEM

Consider a monogenic system of two mass points, m_1 and m_2 (cf. Fig. 3.1), where the only forces are those due to an interaction potential U . We will assume at first that U is any function of the vector between the two particles, $\mathbf{r}_2 - \mathbf{r}_1$, or of their relative velocity, $\dot{\mathbf{r}}_2 - \dot{\mathbf{r}}_1$, or of any higher derivatives of $\mathbf{r}_2 - \mathbf{r}_1$. Such a system has six degrees of freedom and hence six independent generalized coordinates. We choose these to be the three components of the radius vector to the center of mass, \mathbf{R} , plus the three components of the difference vector $\mathbf{r} = \mathbf{r}_2 - \mathbf{r}_1$. The Lagrangian will then have the form

$$L = T(\dot{\mathbf{R}}, \dot{\mathbf{r}}) - U(\mathbf{r}, \dot{\mathbf{r}}, \dots). \quad (3.1)$$

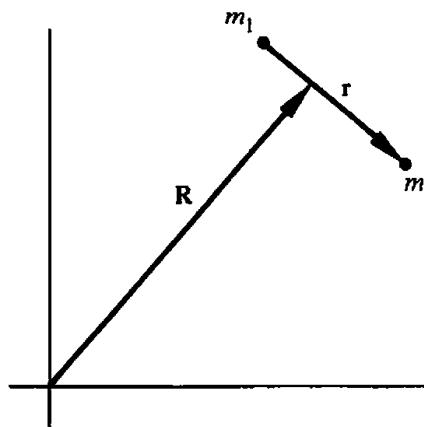


FIGURE 3.1 Coordinates for the two-body problem.

The kinetic energy T can be written as the sum of the kinetic energy of the motion of the center of mass, plus the kinetic energy of motion about the center of mass, T' :

$$T = \frac{1}{2} (m_1 + m_2) \dot{\mathbf{R}}^2 + T'$$

with

$$T' = \frac{1}{2} m_1 \dot{\mathbf{r}}_1'^2 + \frac{1}{2} m_2 \dot{\mathbf{r}}_2'^2.$$

Here \mathbf{r}'_1 and \mathbf{r}'_2 are the radii vectors of the two particles relative to the center of mass and are related to \mathbf{r} by

$$\begin{aligned} \mathbf{r}'_1 &= -\frac{m_2}{m_1 + m_2} \mathbf{r}, \\ \mathbf{r}'_2 &= \frac{m_1}{m_1 + m_2} \mathbf{r} \end{aligned} \quad (3.2)$$

Expressed in terms of \mathbf{r} by means of Eq. (3.2), T' takes on the form

$$T' = \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} \dot{\mathbf{r}}^2$$

and the total Lagrangian (3.1) is

$$L = \frac{m_1 + m_2}{2} \dot{\mathbf{R}}^2 + \frac{1}{2} \frac{m_1 m_2}{m_1 + m_2} \dot{\mathbf{r}}^2 - U(\mathbf{r}, \dot{\mathbf{r}}, \dots). \quad (3.3)$$

It is seen that the three coordinates \mathbf{R} are cyclic, so that the center of mass is either at rest or moving uniformly. None of the equations of motion for \mathbf{r} will contain terms involving \mathbf{R} or $\dot{\mathbf{R}}$. Consequently, the process of integration is particularly simple here. We merely drop the first term from the Lagrangian in all subsequent discussion.

The rest of the Lagrangian is exactly what would be expected if we had a fixed center of force with a single particle at a distance \mathbf{r} from it, having a mass

$$\mu = \frac{m_1 m_2}{m_1 + m_2}, \quad (3.4)$$

where μ is known as the *reduced mass*. Frequently, Eq. (3.4) is written in the form

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}. \quad (3.5)$$

Thus, the central force motion of two bodies about their center of mass can always be reduced to an equivalent one-body problem.

3.2 ■ THE EQUATIONS OF MOTION AND FIRST INTEGRALS

We now restrict ourselves to conservative central forces, where the potential is $V(r)$, a function of r only, so that the force is always along \mathbf{r} . By the results of the preceding section, we need only consider the problem of a single particle of reduced mass m moving about a fixed center of force, which will be taken as the origin of the coordinate system. Since potential energy involves only the radial distance, the problem has spherical symmetry; i.e., any rotation, about any fixed axis, can have no effect on the solution. Hence, an angle coordinate representing rotation about a fixed axis must be cyclic. These symmetry properties result in a considerable simplification in the problem.

Since the problem is spherically symmetric, the total angular momentum vector,

$$\mathbf{L} = \mathbf{r} \times \mathbf{p},$$

is conserved. It therefore follows that \mathbf{r} is always perpendicular to the fixed direction of \mathbf{L} in space. This can be true only if \mathbf{r} always lies in a plane whose normal is parallel to \mathbf{L} . While this reasoning breaks down if \mathbf{L} is zero, the motion in that case must be along a straight line going through the center of force, for $\mathbf{L} = 0$ requires \mathbf{r} to be parallel to $\dot{\mathbf{r}}$, which can be satisfied only in straight-line motion.* Thus, central force motion is always motion in a plane.

Now, the motion of a single particle in space is described by three coordinates; in spherical polar coordinates these are the azimuth angle θ , the zenith angle (or colatitude) ψ , and the radial distance r . By choosing the polar axis to be in the direction of \mathbf{L} , the motion is always in the plane perpendicular to the polar axis. The coordinate ψ then has only the constant value $\pi/2$ and can be dropped from the subsequent discussion. The conservation of the angular momentum vector furnishes three independent constants of motion (corresponding to the three Cartesian components). In effect, two of these, expressing the constant *direction* of the angular momentum, have been used to reduce the problem from three to two degrees of freedom. The third of these constants, corresponding to the conservation of the magnitude of \mathbf{L} , remains still at our disposal in completing the solution.

Expressed now in plane polar coordinates, the Lagrangian is

$$\begin{aligned} L &= T - V \\ &= \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r). \end{aligned} \quad (3.6)$$

As was foreseen, θ is a cyclic coordinate, whose corresponding canonical momentum is the angular momentum of the system:

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta}.$$

*Formally $\dot{\mathbf{r}} = \dot{r}\mathbf{n}_r + r\dot{\theta}\mathbf{n}_\theta$, hence $\mathbf{r} \times \dot{\mathbf{r}} = 0$ requires $\dot{\theta} = 0$.

One of the two equations of motion is then simply

$$\dot{p}_\theta = \frac{d}{dt} (mr^2\dot{\theta}) = 0. \quad (3.7)$$

with the immediate integral

$$mr^2\dot{\theta} = l. \quad (3.8)$$

where l is the constant magnitude of the angular momentum. From (3.7) it also follows that

$$\frac{d}{dt} \left(\frac{1}{2}r^2\dot{\theta} \right) = 0. \quad (3.9)$$

The factor $\frac{1}{2}$ is inserted because $\frac{1}{2}r^2\dot{\theta}$ is just the *areal velocity*—the area swept out by the radius vector per unit time. This interpretation follows from Fig. 3.2, the differential area swept out in time dt being

$$dA = \frac{1}{2}r(r d\theta),$$

and hence

$$\frac{dA}{dt} = \frac{1}{2}r^2 \frac{d\theta}{dt}.$$

The conservation of angular momentum is thus equivalent to saying the areal velocity is constant. Here we have the proof of the well-known Kepler's second law of planetary motion: The radius vector sweeps out equal areas in equal times. It should be emphasized however that the conservation of the areal velocity is a general property of central force motion and is not restricted to an inverse-square law of force.

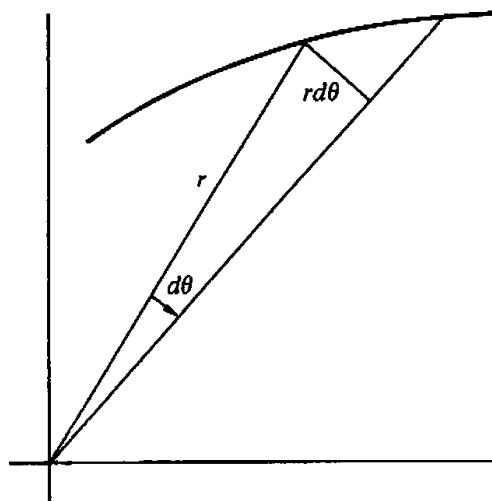


FIGURE 3.2 The area swept out by the radius vector in a time dt .

The remaining Lagrange equation, for the coordinate r , is

$$\frac{d}{dt}(m\dot{r}) - mr\dot{\theta}^2 + \frac{\partial V}{\partial r} = 0. \quad (3.10)$$

Designating the value of the force along \mathbf{r} , $-\partial V/\partial r$, by $f(r)$ the equation can be rewritten as

$$m\ddot{r} - mr\dot{\theta}^2 = f(r). \quad (3.11)$$

By making use of the first integral, Eq. (3.8), $\dot{\theta}$ can be eliminated from the equation of motion, yielding a second-order differential equation involving r only:

$$m\ddot{r} - \frac{l^2}{mr^3} = f(r). \quad (3.12)$$

There is another first integral of motion available, namely the total energy, since the forces are conservative. On the basis of the general energy conservation theorem, we can immediately state that a constant of the motion is

$$E = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + V(r), \quad (3.13)$$

where E is the energy of the system. Alternatively, this first integral could be derived again directly from the equations of motion (3.7) and (3.12). The latter can be written as

$$m\ddot{r} = -\frac{d}{dr}\left(V + \frac{1}{2}\frac{l^2}{mr^2}\right). \quad (3.14)$$

If both sides of Eq. (3.14) are multiplied by \dot{r} the left side becomes

$$m\ddot{r}\dot{r} = \frac{d}{dt}\left(\frac{1}{2}m\dot{r}^2\right).$$

The right side similarly can be written as a total time derivative, for if $g(r)$ is any function of r , then the total time derivative of g has the form

$$\frac{d}{dt}g(r) = \frac{dg}{dr}\frac{dr}{dt}.$$

Hence, Eq. (3.14) is equivalent to

$$\frac{d}{dt}\left(\frac{1}{2}m\dot{r}^2\right) = -\frac{d}{dt}\left(V + \frac{1}{2}\frac{l^2}{mr^2}\right)$$

or

$$\frac{d}{dt}\left(\frac{1}{2}m\dot{r}^2 + \frac{1}{2}\frac{l^2}{mr^2} + V\right) = 0,$$

and therefore

$$\frac{1}{2}m\dot{r}^2 + \frac{1}{2}\frac{l^2}{mr^2} + V = \text{constant}. \quad (3.15)$$

Equation (3.15) is the statement of the conservation of total energy, for by using (3.8) for l , the middle term can be written

$$\frac{1}{2}\frac{l^2}{mr^2} = \frac{1}{2mr^2}m^2r^4\dot{\theta}^2 = \frac{mr^2\dot{\theta}^2}{2}.$$

and (3.15) reduces to (3.13).

These first two integrals give us in effect two of the quadratures necessary to complete the problem. As there are two variables, r and θ , a total of four integrations are needed to solve the equations of motion. The first two integrations have left the Lagrange equations as two first-order equations (3.8) and (3.15); the two remaining integrations can be accomplished (formally) in a variety of ways. Perhaps the simplest procedure starts from Eq. (3.15). Solving for \dot{r} , we have

$$\dot{r} = \sqrt{\frac{2}{m} \left(E - V - \frac{l^2}{2mr^2} \right)}, \quad (3.16)$$

or

$$dt = \frac{dr}{\sqrt{\frac{2}{m} \left(E - V - \frac{l^2}{2mr^2} \right)}}. \quad (3.17)$$

At time $t = 0$, let r have the initial value r_0 . Then the integral of both sides of the equation from the initial state to the state at time t takes the form

$$t = \int_{r_0}^r \frac{dr}{\sqrt{\frac{2}{m} \left(E - V - \frac{l^2}{2mr^2} \right)}}. \quad (3.18)$$

As it stands, Eq. (3.18) gives t as a function of r and the constants of integration E , l , and r_0 . However, it may be inverted, at least formally, to give r as a function of t and the constants. Once the solution for r is found, the solution θ follows immediately from Eq. (3.8), which can be written as

$$d\theta = \frac{l dt}{mr^2}. \quad (3.19)$$

If the initial value of θ is θ_0 , then the integral of (3.19) is simply

$$\theta = l \int_0^t \frac{dt}{mr^2(t)} + \theta_0. \quad (3.20)$$

Equations (3.18) and (3.20) are the two remaining integrations, and formally the problem has been reduced to quadratures, with four constants of integration E , l , r_0 , θ_0 . These constants are not the only ones that can be considered. We might equally as well have taken r_0 , θ_0 , \dot{r}_0 , $\dot{\theta}_0$, but of course E and l can always be determined in terms of this set. For many applications, however, the set containing the energy and angular momentum is the natural one. In quantum mechanics, such constants as the initial values of r and θ , or of \dot{r} and $\dot{\theta}$, become meaningless, but we can still talk in terms of the system energy or of the system angular momentum. Indeed, two salient differences between classical and quantum mechanics appear in the properties of E and l in the two theories. In order to discuss the transition to quantum theories, it is therefore important that the classical description of the system be in terms of its energy and angular momentum.

3.3 ■ THE EQUIVALENT ONE-DIMENSIONAL PROBLEM, AND CLASSIFICATION OF ORBITS

Although we have solved the one-dimensional problem formally, practically speaking the integrals (3.18) and (3.20) are usually quite unmanageable, and in any specific case it is often more convenient to perform the integration in some other fashion. But before obtaining the solution for any specific force laws, let us see what can be learned about the motion in the general case, using only the equations of motion and the conservation theorems, without requiring explicit solutions.

For example, with a system of known energy and angular momentum, the magnitude and direction of the velocity of the particle can be immediately determined in terms of the distance r . The magnitude v follows at once from the conservation of energy in the form

$$E = \frac{1}{2}mv^2 + V(r)$$

or

$$v = \sqrt{\frac{2}{m}(E - V(r))}. \quad (3.21)$$

The radial velocity—the component of $\dot{\mathbf{r}}$ along the radius vector—has been given in Eq. (3.16). Combined with the magnitude v , this is sufficient information to furnish the direction of the velocity.* These results, and much more, can also be obtained from consideration of an equivalent one-dimensional problem.

The equation of motion in r , with $\dot{\theta}$ expressed in terms of l , Eq. (3.12), involves only r and its derivatives. It is the same equation as would be obtained for a

*Alternatively, the conservation of angular momentum furnishes $\dot{\theta}$, the angular velocity, and this together with \dot{r} gives both the magnitude and direction of $\dot{\mathbf{r}}$.

fictitious one-dimensional problem in which a particle of mass m is subject to a force

$$f' = f + \frac{l^2}{mr^3}. \quad (3.22)$$

The significance of the additional term is clear if it is written as $mr\dot{\theta}^2 = mv_\theta^2/r$, which is the familiar centrifugal force. An equivalent statement can be obtained from the conservation theorem for energy. By Eq. (3.15) the motion of the particle in r is that of a one-dimensional problem with a fictitious potential energy:

$$V' = V + \frac{1}{2} \frac{l^2}{mr^2}. \quad (3.22')$$

As a check, note that

$$f' = -\frac{\partial V'}{\partial r} = f(r) + \frac{l^2}{mr^3},$$

which agrees with Eq. (3.22). The energy conservation theorem (3.15) can thus also be written as

$$E = V' + \frac{1}{2}m\dot{r}^2. \quad (3.15')$$

As an illustration of this method of examining the motion, consider a plot of V' against r for the specific case of an attractive inverse-square law of force:

$$f = -\frac{k}{r^2}.$$

(For positive k , the minus sign ensures that the force is *toward* the center of force.) The potential energy for this force is

$$V = -\frac{k}{r},$$

and the corresponding fictitious potential is

$$V' = -\frac{k}{r} + \frac{l^2}{2mr^2}.$$

Such a plot is shown in Fig. 3.3; the two dashed lines represent the separate components

$$-\frac{k}{r} \quad \text{and} \quad \frac{l^2}{2mr^2},$$

and the solid line is the sum V' .

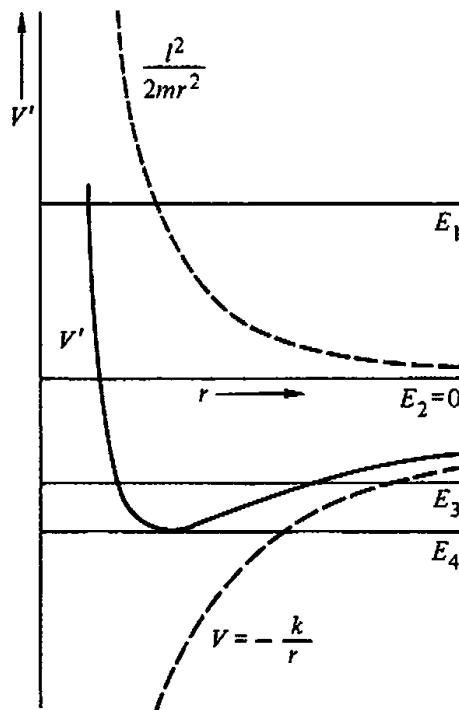


FIGURE 3.3 The equivalent one-dimensional potential for attractive inverse-square law of force.

Let us consider now the motion of a particle having the energy E_1 , as shown in Figs. 3.3 and 3.4. Clearly this particle can never come closer than r_1 (cf. Fig. 3.4). Otherwise with $r < r_1$, V' exceeds E_1 and by Eq. (3.15') the kinetic energy would have to be negative, corresponding to an imaginary velocity! On the other hand, there is no upper limit to the possible value of r , so the orbit is not bounded. A particle will come in from infinity, strike the “repulsive centrifugal barrier,” be repelled, and travel back out to infinity (cf. Fig. 3.5). The distance between E and V' is $\frac{1}{2}mr^2$, i.e., proportional to the square of the radial velocity, and becomes zero, naturally, at the *turning point* r_1 . At the same time, the distance between E and V on the plot is the kinetic energy $\frac{1}{2}mv^2$ at the given value of r . Hence, the distance between the V and V' curves is $\frac{1}{2}mr^2\dot{\theta}^2$. These curves therefore supply the magnitude of the particle velocity and its components for any distance r , at the given energy and angular momentum. This information is sufficient to produce an approximate picture of the form of the orbit.

For the energy $E_2 = 0$ (cf. Fig. 3.3), a roughly similar picture of the orbit behavior is obtained. But for any lower energy, such as E_3 indicated in Fig. 3.6, we have a different story. In addition to a lower bound r_1 , there is also a maximum value r_2 that cannot be exceeded by r with positive kinetic energy. The motion is then “bounded,” and there are two turning points, r_1 and r_2 , also known as *apsidal distances*. This does not necessarily mean that the orbits are closed. All that can be said is that they are bounded, contained between two circles of radius r_1 and r_2 with turning points always lying on the circles (cf. Fig. 3.7).

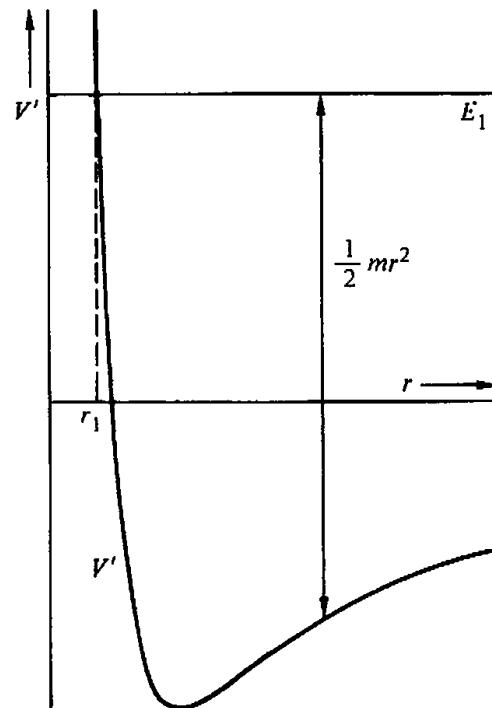


FIGURE 3.4 Unbounded motion at positive energies for inverse-square law of force.

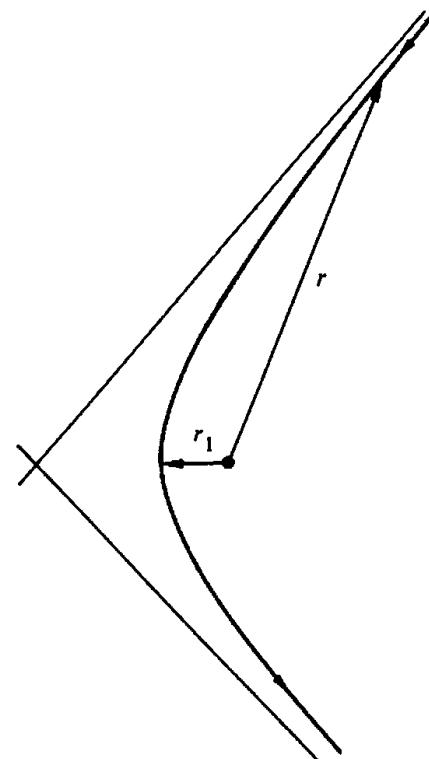


FIGURE 3.5 The orbit for E_1 corresponding to unbounded motion.

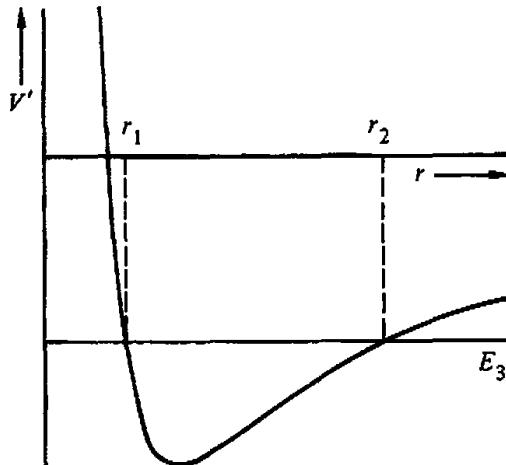


FIGURE 3.6 The equivalent one-dimensional potential for inverse-square law of force, illustrating bounded motion at negative energies.

If the energy is E_4 at the minimum of the fictitious potential as shown in Fig. 3.8, then the two bounds coincide. In such case, motion is possible at only one radius; $\dot{r} = 0$, and the orbit is a circle. Remembering that the effective “force” is the negative of the slope of the V' curve, the requirement for circular orbits is simply that f' be zero, or

$$f(r) = -\frac{l^2}{mr^3} = -mr\dot{\theta}^2.$$

We have here the familiar elementary condition for a circular orbit, that the applied force be equal and opposite to the “reversed effective force” of centripetal

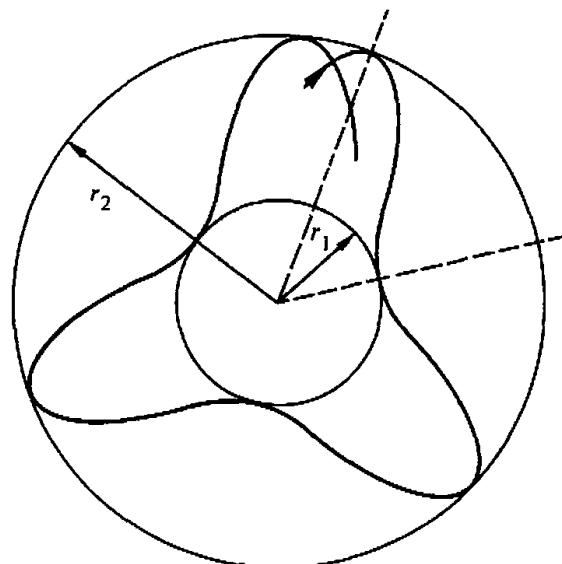


FIGURE 3.7 The nature of the orbits for bounded motion.

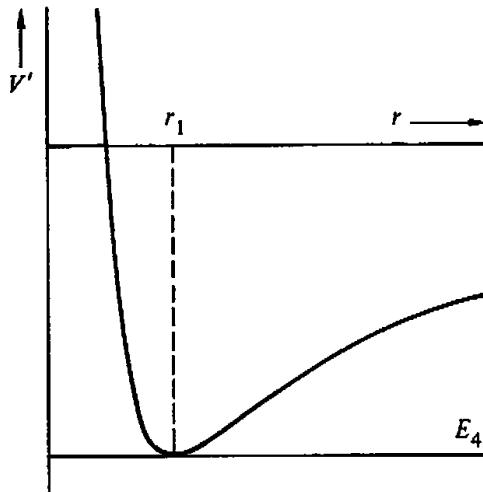


FIGURE 3.8 The equivalent one-dimensional potential of inverse-square law of force, illustrating the condition for circular orbits.

acceleration.* The properties of circular orbits and the conditions for them will be studied in greater detail in Section 3.6.

Note that all of this discussion of the orbits for various energies has been at one value of the angular momentum. Changing l changes the quantitative details of the V' curve, but it does not affect the general classification of the types of orbits.

For the attractive inverse-square law of force discussed above, we shall see that the orbit for E_1 is a hyperbola, for E_2 a parabola, and for E_3 an ellipse. With other forces the orbits may not have such simple forms. However, the same general qualitative division into open, bounded, and circular orbits will be true for any attractive potential that (1) falls off slower than $1/r^2$ as $r \rightarrow \infty$, and (2) becomes infinite slower than $1/r^2$ as $r \rightarrow 0$. The first condition ensures that the potential predominates over the centrifugal term for large r , while the second condition is such that for small r it is the centrifugal term that is important.

The qualitative nature of the motion will be altered if the potential does not satisfy these requirements, but we may still use the method of the equivalent potential to examine features of the orbits. As an example, let us consider the attractive potential

$$V(r) = -\frac{a}{r^3}, \quad \text{with} \quad f = -\frac{3a}{r^4}.$$

The energy diagram is then as shown in Fig. 3.9. For an energy E , there are two possible types of motion, depending upon the initial value of r . If r_0 is less than r_1 the motion will be bounded, r will always remain less than r_1 , and the particle will pass through the center of force. If r is initially greater than r_2 , then it will

*The case $E < E_4$ does not correspond to physically possible motion, for then r^2 would have to be negative, or r imaginary.

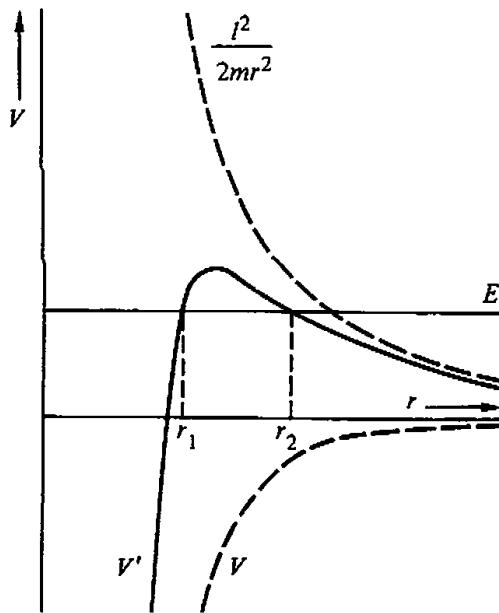


FIGURE 3.9 The equivalent one-dimensional potential for an attractive inverse-fourth law of force

always remain so; the motion is unbounded, and the particle can never get inside the “potential” hole. The initial condition $r_1 < r_0 < r_2$ is again not physically possible.

Another interesting example of the method occurs for a linear restoring force (isotropic harmonic oscillator):

$$f = -kr, \quad V = \frac{1}{2}kr^2.$$

For zero angular momentum, corresponding to motion along a straight line, $V' = V$ and the situation is as shown in Fig. 3.10. For any positive energy the motion is bounded and, as we know, simple harmonic. If $l \neq 0$, we have the state of affairs shown in Fig. 3.11. The motion then is always bounded for all physically possible

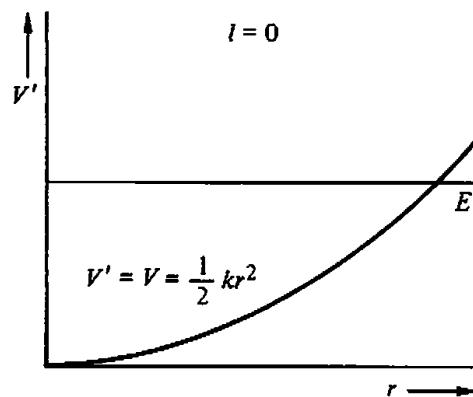


FIGURE 3.10 Effective potential for zero angular momentum.

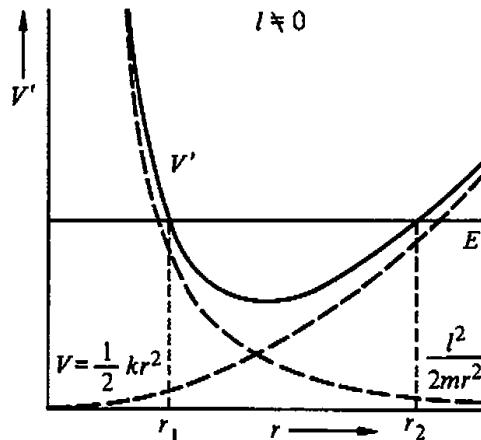


FIGURE 3.11 The equivalent one-dimensional potential for a linear restoring force.

energies and does not pass through the center of force. In this particular case, it is easily seen that the orbit is elliptic, for if $\mathbf{f} = -k\mathbf{r}$, the x - and y -components of the force are

$$f_x = -kx, \quad f_y = -ky.$$

The total motion is thus the resultant of two simple harmonic oscillations at right angles, and of the same frequency, which in general leads to an elliptic orbit.

A well-known example is the spherical pendulum for small amplitudes. The familiar Lissajous figures are obtained as the composition of two sinusoidal oscillations at right angles where the ratio of the frequencies is a rational number. For two oscillations at the same frequency, the figure is a straight line when the oscillations are in phase, a circle when they are 90° out of phase, and an elliptic shape otherwise. Thus, central force motion under a linear restoring force therefore provides the simplest of the Lissajous figures.

3.4 ■ THE VIRIAL THEOREM

Another property of central force motion can be derived as a special case of a general theorem valid for a large variety of systems—the *virial theorem*. It differs in character from the theorems previously discussed in being *statistical* in nature; i.e., it is concerned with the time averages of various mechanical quantities.

Consider a general system of mass points with position vectors \mathbf{r}_i and applied forces \mathbf{F}_i (including any forces of constraint). The fundamental equations of motion are then

$$\dot{\mathbf{p}}_i = \mathbf{F}_i. \quad (1.3)$$

We are interested in the quantity

$$G = \sum_i \mathbf{p}_i \cdot \mathbf{r}_i,$$

where the summation is over all particles in the system. The total time derivative of this quantity is

$$\frac{dG}{dt} = \sum_i \dot{\mathbf{r}}_i \cdot \mathbf{p}_i + \sum_i \dot{\mathbf{p}}_i \cdot \mathbf{r}_i. \quad (3.23)$$

The first term can be transformed to

$$\sum_i \dot{\mathbf{r}}_i \cdot \mathbf{p}_i = \sum_i m_i \dot{\mathbf{r}}_i \cdot \dot{\mathbf{r}}_i = \sum_i m_i v_i^2 = 2T,$$

while the second term by (1.3) is

$$\sum_i \dot{\mathbf{p}}_i \cdot \mathbf{r}_i = \sum_i \mathbf{F}_i \cdot \mathbf{r}_i.$$

Equation (3.23) therefore reduces to

$$\frac{d}{dt} \sum_i \mathbf{p}_i \cdot \mathbf{r}_i = 2T + \sum_i \mathbf{F}_i \cdot \mathbf{r}_i. \quad (3.24)$$

The time average of Eq. (3.24) over a time interval τ is obtained by integrating both sides with respect to t from 0 to τ , and dividing by τ :

$$\frac{1}{\tau} \int_0^\tau \frac{dG}{dt} dt \equiv \overline{\frac{dG}{dt}} = \overline{2T} + \overline{\sum_i \mathbf{F}_i \cdot \mathbf{r}_i}$$

or

$$\overline{2T} + \overline{\sum_i \mathbf{F}_i \cdot \mathbf{r}_i} = \frac{1}{\tau} [G(\tau) - G(0)]. \quad (3.25)$$

If the motion is periodic, i.e., all coordinates repeat after a certain time, and if τ is chosen to be the period, then the right-hand side of (3.25) vanishes. A similar conclusion can be reached even if the motion is not periodic, provided that the coordinates and velocities for all particles remain finite so that there is an upper bound to G . By choosing τ sufficiently long, the right-hand side of Eq. (3.25) can be made as small as desired. In both cases, it then follows that

$$\overline{T} = -\frac{1}{2} \overline{\sum_i \mathbf{F}_i \cdot \mathbf{r}_i}. \quad (3.26)$$

Equation (3.26) is known as the *virial theorem*, and the right-hand side is called the *virial of Clausius*. In this form the theorem is important in the kinetic theory

of gases since it can be used to derive ideal gas law for perfect gases by means of the following brief argument.

We consider a gas consisting of N atoms confined within a container of volume V . The gas is further assumed to be at a Kelvin temperature T (not to be confused with the symbol for kinetic energy). Then by the equipartition theorem of kinetic theory, the average kinetic energy of each atom is given by $\frac{3}{2}k_B T$, k_B being the Boltzmann constant, a relation that in effect is the definition of temperature. The left-hand side of Eq. (3.26) is therefore

$$\frac{3}{2}Nk_B T.$$

On the right-hand side of Eq. (3.26), the forces \mathbf{F}_i include both the forces of interaction between atoms and the forces of constraint on the system. A perfect gas is defined as one for which the forces of interaction contribute negligibly to the virial. This occurs, e.g., if the gas is so tenuous that collisions between atoms occur rarely, compared to collisions with the walls of the container. It is these walls that constitute the constraint on the system, and the forces of constraint, \mathbf{F}_c , are localized at the wall and come into existence whenever a gas atom collides with the wall. The sum on the right-hand side of Eq. (3.26) can therefore be replaced in the average by an integral over the surface of the container. The force of constraint represents the reaction of the wall to the collision forces exerted by the atoms on the wall, i.e., to the pressure P . With the usual outward convention for the unit vector \mathbf{n} in the direction of the normal to the surface, we can therefore write

$$d\mathbf{F}_i = -P\mathbf{n} dA,$$

or

$$\frac{1}{2} \sum_i \mathbf{F}_i \cdot \mathbf{r}_i = -\frac{P}{2} \int \mathbf{n} \cdot \mathbf{r} dA.$$

But, by Gauss's theorem,

$$\int \mathbf{n} \cdot \mathbf{r} dA = \int \nabla \cdot \mathbf{r} dV = 3V.$$

The virial theorem, Eq. (3.26), for the system representing a perfect gas can therefore be written

$$\frac{3}{2}Nk_B T = \frac{3}{2}PV,$$

which, cancelling the common factor of $\frac{3}{2}$ on both sides, is the familiar ideal gas law. Where the interparticle forces contribute to the virial, the perfect gas law of course no longer holds. The virial theorem is then the principal tool, in classical kinetic theory, for calculating the equation of state corresponding to such imperfect gases.

We can further show that if the forces \mathbf{F}_i are the sum of nonfrictional forces \mathbf{F}'_i and frictional forces \mathbf{f}_i proportional to the velocity, then the virial depends only on the \mathbf{F}'_i ; there is no contribution from the \mathbf{f}_i . Of course, the motion of the system must not be allowed to die down as a result of the frictional forces. Energy must constantly be pumped into the system to maintain the motion; otherwise *all* time averages would vanish as τ increases indefinitely (cf. Derivation 1.)

If the forces are derivable from a potential, then the theorem becomes

$$\bar{T} = \frac{1}{2} \sum_i \nabla V \cdot \mathbf{r}_i, \quad (3.27)$$

and for a single particle moving under a central force it reduces to

$$\bar{T} = \frac{1}{2} \frac{\partial V}{\partial r} r. \quad (3.28)$$

If V is a power-law function of r ,

$$V = ar^{n+1},$$

where the exponent is chosen so that the force law goes as r^n , then

$$\frac{\partial V}{\partial r} r = (n+1)V,$$

and Eq. (3.28) becomes

$$\bar{T} = \frac{n+1}{2} \bar{V}. \quad (3.29)$$

By an application of Euler's theorem for homogeneous functions (cf. p. 62), it is clear that Eq. (3.29) also holds whenever V is a homogeneous function in r of degree $n+1$. For the further special case of inverse-square law forces, n is -2 , and the virial theorem takes on a well-known form:

$$\bar{T} = -\frac{1}{2} \bar{V}. \quad (3.30)$$

3.5 ■ THE DIFFERENTIAL EQUATION FOR THE ORBIT, AND INTEGRABLE POWER-LAW POTENTIALS

In treating specific details of actual central force problems, a change in the orientation of our discussion is desirable. Hitherto solving a problem has meant finding r and θ as functions of time with E , l , etc., as constants of integration. But most often what we really seek is the equation of the orbit, i.e., the dependence of r upon θ , eliminating the parameter t . For central force problems, the elimination is particularly simple, since t occurs in the equations of motion only as a variable of differentiation. Indeed, one equation of motion, (3.8), simply provides a definite

relation between a differential change dt and the corresponding change $d\theta$:

$$l dt = mr^2 d\theta. \quad (3.31)$$

The corresponding relation between derivatives with respect to t and θ is

$$\frac{d}{dt} = \frac{l}{mr^2} \frac{d}{d\theta}. \quad (3.32)$$

These relations may be used to convert the equation of motion (3.12) or (3.16) to a differential equation for the orbit. A substitution into Eq. (3.12) gives a second-order differential equation, while a substitution into Eq. (3.17) gives a simpler first-order differential equation.

The substitution into Eq. (3.12) yields

$$\frac{1}{r^2} \frac{d}{d\theta} \left(\frac{1}{mr^2} \frac{dr}{d\theta} \right) - \frac{l^2}{mr^3} = f(r), \quad (3.33)$$

which upon substituting $u = 1/r$ and expressing the results in terms of the potential gives

$$\frac{d^2 u}{d\theta^2} + u = -\frac{m}{l^2} \frac{d}{du} V \left(\frac{1}{u} \right). \quad (3.34)$$

The preceding equation is such that the resulting orbit is symmetric about two adjacent turning points. To prove this statement, note that if the orbit is symmetrical it should be possible to reflect it about the direction of the turning angle without producing any change. If the coordinates are chosen so that the turning point occurs for $\theta = 0$, then the reflection can be effected mathematically by substituting $-\theta$ for θ . The differential equation for the orbit, (3.34), is obviously invariant under such a substitution. Further the initial conditions, here

$$u = u(0), \quad \left(\frac{du}{d\theta} \right)_0 = 0, \quad \text{for } \theta = 0,$$

will likewise be unaffected. Hence, the orbit equation must be the same whether expressed in terms of θ or $-\theta$, which is the desired conclusion. *The orbit is therefore invariant under reflection about the apsidal vectors.* In effect, this means that the complete orbit can be traced if the portion of the orbit between any two turning points is known. Reflection of the given portion about one of the apsidal vectors produces a neighboring stretch of the orbit, and this process can be repeated indefinitely until the rest of the orbit is completed, as illustrated in Fig. 3.12.

For any particular force law, the actual equation of the orbit can be obtained by eliminating t from the solution (3.17) by means of (3.31), resulting in

$$d\theta = \frac{l dr}{mr^2 \sqrt{\frac{2}{m} \left(E - V(r) - \frac{l^2}{2mr^2} \right)}}. \quad (3.35)$$

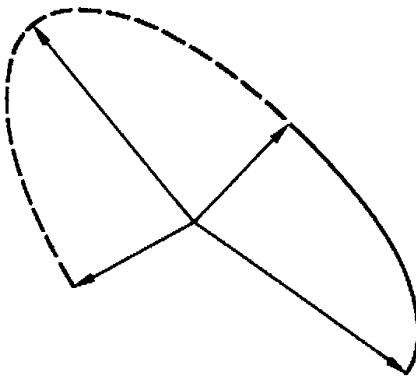


FIGURE 3.12 Extension of the orbit by reflection of a portion about the apsidal vectors.

With slight rearrangements, the integral of (3.35) is

$$\theta = \int_{r_0}^r \frac{dr}{r^2 \sqrt{\frac{2mE}{l^2} - \frac{2mV}{l^2} - \frac{1}{r^2}}} + \theta_0, \quad (3.36)$$

or, if the variable of integration is changed to $u = 1/r$,

$$\theta = \theta_0 - \int_{u_0}^u \frac{du}{\sqrt{\frac{2mE}{l^2} - \frac{2mV}{l^2} - u^2}}. \quad (3.37)$$

As in the case of the equation of motion, Eq. (3.37), while solving the problem formally, is not always a practicable solution, because the integral often cannot be expressed in terms of well-known functions. In fact, only certain types of force laws have been investigated. The most important are the power-law functions of r ,

$$V = ar^{n+1} \quad (3.38)$$

so that the force varies at the n th power of r .* With this potential, (3.37) becomes

$$\theta = \theta_0 - \int_{u_0}^u \frac{du}{\sqrt{\frac{2mE}{l^2} - \frac{2ma}{l^2} u^{-n-1} - u^2}}. \quad (3.39)$$

This again is integrable in terms of simple functions only in certain cases. The particular power-law exponents for which the results can be expressed in terms of trigonometric functions are

$$n = 1, -2, -3.$$

*The case $n = -1$ is to be excluded from the discussion. In the potential (3.38), it corresponds to a constant potential, i.e., no force at all. It is an equally anomalous case if the exponent is used in the force law directly, since a force varying as r^{-1} corresponds to a logarithmic potential, which is not a power law at all. A logarithmic potential is unusual for motion about a point, it is more characteristic of a *line* source. Further details of these cases are given in the second edition of this text.

The results of the integral for

$$n = 5, 3, 0, -4, -5, -7$$

can be expressed in terms of elliptic functions. These are all the possibilities for an integer exponent where the formal integrations are expressed in terms of simple well-known functions. Some fractional exponents can be shown to lead to elliptic functions, and many other exponents can be expressed in terms of the hypergeometric function. The trigonometric and elliptical functions are special cases of generalized hypergeometric function integrals. Equation (3.39) can of course be numerically integrated for any nonpathological potential, but this is beyond the scope of the text.

3.6 ■ CONDITIONS FOR CLOSED ORBITS (BERTRAND'S THEOREM)

We have not yet extracted all the information that can be obtained from the equivalent one-dimensional problem or from the orbit equation without explicitly solving for the motion. In particular, it is possible to derive a powerful and thought-provoking theorem on the types of attractive central forces that lead to *closed orbits*, i.e., orbits in which the particle eventually retraces its own footsteps.

Conditions have already been described for one kind of closed orbit, namely a circle about the center of force. For any given l , this will occur if the equivalent potential $V'(r)$ has a minimum or maximum at some distance r_0 and if the energy E is just equal to $V'(r_0)$. The requirement that V' have an extremum is equivalent to the vanishing of f' at r_0 , leading to the condition derived previously (cf. Section 3.3).

$$f(r_0) = -\frac{l^2}{mr_0^3}, \quad (3.40)$$

which says the force must be attractive for circular orbits to be possible. In addition, the energy of the particle must be given by

$$E = V(r_0) + \frac{l^2}{2mr_0^2}. \quad (3.41)$$

which, by Eq. (3.15), corresponds to the requirement that for a circular orbit \dot{r} is zero. Equations (3.40) and (3.41) are both elementary and familiar. Between them they imply that for any attractive central force it is possible to have a circular orbit at some arbitrary radius r_0 , provided the angular momentum l is given by Eq. (3.40) and the particle energy by Eq. (3.41).

The character of the circular orbit depends on whether the extremum of V' is a minimum, as in Fig. 3.8, or a maximum, as in Fig. 3.9. If the energy is slightly above that required for a circular orbit at the given value of l , then for a minimum in V' the motion, though no longer circular, will still be bounded. However, if

V' exhibits a maximum, then the slightest raising of E above the circular value, Eq. (3.34), results in motion that is unbounded, with the particle moving both through the center of force and out to infinity for the potential shown in Fig. 3.9. Borrowing the terminology from the case of static equilibrium, the circular orbit arising in Fig. 3.8 is said to be *stable*; that in Fig. 3.9 is *unstable*. The stability of the circular orbit is thus determined by the sign of the second derivative of V' at the radius of the circle, being stable for positive second derivative (V' concave up) and unstable for V' concave down. A stable orbit therefore occurs if

$$\frac{\partial^2 V'}{\partial r^2} \Big|_{r=r_0} = -\frac{\partial f}{\partial r} \Big|_{r=r_0} + \frac{3l^2}{mr_0^4} > 0. \quad (3.42)$$

Using Eq. (3.40), this condition can be written

$$\frac{\partial f}{\partial r} \Big|_{r=r_0} < -\frac{3f(r_0)}{r_0}, \quad (3.43)$$

or

$$\frac{d \ln f}{d \ln r} \Big|_{r=r_0} > -3 \quad (3.43')$$

where $f(r_0)/r_0$ is assumed to be negative and given by dividing Eq. (3.40) by r_0 . If the force behaves like a power law of r in the vicinity of the circular radius r_0 ,

$$f = -kr^n,$$

then the stability condition, Eq. (3.43), becomes

$$-knr^{n-1} < 3kr^{n-1}$$

or

$$n > -3, \quad (3.44)$$

where k is assumed to be positive. A power-law attractive potential varying more slowly than $1/r^2$ is thus capable of stable circular orbits for all values of r_0 .

If the circular orbit is stable, then a small increase in the particle energy above the value for a circular orbit results in only a slight variation of r about r_0 . It can be easily shown that for such small deviations from the circularity conditions, the particle executes a simple harmonic motion in $u (\equiv 1/r)$ about u_0 :

$$u = u_0 + a \cos \beta \theta. \quad (3.45)$$

Here a is an amplitude that depends upon the deviation of the energy from the value for circular orbits, and β is a quantity arising from a Taylor series expansion

of the force law $f(r)$ about the circular orbit radius r_0 . Direct substitution into the force law gives

$$\beta^2 = 3 + \frac{r}{f} \frac{df}{dr} \Big|_{r=r_0}. \quad (3.46)$$

As the radius vector of the particle sweeps completely around the plane, u goes through β cycles of its oscillation (cf. Fig. 3.13). If β is a rational number, the ratio of two integers, p/q , then after q revolutions of the radius vector the orbit would begin to retrace itself so that the orbit is *closed*.

At each r_0 such that the inequality in Eq. (3.43) is satisfied, it is possible to establish a stable circular orbit by giving the particle an initial energy and angular momentum prescribed by Eqs. (3.40) and (3.41). The question naturally arises as to what form the force law must take in order that the slightly perturbed orbit about any of these circular orbits should be closed. It is clear that under these conditions β must not only be a rational number, it must also be the *same* rational number at all distances that a circular orbit is possible. Otherwise, since β can take on only discrete values, the number of oscillatory periods would change discontinuously with r_0 , and indeed the orbits could not be closed at the discontinuity. With β^2 everywhere constant, the defining equation for β^2 , Eq. (3.46), becomes in effect a differential equation for the force law f in terms of the independent variable r_0 .

We can indeed consider Eq. (3.46) to be written in terms of r if we keep in mind that the equation is valid only over the ranges in r for which stable circular orbits are possible. A slight rearrangement of Eq. (3.46) leads to the equation

$$\frac{d \ln f}{d \ln r} = \beta^2 - 3, \quad (3.47)$$

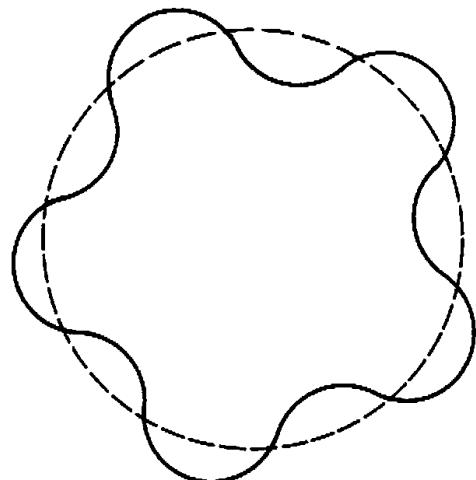


FIGURE 3.13 Orbit for motion in a central force deviating slightly from a circular orbit for $\beta = 5$.

which can be immediately integrated to give a force law:

$$f(r) = -\frac{k}{r^{3-\beta^2}}. \quad (3.48)$$

All force laws of this form, with β a rational number, lead to closed stable orbits for initial conditions that differ only *slightly* from conditions defining a circular orbit. Included within the possibilities allowed by Eq. (3.48) are some familiar forces such as the inverse-square law ($\beta \equiv 1$), but of course many other behaviors, such as $f = -kr^{-2/9}$ ($\beta = \frac{5}{3}$), are also permitted.

Suppose the initial conditions deviate more than slightly from the requirements for circular orbits; will these same force laws still give circular orbits? The question can be answered directly by keeping an additional term in the Taylor series expansion of the force law and solving the resultant orbit equation.

J. Bertrand solved this problem in 1873 and found that for more than first-order deviations from circularity, the orbits are closed only for $\beta^2 = 1$ and $\beta^2 = 4$. The first of these values of β^2 , by Eq. (3.48), leads to the familiar attractive inverse-square law; the second is an attractive force proportional to the radial distance—Hooke's law! These force laws, and only these, could possibly produce closed orbits for any arbitrary combination of l and E ($E < 0$), and in fact we know from direct solution of the orbit equation that they do. Hence, we have Bertrand's theorem: *The only central forces that result in closed orbits for all bound particles are the inverse-square law and Hooke's law.*

This is a remarkable result, well worth the tedious algebra required. It is a commonplace astronomical observation that bound celestial objects move in orbits that are in first approximation closed. For the most part, the small deviations from a closed orbit are traceable to perturbations such as the presence of other bodies. The prevalence of closed orbits holds true whether we consider only the solar system, or look to the many examples of true binary stars that have been observed. Now, Hooke's law is a most unrealistic force law to hold at all distances, for it implies a force increasing indefinitely to infinity. Thus, the existence of closed orbits for a wide range of initial conditions by itself leads to the conclusion that the gravitational force varies as the inverse-square of the distance.

We can phrase this conclusion in a slightly different manner, one that is of somewhat more significance in modern physics. The orbital motion in a plane can be looked on as compounded of two oscillatory motions, one in r and one in θ with the same period. The *character of orbits in a gravitational field fixes the form of the force law*. Later on we shall encounter other formulations of the relation between degeneracy and the nature of the potential.

3.7 ■ THE KEPLER PROBLEM: INVERSE-SQUARE LAW OF FORCE

The inverse-square law is the most important of all the central force laws, and it deserves detailed treatment. For this case, the force and potential can be written

as

$$f = -\frac{k}{r^2} \quad V = -\frac{k}{r}. \quad (3.49)$$

There are several ways to integrate the equation for the orbit, the simplest being to substitute (3.49) in the differential equation for the orbit (3.33). Another approach is to start with Eq. (3.39) with n set equal to -2 for the gravitational force

$$\theta = \theta' - \int \frac{du}{\sqrt{\frac{2mE}{l^2} + \frac{2mk u}{l^2} - u^2}}, \quad (3.50)$$

where the integral is now taken as indefinite. The quantity θ' appearing in (3.50) is a constant of integration determined by the initial conditions and will not necessarily be the same as the initial angle θ_0 at time $t = 0$. The indefinite integral is of the standard form,

$$\int \frac{dx}{\sqrt{\alpha + \beta x + \gamma x^2}} = \frac{1}{\sqrt{-\gamma}} \arccos -\frac{\beta + 2\gamma x}{\sqrt{q}}, \quad (3.51)$$

where

$$q = \beta^2 - 4\alpha\gamma.$$

To apply this to (3.50), we must set

$$\alpha = \frac{2mE}{l^2}, \quad \beta = \frac{2mk}{l^2} \quad \gamma = -1, \quad (3.52)$$

and the discriminant q is therefore

$$q = \left(\frac{2mk}{l^2} \right)^2 \left(1 + \frac{2El^2}{mk^2} \right). \quad (3.53)$$

With these substitutes, Eq. (3.50) becomes

$$\theta = \theta' - \arccos \frac{\frac{l^2 u}{mk} - 1}{\sqrt{1 + \frac{2El^2}{mk^2}}}. \quad (3.54)$$

Finally, by solving for u , $\equiv 1/r$, the equation of the orbit is found to be

$$\frac{1}{r} = \frac{mk}{l^2} \left(1 + \sqrt{1 + \frac{2El^2}{mk^2}} \cos(\theta - \theta') \right). \quad (3.55)$$

The constant of integration θ' can now be identified from Eq. (3.55) as one of the turning angles of the orbit. Note that only three of the four constants of integration appear in the orbit equation; this is always a characteristic property of the orbit. In

effect, the fourth constant locates the initial position of the particle on the orbit. If we are interested solely in the orbit equation, this information is clearly irrelevant and hence does not appear in the answer. Of course, the missing constant has to be supplied if we wish to complete the solution by finding r and θ as functions of time. Thus, if we choose to integrate the conservation theorem for angular momentum,

$$mr^2 d\theta = l dt,$$

by means of (3.55), we must additionally specify the initial angle θ_0 .

Now, the general equation of a conic with one focus at the origin is

$$\frac{1}{r} = C[1 + e \cos(\theta - \theta')], \quad (3.56)$$

where e is the eccentricity of the conic section. By comparison with Eq. (3.55), it follows that the orbit is always a conic section, with the eccentricity

$$e = \sqrt{1 + \frac{2El^2}{mk^2}}. \quad (3.57)$$

The nature of the orbit depends upon the magnitude of e according to the following scheme:

$e > 1.$	$E > 0:$	hyperbola,
$e = 1,$	$E = 0:$	parabola,
$e < 1,$	$E < 0:$	ellipse,
$e = 0,$	$E = -\frac{mk^2}{2l^2}:$	circle.

This classification agrees with the qualitative discussion of the orbits on the energy diagram of the equivalent one-dimensional potential V' . The condition for circular motion appears here in a somewhat different form, but it can easily be derived as a consequence of the previous conditions for circularity. For a circular orbit, T and V are constant in time, and from the virial theorem

$$E \equiv T + V = -\frac{V}{2} + V = \frac{V}{2}.$$

Hence

$$E = -\frac{k}{2r_0}. \quad (3.58)$$

But from Eq. (3.41), the statement of equilibrium between the central force and the "effective force," we can write

$$\frac{k}{r_0^2} = \frac{l^2}{mr_0^3},$$

or

$$r_0 = \frac{l^2}{mk}. \quad (3.59)$$

With this formula for the orbital radius, Eq. (3.58) becomes

$$E = -\frac{mk^2}{2l^2},$$

the above condition for circular motion.

In the case of elliptic orbits, it can be shown the major axis depends solely upon the energy, a theorem of considerable importance in the Bohr theory of the atom. The semimajor axis is one-half the sum of the two apsidal distances r_1 and r_2 (cf. Fig. 3.6). By definition, the radial velocity is zero at these points, and the conservation of energy implies that the apsidal distances are therefore the roots of the equation (cf. Eq. (3.15))

$$E - \frac{l^2}{2mr^2} + \frac{k}{r} = 0,$$

or

$$r^2 + \frac{k}{E}r - \frac{l^2}{2mE} = 0. \quad (3.60)$$

Now, the coefficient of the linear term in a quadratic equation is the negative of the sum of the roots. Hence, the semimajor axis is given by

$$a = \frac{r_1 + r_2}{2} = -\frac{k}{2E}. \quad (3.61)$$

Note that in the circular limit, Eq. (3.61) agrees with Eq. (3.58). In terms of the semimajor axis, the eccentricity of the ellipse can be written

$$e = \sqrt{1 - \frac{l^2}{mka}}, \quad (3.62)$$

(a relation we will have use for in a later chapter). Further, from Eq. (3.62) we have the expression

$$\frac{l^2}{mk} = a(1 - e^2), \quad (3.63)$$

in terms of which the elliptical orbit equation (3.55) can be written

$$r = \frac{a(1 - e^2)}{1 + e \cos(\theta - \theta')}. \quad (3.64)$$

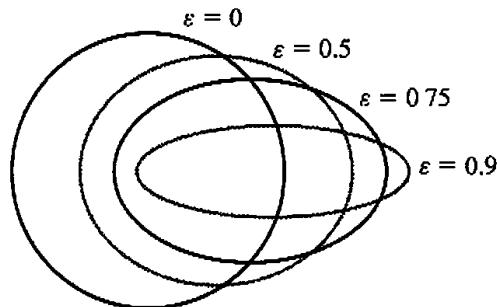


FIGURE 3.14 Ellipses with the same major axes and eccentricities from 0.0 to 0.9.

From Eq. (3.64), it follows that the two apsidal distances (which occur when $\theta - \theta'$ is 0 and π , respectively) are equal to $a(1 - e)$ and $a(1 + e)$, as is to be expected from the properties of an ellipse.

Figure 3.14 shows sketches of four elliptical orbits with the same major axis a , and hence the same energy, but with eccentricities $\varepsilon = 0.0, 0.5, 0.75$, and 0.9 . Figure 3.15 shows how r_1 and r_2 depend on the eccentricity ε .

The velocity vector v_{\parallel} of the particle along the elliptical path can be resolved into a radial component $v_r = \dot{r} = p_r/m$ plus an angular component $v_\theta = r\dot{\theta} = l/mr$

$$v_{\parallel} = v_r \hat{r} + v_\theta \hat{\theta}.$$

The radial component with the magnitude $v_r = \varepsilon v_0 \sin \theta / (1 - \varepsilon^2)$ vanishes at the two apsidal distances, while v_θ attains its maximum value at perihelion and its minimum at aphelion. Table 3.1 lists angular velocity values at the apsidal distances for several eccentricities. Figure 3.16 presents plots of the radial velocity component v_r versus the radius vector r for the half cycle when v_r points outward, i.e., it is positive. During the remaining half cycle v_r is nega-

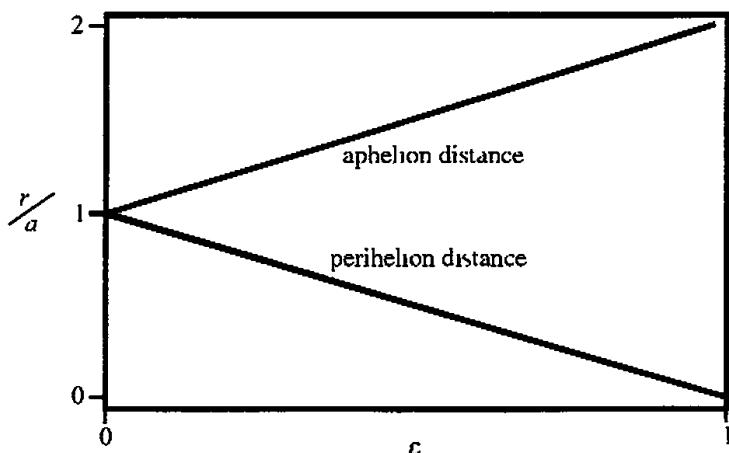


FIGURE 3.15 Dependence of normalized apsidal distances r_1 (lower line) and r_2 (upper line) on the eccentricity ε .

TABLE 3.1 Normalized angular speeds $\dot{\theta}$ and $v_{\theta} = r\dot{\theta}$ at perihelion (r_1) and aphelion (r_2), respectively, in Keplerian orbits of various eccentricities (ε). The normalized radial distances at perihelion and aphelion are listed in columns 2 and 3, respectively. The normalization is with respect to motion in a circle with the radius a and the angular momentum $l = mav_0 = ma^2\dot{\theta}_0$.

Eccentricity ε	Perihelion	Aphelion	Angular speed		Linear angular speed	
	r_1/a	r_2/a	$\dot{\theta}_1/\dot{\theta}_0$	$\dot{\theta}_2/\dot{\theta}_0$	$v_{\theta1}/v_0$	$v_{\theta2}/v_0$
	$1 - \varepsilon$	$1 + \varepsilon$	$\frac{1}{(1 - \varepsilon)^2}$	$\frac{1}{(1 + \varepsilon)^2}$	$\frac{1}{1 - \varepsilon}$	$\frac{1}{1 + \varepsilon}$
0	1	1	1	1	1	1
0.1	0.9	1.1	1.234	0.826	1.111	0.909
0.3	0.7	1.3	2.041	0.592	1.429	0.769
0.5	0.5	1.5	4 000	0.444	2.000	0.667
0.7	0.3	1.7	11.111	0.346	3.333	0.588
0.9	0.1	1.9	100.000	0.277	10.000	0.526

tive, and the plot of Fig. 3.16 repeats itself for the negative range below $v_r = 0$ (not shown). Figure 3.17 shows analogous plots of the angular velocity component v_{θ} versus the angle θ . In these plots and in the table the velocities are normalized relative to the quantities v_0 and $\dot{\theta}_0$ obtained from the expressions $l = mr^2\dot{\theta} = mr v_{\theta} = ma^2\dot{\theta}_0 = mav_0$ for the conservation of angular momentum in the elliptic orbits of semimajor axis a , and in the circle of radius a .

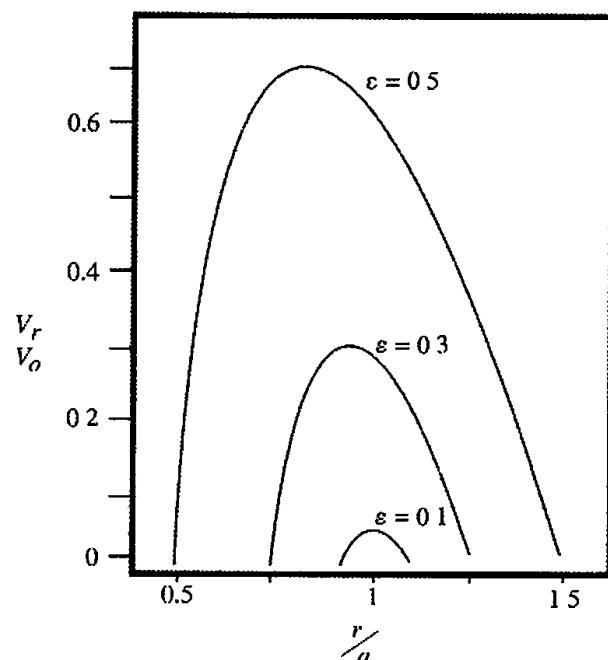


FIGURE 3.16 Normalized radial velocity, v_r , versus r for three values of the eccentricity ε .

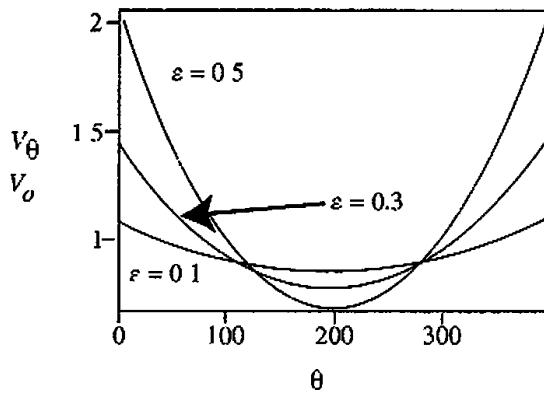


FIGURE 3.17 Normalized orbital velocity, v_θ , versus θ for three values of the eccentricity ϵ .

3.8 ■ THE MOTION IN TIME IN THE KEPLER PROBLEM

The orbital equation for motion in a central inverse-square force law can thus be solved in a fairly straightforward manner with results that can be stated in simple closed expressions. Describing the motion of the particle in time as it traverses the orbit is however a much more involved matter. In principle, the relation between the radial distance of the particle r and the time (relative to some starting point) is given by Eq. (3.18), which here takes on the form

$$t = \sqrt{\frac{m}{2}} \int_{r_0}^r \frac{dr}{\sqrt{\frac{k}{r} - \frac{l^2}{2mr^2} + E}}. \quad (3.65)$$

Similarly, the polar angle θ and the time are connected through the conservation of angular momentum,

$$dt = \frac{mr^2}{l} d\theta,$$

which combined with the orbit equation (3.51) leads to

$$t = \frac{l^3}{mk^2} \int_{\theta_0}^{\theta} \frac{d\theta}{[1 + e \cos(\theta - \theta')]^2}. \quad (3.66)$$

Either of these integrals can be carried out in terms of elementary functions. However, the relations are very complex, and their inversions to give r or θ as functions of t pose formidable problems, especially when one wants the high precision needed for astronomical observations.

To illustrate some of these involvements, let us consider the situation for parabolic motion ($e = 1$), where the integrations can be most simply carried out. It is customary to measure the plane polar angle from the radius vector at

the point of closest approach—a point most usually designated as the *perihe-
lion*.* This convention corresponds to setting θ' in the orbit equation (3.51) equal to zero. Correspondingly, time is measured from the moment, T , of perihelion passage. Using the trigonometric identity

$$1 + \cos \theta = 2 \cos^2 \frac{\theta}{2},$$

Eq. (3.66) then reduces for parabolic motion to the form

$$t = \frac{l^3}{4mk^2} \int_0^\theta \sec^4 \frac{\theta}{2} d\theta.$$

The integration is easily performed by a change of variable to $x = \tan(\theta/2)$, leading to the integral

$$t = \frac{l^3}{2mk^2} \int_0^{\tan(\theta/2)} (1 + x^2) dx,$$

or

$$t = \frac{l^3}{2mk^2} \left(\tan \frac{\theta}{2} + \frac{1}{3} \tan^3 \frac{\theta}{2} \right). \quad (3.67)$$

In this equation, $-\pi < \theta < \pi$, where for $t \rightarrow -\infty$ the particle starts approaching from infinitely far away located at $\theta = -\pi$. The time $t = 0$ corresponds to $\theta = 0$, where the particle is at perihelion. Finally $t \rightarrow +\infty$ corresponds to $\theta \rightarrow \pi$ as the particle moves infinitely far away. This is a straightforward relation for t as a function of θ ; inversion to obtain θ at a given time requires solving a cubic equation for $\tan(\theta/2)$, then finding the corresponding arctan. The radial distance at a given time is given through the orbital equation.

For elliptical motion, Eq. (3.65) is most conveniently integrated through an auxiliary variable ψ , denoted as the *eccentric anomaly*,* and defined by the relation

$$r = a(1 - e \cos \psi). \quad (3.68)$$

By comparison with the orbit equation, (3.64), it is clear that ψ also covers the interval 0 to 2π as θ goes through a complete revolution, and that the perihelion occurs at $\psi = 0$ (where $\theta = 0$ by convention) and the aphelion at $\psi = \pi = \theta$.

*Literally, the term should be restricted to orbits around the Sun, while the more general term should be *periapsis*. However, it has become customary to use perihelion no matter where the center of force is. Even for space craft orbiting the Moon, official descriptions of the orbital parameters refer to perihelion where pericynthion would be the pedantic term.

*Medieval astronomers expected the angular motion to be constant. The angle calculated by multiplying this average angular velocity ($2\pi/\text{period}$) by the time since the last perihelion passage was called the mean anomaly. From the mean anomaly the eccentric anomaly could be calculated and then used to calculate the true anomaly. The angle θ is called the true anomaly just as it was in medieval astronomy.

Expressing E and ℓ in terms of a , e , and k , Eq. (3.65) can be rewritten for elliptic motion as

$$t = -\sqrt{\frac{m}{2k}} \int_{r_0}^r \frac{r dr}{\sqrt{r - \frac{r^2}{2a} - \frac{a(1-e^2)}{2}}}, \quad (3.69)$$

where, by the convention on the starting time, r_0 is the perihelion distance. Substitution of r in terms of ψ from Eq. (3.68) reduces this integral, after some algebra, to the simple form

$$t = \sqrt{\frac{ma^3}{k}} \int_0^\psi (1 - e \cos \psi) d\psi. \quad (3.70)$$

First, we may note that Eq. (3.70) provides an expression for the period, τ , of elliptical motion, if the integral is carried over the full range in ψ of 2π :

$$\tau = 2\pi a^{3/2} \sqrt{\frac{m}{k}}. \quad (3.71)$$

This important result can also be obtained directly from the properties of an ellipse. From the conservation of angular momentum, the areal velocity is constant and is given by

$$\frac{dA}{dt} = \frac{1}{2} r^2 \theta = \frac{l}{2m}. \quad (3.72)$$

The area of the orbit, A , is to be found by integrating (3.72) over a complete period τ :

$$\int_0^\tau \frac{dA}{dt} dt = A = \frac{l\tau}{2m}.$$

Now, the area of an ellipse is

$$A = \pi ab,$$

where, by the definition of eccentricity, the semiminor axis b is related to a according to the formula

$$b = a\sqrt{1 - e^2}.$$

By (3.62), the semiminor axis can also be written as

$$b = a^{1/2} \sqrt{\frac{l^2}{mk}},$$

and the period is therefore

$$\tau = \frac{2m}{l} \pi a^{3/2} \sqrt{\frac{l^2}{mk}} = 2\pi a^{3/2} \sqrt{\frac{m}{k}},$$

as was found previously. Equation (3.71) states that, other things being equal, the square of the period is proportional to the cube of the major axis, and this conclusion is often referred to as the third of Kepler's laws.* Actually, Kepler was concerned with the specific problem of planetary motion in the gravitational field of the Sun. A more precise statement of this third law would therefore be: *The square of the periods of the various planets are proportional to the cube of their major axes.* In this form, the law is only approximately true. Recall that the motion of a planet about the Sun is a two-body problem and m in (3.71) must be replaced by the reduced mass: (cf. Eq. (3.4))

$$\mu = \frac{m_1 m_2}{m_1 + m_2},$$

where m_1 may be taken as referring to the planet and m_2 to the Sun. Further, the gravitational law of attraction is

$$f = -G \frac{m_1 m_2}{r^2},$$

so that the constant k is

$$k = G m_1 m_2. \quad (3.73)$$

Under these conditions, (3.71) becomes

$$\tau = \frac{2\pi a^{3/2}}{\sqrt{G(m_1 + m_2)}} \approx \frac{2\pi a^{3/2}}{\sqrt{Gm_2}}, \quad (3.74)$$

if we neglect the mass of the planet compared to the Sun. It is the approximate version of Eq. (3.74) that is Kepler's third law, for it states that τ is proportional to $a^{3/2}$, with the same constant of proportionality for all planets. However, the planetary mass m_1 is not always completely negligible compared to the Sun's; for example, Jupiter has a mass of about 0.1% of the mass of the Sun. On the other hand, Kepler's third law is rigorously true for the electron orbits in the Bohr atom, since μ and k are then the same for all orbits in a given atom.

To return to the general problem of the position in time for an elliptic orbit, we may rewrite Eq. (3.70) slightly by introducing the frequency of revolution ω as

*Kepler's three laws of planetary motion, published around 1610, were the result of his pioneering analysis of planetary observations and laid the groundwork for Newton's great advances. The second law, the conservation of areal velocity, is a general theorem for central force motion, as has been noted previously. However, the first—that the planets move in elliptical orbits about the Sun at one focus—and the third are restricted specifically to the inverse-square law of force.

$$\omega = \frac{2\pi}{\tau} = \sqrt{\frac{k}{ma^3}}. \quad (3.75)$$

The integration in Eq. (3.70) is of course easily performed, resulting in the relation

$$\omega t = \psi - e \sin \psi, \quad (3.76)$$

known as *Kepler's equation*. The quantity ωt goes through the range 0 to 2π , along with ψ and θ , in the course of a complete orbital revolution and is therefore also denoted as an anomaly, specifically the *mean anomaly*.

To find the position in orbit at a given time t , Kepler's equation, (3.76), would first be inverted to obtain the corresponding eccentric anomaly ψ . Equation (3.68) then yields the radial distance, while the polar angle θ can be expressed in terms of ψ by comparing the defining equation (3.68) with the orbit equation (3.64):

$$1 + e \cos \theta = \frac{1 - e^2}{1 - e \cos \psi}.$$

With a little algebraic manipulation, this can be simplified, to

$$\cos \theta = \frac{\cos \psi - e}{1 - e \cos \psi}. \quad (3.77)$$

By successively adding and subtracting both sides of Eq. (3.77) from unity and taking the ratio of the resulting two equations, we are led to the alternative form

$$\tan \frac{\theta}{2} = \sqrt{\frac{1+e}{1-e}} \tan \frac{\psi}{2}. \quad (3.78)$$

Either Eq. (3.77) or (3.78) thus provides θ , once ψ is known. The solution of the transcendental Kepler's equation (3.76) to give the value of ψ corresponding to a given time is a problem that has attracted the attention of many famous mathematicians ever since Kepler posed the question early in the seventeenth century. Newton, for example, contributed what today would be called an analog solution. Indeed, it can be claimed that the practical need to solve Kepler's equation to accuracies of a second of arc over the whole range of eccentricity fathered many of the developments in numerical mathematics in the eighteenth and nineteenth centuries. A few of the more than 100 methods of solution developed in the pre-computer era are considered in the exercises to this chapter.

3.9 ■ THE LAPLACE–RUNGE–LENZ VECTOR

The Kepler problem is also distinguished by the existence of an additional conserved vector besides the angular momentum. For a general central force, New-

ton's second law of motion can be written vectorially as

$$\dot{\mathbf{p}} = f(r) \frac{\mathbf{r}}{r}. \quad (3.79)$$

The cross product of $\dot{\mathbf{p}}$ with the constant angular momentum vector \mathbf{L} therefore can be expanded as

$$\begin{aligned} \dot{\mathbf{p}} \times \mathbf{L} &= \frac{mf(r)}{r} [\mathbf{r} \times (\mathbf{r} \times \dot{\mathbf{r}})] \\ &= \frac{mf(r)}{r} [\mathbf{r}(\mathbf{r} \cdot \dot{\mathbf{r}}) - r^2 \dot{\mathbf{r}}]. \end{aligned} \quad (3.80)$$

Equation (3.80) can be further simplified by noting that

$$\mathbf{r} \cdot \dot{\mathbf{r}} = \frac{1}{2} \frac{d}{dt} (\mathbf{r} \cdot \mathbf{r}) = r\dot{r}$$

(or, in less formal terms, the component of the velocity in the radial direction is \dot{r}). As \mathbf{L} is constant, Eq. (3.80) can then be rewritten, after a little manipulation, as

$$\frac{d}{dt} (\mathbf{p} \times \mathbf{L}) = -mf(r)r^2 \left(\frac{\dot{\mathbf{r}}}{r} - \frac{\mathbf{r}\dot{r}}{r^2} \right).$$

or

$$\frac{d}{dt} (\mathbf{p} \times \mathbf{L}) = -mf(r)r^2 \frac{d}{dt} \left(\frac{\mathbf{r}}{r} \right). \quad (3.81)$$

Without specifying the form of $f(r)$, we can go no further. But Eq. (3.81) can be immediately integrated if $f(r)$ is inversely proportional to r^2 —the Kepler problem. Writing $f(r)$ in the form prescribed by Eq. (3.49), Eq. (3.81) then becomes

$$\frac{d}{dt} (\mathbf{p} \times \mathbf{L}) = \frac{d}{dt} \left(\frac{mkr}{r} \right),$$

which says that for the Kepler problem there exists a *conserved vector* \mathbf{A} defined by

$$\mathbf{A} = \mathbf{p} \times \mathbf{L} - mk \frac{\mathbf{r}}{r}. \quad (3.82)$$

The relationships between the three vectors in Eq. (3.82) and the conservation of \mathbf{A} are illustrated in Fig. 3.18, which shows the three vectors at different positions in the orbit. In recent times, the vector \mathbf{A} has become known amongst physicists as the Runge–Lenz vector, but priority belongs to Laplace.

From the definition of \mathbf{A} , we can easily see that

$$\mathbf{A} \cdot \mathbf{L} = 0, \quad (3.83)$$

since \mathbf{L} is perpendicular to $\mathbf{p} \times \mathbf{L}$ and \mathbf{r} is perpendicular to $\mathbf{L} = \mathbf{r} \times \mathbf{p}$. It follows from this orthogonality of \mathbf{A} to \mathbf{L} that \mathbf{A} must be some fixed vector in the plane of

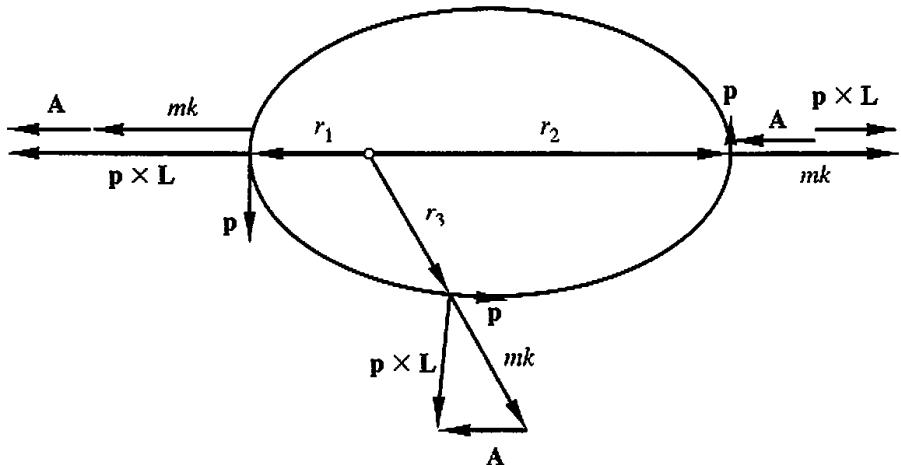


FIGURE 3.18 The vectors \mathbf{p} , \mathbf{L} , and \mathbf{A} at three positions in a Keplerian orbit. At perihelion (extreme left) $|\mathbf{p} \times \mathbf{L}| = mk(1+e)$ and at aphelion (extreme right) $|\mathbf{p} \times \mathbf{L}| = mk(1-e)$. The vector \mathbf{A} always points in the same direction with a magnitude mke .

the orbit. If θ is used to denote the angle between \mathbf{r} and the fixed direction of \mathbf{A} , then the dot product of \mathbf{r} and \mathbf{A} is given by

$$\mathbf{A} \cdot \mathbf{r} = Ar \cos \theta = \mathbf{r} \cdot (\mathbf{p} \times \mathbf{L}) - mkr. \quad (3.84)$$

Now, by permutation of the terms in the triple dot product, we have

$$\mathbf{r} \cdot (\mathbf{p} \times \mathbf{L}) = \mathbf{L} \cdot (\mathbf{r} \times \mathbf{p}) = l^2,$$

so that Eq. (3.84) becomes

$$Ar \cos \theta = l^2 - mkr,$$

or

$$\frac{l}{r} = \frac{mk}{l^2} \left(1 + \frac{\mathbf{A}}{mk} \cos \theta \right). \quad (3.85)$$

The Laplace–Runge–Lenz vector thus provides still another way of deriving the orbit equation for the Kepler problem! Comparing Eq. (3.85) with the orbit equation in the form of Eq. (3.55) shows that \mathbf{A} is in the direction of the radius vector to the perihelion point on the orbit, and has a magnitude

$$\mathbf{A} = mke. \quad (3.86)$$

For the Kepler problem we have thus identified two vector constants of the motion \mathbf{L} and \mathbf{A} , and a scalar E . Since a vector must have all three independent components, this corresponds to seven conserved quantities in all. Now, a system such as this with three degrees of freedom has six independent constants of the motion, corresponding, say to the three components of both the initial position

and the initial velocity of the particle. Further, the constants of the motion we have found are all algebraic functions of \mathbf{r} and \mathbf{p} that describe the orbit as a whole (orientation in space, eccentricity, etc.); none of these seven conserved quantities relate to where the particle is located in the orbit at the initial time. Since one constant of the motion must relate to this information, say in the form of T , the time of the perihelion passage, there can be only five independent constants of the motion describing the size, shape, and orientation of the orbit. We can therefore conclude that not all of the quantities making up \mathbf{L} , \mathbf{A} , and E can be independent; there must in fact be two relations connecting these quantities. One such relation has already been obtained as the orthogonality of \mathbf{A} and \mathbf{L} , Eq. (3.83). The other follows from Eq. (3.86) when the eccentricity is expressed in terms of E and l from Eq. (3.57), leading to

$$\mathbf{A}^2 = m^2 k^2 + 2mEl^2, \quad (3.87)$$

thus confirming that there are only five *independent* constants out of the seven.

The angular momentum vector and the energy alone contain only four independent constants of the motion: The Laplace–Runge–Lenz vector thus adds one more. It is natural to ask why there should not exist for any general central force law some conserved quantity that together with \mathbf{L} and E serves to define the orbit in a manner similar to the Laplace–Runge–Lenz vector for the special case of the Kepler problem. The answer seems to be that such conserved quantities can in fact be constructed, but that they are in general rather peculiar functions of the motion. The constants of the motion relating to the orbit between them define the orbit, i.e., lead to the orbit equation giving r as a function of θ . We have seen that in general orbits for central force motion are not closed; the arguments of Section 3.6 show that closed orbits imply rather stringent conditions on the form of the force law. It is a property of nonclosed orbits that the curve will eventually pass through any arbitrary (r, θ) point that lies between the bounds of the turning points of r . Intuitively this can be seen from the nonclosed nature of the orbit; as θ goes around a full cycle, the particle must never retrace its footsteps on any previous orbit. Thus, the orbit equation is such that r is a multivalued function of θ (modulo 2π); in fact, it is an *infinite-valued function* of θ . The corresponding conserved quantity additional to \mathbf{L} and E defining the orbit must similarly involve an infinite-valued function of the particle motion. Suppose the \mathbf{r} variable is periodic with angular frequency ω_r and the angular coordinate θ is periodic with angular frequency ω_θ . If these two frequencies have a ratio (ω_r/ω_θ) that is an integer or integer fraction, periods are said to be *commensurate*. Commensurate orbits are closed with the orbiting mass continually retracing its path. When $\omega_\theta > \omega_r$, the orbit will spiral about the origin as the distance varies between the apsidal (maximum and minimum) values, closing only if the frequencies are commensurate. If, as in the Kepler problem, $\omega_r = \omega_e$, the periods are said to be degenerate. If the orbits are degenerate there exists an additional conserved quantity that is an algebraic function of \mathbf{r} and \mathbf{p} , such as the Runge–Lenz vector.

From these arguments we would expect a simple analog of such a vector to exist for the case of a Hooke's law force, where, as we have seen, the orbits are

also degenerate. This is indeed the case, except that the natural way to formulate the constant of the motion leads not to a vector but to a tensor of the second rank (cf. Section 7.5). Thus, the existence of an additional constant or integral of the motion, beyond E and \mathbf{L} , that is a simple algebraic function of the motion is sufficient to indicate that the motion is degenerate and the bounded orbits are closed.

3.10 ■ SCATTERING IN A CENTRAL FORCE FIELD

Historically, the interest in central forces arose out of the astronomical problems of planetary motion. There is no reason, however, why central force motion must be thought of only in terms of such problems; mention has already been made of the orbits in the Bohr atom. Another field that can be investigated in terms of classical mechanics is the *scattering* of particles by central force fields. Of course, if the particles are on the atomic scale, it must be expected that the specific results of a classical treatment will often be incorrect physically, for quantum effects are usually large in such regions. Nevertheless, many classical predictions remain valid to a good approximation. More important, the procedures for *describing* scattering phenomena are the same whether the mechanics is classical or quantum; we can learn to speak the language equally as well on the basis of classical physics.

In its one-body formulation, the scattering problem is concerned with the scattering of particles by a *center of force*. We consider a uniform beam of particles—whether electrons, or α -particles, or planets is irrelevant—all of the same mass and energy incident upon a center of force. It will be assumed that the force falls off to zero for very large distances. The incident beam is characterized by specifying its *intensity* I (also called flux density), which gives the number of particles crossing unit area normal to the beam in unit time. As a particle approaches the center of force, it will be either attracted or repelled, and its orbit will deviate from the incident straight-line trajectory. After passing the center of force, the force acting on the particle will eventually diminish so that the orbit once again approaches a straight line. In general, the final direction of motion is not the same as the incident direction, and the particle is said to be scattered. The *cross section for scattering in a given direction*, $\sigma(\Omega)$, is defined by

$$\sigma(\Omega) d\Omega = \frac{\text{number of particles scattered into solid angle } d\Omega \text{ per unit time}}{\text{incident intensity}}, \quad (3.88)$$

where $d\Omega$ is an element of solid angle in the direction Ω . Often $\sigma(\Omega)$ is also designated as the *differential scattering cross section*. With central forces there must be complete symmetry around the axis of the incident beam; hence the element of solid angle can be written

$$d\Omega = 2\pi \sin \Theta d\Theta, \quad (3.89)$$

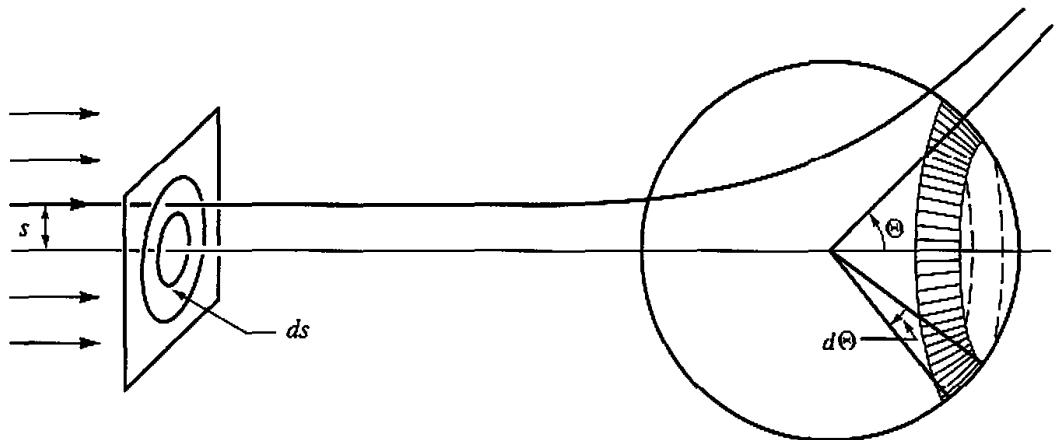


FIGURE 3.19 Scattering of an incident beam of particles by a center of force.

where Θ is the angle between the scattered and incident directions, known as the *scattering angle* (cf. Fig. 3.19, where repulsive scattering is illustrated). Note that the name “cross section” is deserved in that $\sigma(\Omega)$ has the dimensions of an area.

For any given particle the constants of the orbit, and hence the amount of scattering, are determined by its energy and angular momentum. It is convenient to express the angular momentum in terms of the energy and a quantity known as the *impact parameter*, s , defined as the perpendicular distance between the center of force and the incident velocity. If v_0 is the incident speed of the particle, then

$$l = mv_0 s = s\sqrt{2mE}. \quad (3.90)$$

Once E and s are fixed, the angle of scattering Θ is then determined uniquely.* For the moment, it will be assumed that different values of s cannot lead to the same scattering angle. Therefore, the number of particles scattered into a solid angle $d\Omega$ lying between Θ and $\Theta + d\Theta$ must be equal to the number of the incident particles with impact parameter lying between the corresponding s and $s + ds$:

$$2\pi Is|ds| = 2\pi\sigma(\Theta)I \sin \Theta |d\Theta|. \quad (3.91)$$

Absolute value signs are introduced in Eq. (3.91) because numbers of particles must of course always be positive, while s and Θ often vary in opposite directions. If s is considered as a function of the energy and the corresponding scattering angle,

$$s = s(\Theta, E), \quad (3.92)$$

*It is at this point in the formulation that classical and quantum mechanics part company. Indeed, it is fundamentally characteristic of quantum mechanics that we cannot unequivocally predict the trajectory of any particular particle. We can only give probabilities for scattering in various directions.

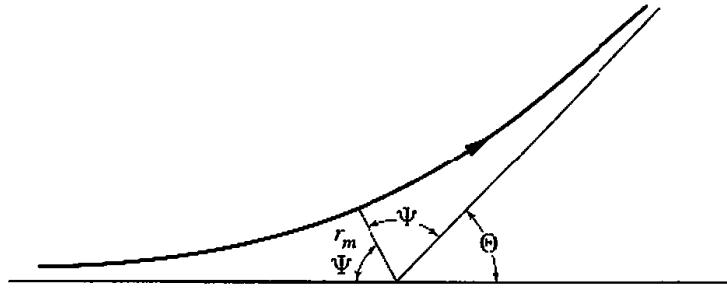


FIGURE 3.20 Relation of orbit parameters and scattering angle in an example of repulsive scattering.

then the dependence of the differential cross section on Θ is given by

$$\sigma(\Theta) = \frac{s}{\sin \Theta} \left| \frac{ds}{d\Theta} \right|. \quad (3.93)$$

A formal expression for the scattering angle Θ as a function of s can be directly obtained from the orbit equation, Eq. (3.36). Again, for simplicity, we will consider the case of purely repulsive scattering (cf. Fig. 3.20). As the orbit must be symmetric about the direction of the periapsis, the scattering angle is given by

$$\Theta = \pi - 2\Psi, \quad (3.94)$$

where Ψ is the angle between the direction of the incoming asymptote and the periapsis (closest approach) direction. In turn, Ψ can be obtained from Eq. (3.36) by setting $r_0 = \infty$ when $\theta_0 = \pi$ (the incoming direction), whence $\theta = \pi - \Psi$ when $r = r_m$, the distance of closest approach. A trivial rearrangement then leads to

$$\Psi = \int_{r_m}^{\infty} \frac{dr}{r^2 \sqrt{\frac{2mE}{l^2} - \frac{2mV}{l^2} - \frac{1}{r^2}}}. \quad (3.95)$$

Expressing l in terms of the impact parameter s (Eq. (3.90)), the resultant expression for $\Theta(s)$ is

$$\Theta(s) = \pi - 2 \int_{r_m}^{\infty} \frac{s dr}{r \sqrt{r^2 \left(1 - \frac{V(r)}{E}\right) - s^2}}. \quad (3.96)$$

or, changing r to $1/u$

$$\Theta(s) = \pi - 2 \int_0^{u_m} \frac{s du}{\sqrt{1 - \frac{V(u)}{E} - s^2 u^2}}. \quad (3.97)$$

Equations (3.96) and (3.97) are rarely used except for direct numerical computation of the scattering angle. However, when an analytic expression is available for the orbits, the relation between Θ and s can often be obtained almost by inspection. An historically important illustration of such a procedure is the repulsive scattering of charged particles by a Coulomb field. The scattering force field is that produced by a fixed charge $-Ze$ acting on the incident particles having a charge $-Z'e$ so that the force can be written as

$$f = \frac{ZZ'e^2}{r^2},$$

i.e., a repulsive inverse-square law. The results of Section 3.7 can be taken over here with no more change than writing the force constant as

$$k = -ZZ'e^2. \quad (3.98)$$

The energy E is greater than zero, and the orbit is a hyperbola with the eccentricity given by^x

$$\epsilon = \sqrt{1 + \frac{2El^2}{m(ZZ'e^2)^2}} = \sqrt{1 + \left(\frac{2Es}{ZZ'e}\right)^2}, \quad (3.99)$$

where use has been made of Eq. (3.90). If θ' in Eq. (3.55) is chosen to be π , periapsis corresponds to $\theta = 0$ and the orbit equation becomes

$$\frac{1}{r} = \frac{mZZ'e}{l^2}(\epsilon \cos \theta - 1). \quad (3.100)$$

This hyperbolic orbit equation has the same form as the elliptic orbit equation (3.56) except for a change in sign. The direction of the incoming asymptote, Ψ , is then determined by the condition $r \rightarrow \infty$:

$$\cos \Psi = \frac{1}{\epsilon}$$

or, by Eq. (3.94),

$$\sin \frac{\Theta}{2} = \frac{1}{\epsilon}.$$

Hence,

$$\cot^2 \frac{\Theta}{2} = \epsilon^2 - 1,$$

and using Eq. (3.99)

^xTo avoid confusion with the electron charge e , the eccentricity will temporarily be denoted by ϵ .

$$\cot \frac{\Theta}{2} = \frac{2Es}{ZZ'e}.$$

The desired functional relationship between the impact parameter and the scattering angle is therefore

$$s = \frac{ZZ'e^2}{2E} \cot \frac{\Theta}{2}, \quad (3.101)$$

so that on carrying through the manipulation required by Eq. (3.93), we find that $\sigma(\Theta)$ is given by

$$\sigma(\Theta) = \frac{1}{4} \left(\frac{ZZ'e^2}{2E} \right)^2 \csc^4 \frac{\Theta}{2}. \quad (3.102)$$

Equation (3.102) gives the famous Rutherford scattering cross section, originally derived by Rutherford for the scattering of α particles by atomic nuclei. Quantum mechanics in the nonrelativistic limit yields a cross section identical with this classical result.

In atomic physics, the concept of a *total scattering cross section* σ_T , defined as

$$\sigma_T = \int_{4\pi} \sigma(\Omega) d\Omega = 2\pi \int_0^\pi \sigma(\Theta) \sin \Theta d\Theta.$$

is of considerable importance. However, if we attempt to calculate the total cross section for Coulomb scattering by substituting Eq. (3.102) in this definition, we obtain an infinite result! The physical reason behind this behavior is not difficult to discern. From its definition the total cross section is the number of particles scattered in all directions per unit time for unit incident intensity. Now, the Coulomb field is an example of a “long-range” force; its effects extend to infinity. The very small deflections occur only for particles with very large impact parameters. Hence, all particles in an incident beam of infinite lateral extent will be scattered to some extent and must be included in the total scattering cross section. It is therefore clear that the infinite value for σ_T is not peculiar to the Coulomb field; it occurs in classical mechanics whenever the scattering field is different from zero at all distances, no matter how large.* Only if the force field “cuts off,” i.e., is zero beyond a certain distance, will the scattering cross section be finite. Physically, such a cut-off occurs for the Coulomb field of a nucleus as a result of the presence of the atomic electrons, which “screen” the nucleus and effectively cancel its charge outside the atom.

* σ_T is also infinite for the Coulomb field in quantum mechanics, since it has been stated that Eq. (3.102) remains valid there. However, not all “long-range” forces give rise to infinite total cross sections in quantum mechanics. It turns out that all potentials that fall off faster at larger distances than $1/r^2$ produce a finite quantum-mechanical total scattering cross section

In Rutherford scattering, the scattering angle Θ is a smooth monotonic function of the impact parameter s . From Eq. (3.101) we see that as s decreases from infinity, Θ increases monotonically from zero, reaching the value π as s goes to zero. However, other types of behavior are possible in classical systems, requiring some modification in the prescription, Eq. (3.93), for the classical cross section. For example, with a repulsive potential and particle energy qualitatively of the nature shown in Fig. 3.21(a), it is easy to see physically that the curve of Θ versus s may behave as indicated in Fig. 3.21(b). Thus, with very large values of the impact parameter, as noted above, the particle always remains at large radial distances from the center of force and suffers only minor deflection. At the other extreme, for $s = 0$, the particle travels in a straight line into the center of force, and if the energy is greater than the maximum of the potential, it will continue on through the center without being scattered at all. Hence, for both limits in s , the scattering angle goes to zero. For some intermediate value of s , the scattering angle must pass through a maximum Θ_m . When $\Theta < \Theta_m$, there will be two values of s that can give rise to the same scattering angle. Each will contribute to the scattering cross section at that angle, and Eq. (3.93) should accordingly be modified to the form

$$\sigma(\Theta) = \sum_i \frac{s_i}{\sin \Theta} \left| \frac{ds}{d\Theta} \right|_i, \quad (3.103)$$

where for $\Theta \neq \Theta_m$ the index i takes on the values 1 and 2. Here the subscript i distinguishes the various values of s giving rise to the same value of Θ .

Of particular interest is the cross section at the maximum angle of scattering Θ_m . As the derivative of Θ with respect to s vanishes at this angle, it follows from Eq. (3.93) or (3.103) that the cross section must become infinite at $\Theta \rightarrow \Theta_m$. But for all larger angles the cross section is zero, since the scattering angle cannot exceed Θ_m . The phenomenon of the infinite rise of the cross section followed by abrupt disappearance is very similar to what occurs in the geometrical optics of the scattering of sunlight by raindrops. On the basis of this similarity, the phenomenon is called *rainbow scattering*.

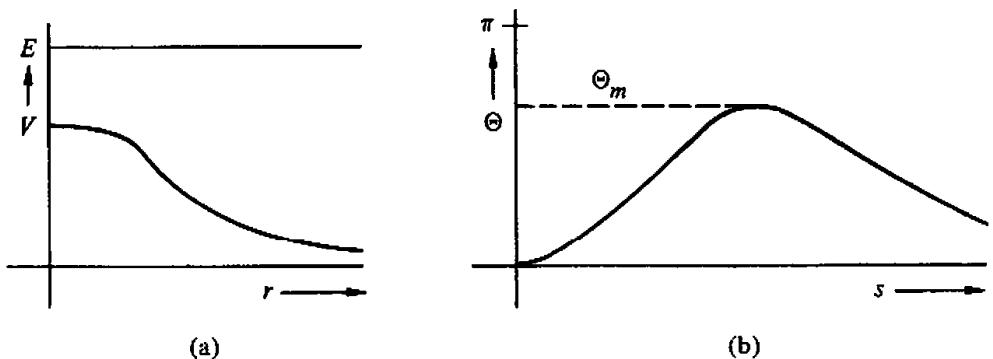


FIGURE 3.21 Repulsive nonsingular scattering potential and double-valued curve of scattering angle Θ versus impact parameter s_0 for sufficiently high energy.

So far, the examples have been for purely repulsive scattering. If the scattering involves attractive forces, further complications may arise. The effect of attraction will be to pull the particle in toward the center instead of the repulsive deflection outward shown in Fig. 3.20. In consequence, the angle Ψ between the incoming direction and the periapsis direction may be greater than $\pi/2$, and the scattering angle as given by Eq. (3.94) is then negative. This in itself is no great difficulty as clearly it is the magnitude of Θ that is involved in finding the cross section. But, under circumstances Θ as calculated by Eq. (3.96) may be greater than 2π . That is, the particle undergoing scattering may circle the center of force for one or more revolutions before going off finally in the scattered direction.

To see how this may occur physically, consider a scattering potential shown as the $s = 0$ curve in Fig. 3.22. It is typical of the intermolecular potentials assumed in many kinetic theory problems—an attractive potential at large distances falling off more rapidly than $1/r^2$, with a rapidly rising repulsive potential at small distances. The other curves in Fig. 3.22 show the effective one-dimensional potential $V'(r)$, Eq. (3.22'), for various values of the impact parameter s (equivalently various values of I). Since the repulsive centrifugal barrier dominates at large r for all values of $s > 0$, the equivalent potential for small s will exhibit a hump.

Now let us consider an incoming particle with impact parameter s_1 and at the energy E_1 corresponding to the maximum of the hump. As noted in Section 3.3, the difference between E_1 and $V'(r)$ is proportional to the square of the radial velocity at that distance. When the incoming particle reaches r_1 , the location of the maximum in V' , the radial velocity is zero. Indeed, recall from the discussion

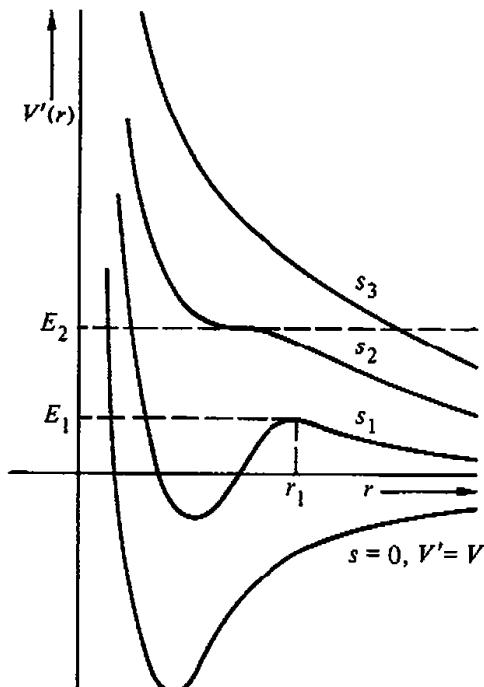


FIGURE 3.22 A combined attractive and repulsive scattering potential, and the corresponding equivalent one-dimensional potential at several values of the impact parameter s .

in Section 3.6 that we have here the conditions for an unstable circular orbit at the distance r_1 . In the absence of any perturbation, the incoming particle with parameters E_1 and s_1 , once having reached r , would circle around the center of force indefinitely at that distance without ever emerging! For the same impact parameter but at an energy E slightly higher than E_1 , no true circular orbit would be established. However, when the particle is in the immediate vicinity of r_1 the radial speed would be very small, and the particle would spend a disproportionately large time in the neighbourhood of the hump. The angular velocity, $\dot{\theta}$, meanwhile would not be affected by the existence of a maximum, being given at r , by (3.90)

$$\dot{\theta} = \frac{l}{mr_1^2} = \frac{s_1}{r_1^2} \sqrt{\frac{2E}{m}}.$$

Thus, in the time it takes the particle to get through the region of the hump, the angular velocity may have carried the particle through angles larger than 2π or even multiples thereof. In such instances, the classical scattering is said to exhibit *orbiting* or *spiraling*.

As the impact parameter is increased, the well and hump in the equivalent potential V' tend to flatten out, until at some parameter s_2 there is only a point of inflection in V' at an energy E_2 (cf. Fig. 3.22). For particle energies above E_2 , there will no longer be orbiting. But the combined effects of the attractive and repulsive components of the effective potential can lead even in such cases to zero deflection for some finite value of the impact parameter. At large energies and small impact parameters, the major scattering effects are caused by the strongly repulsive potentials at small distances, and the scattering qualitatively resembles the behavior of Rutherford scattering.

We have seen that the scattered particle may be deflected by more than π when orbiting takes place. On the other hand, the observed scattering angle in the laboratory lies between 0 and π . It is therefore helpful in such ambiguous cases to distinguish between a *deflection angle* Φ , as calculated by the right-hand sides of Eqs. (3.96) or (3.97), and the observed scattering angle Θ . For given Φ , the angle Θ is to be determined from the relation

$$\Theta = \pm\Phi - 2m\pi, \quad m \text{ a positive integer.}$$

The sign and the value of m are to be chosen so that Θ lies between 0 and π . The sum in Eq. (3.103) then covers all values of Φ leading to the same Θ . Figure 3.23 shows curves of Θ versus s for the potential of Fig. 3.22 at two different energies. The orbiting that takes place for $E = E_1$ shows up as a singularity in the curve at $s = s_1$. When $E > E_2$, orbiting no longer takes place, but there is a rainbow effect at $\Theta = -\Phi'$ (although there is a nonvanishing cross section at higher scattering angles). Note that Θ vanishes at $s = s_3$, which means from Eq. (3.93) that the cross section becomes infinite in the forward direction through the vanishing of $\sin \Theta$. The cross section can similarly become infinite in the backward direction

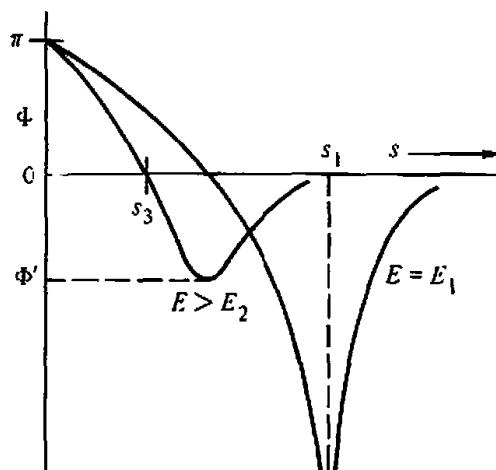


FIGURE 3.23 Curves of deflection angle Φ versus s , for the potential of Fig. 3.22 at two different energies.

providing

$$s \left| \frac{ds}{d\Theta} \right|$$

remains finite at $\Theta = \pi$. These infinities in the forward or backward scattering angles are referred to as *glory scattering*, again in analogy to the corresponding phenomenon in meteorological optics.*

A more general treatment would involve quantum corrections, but in some instances quantum effects are small, as in the scattering of low-energy ions in crystal lattices, and the classical calculations are directly useful. Even when quantum-mechanical corrections are important, it often suffices to use an approximation method (the “semiclassical” approximation) for which a knowledge of the classical trajectory is required. For almost all potentials of practical interest, it is impossible to find an analytic form for the orbit, and Eq. (3.96) (or variant forms) is either approximated for particular regions of s or integrated numerically.

3.11 ■ TRANSFORMATION OF THE SCATTERING PROBLEM TO LABORATORY COORDINATES

In the previous section we were concerned with the one-body problem of the scattering of a particle by a fixed center of force. In practice, the scattering always involved two bodies; e.g., in Rutherford scattering we have the α particle and the atomic nucleus. The second particle, m_2 , is not fixed but recoils from its initial position as a result of the scattering. Since it has been shown that any two-body

*The backward glory is familiar to airplane travelers as the ring of light observed to encircle the shadow of the plane projected on clouds underneath.

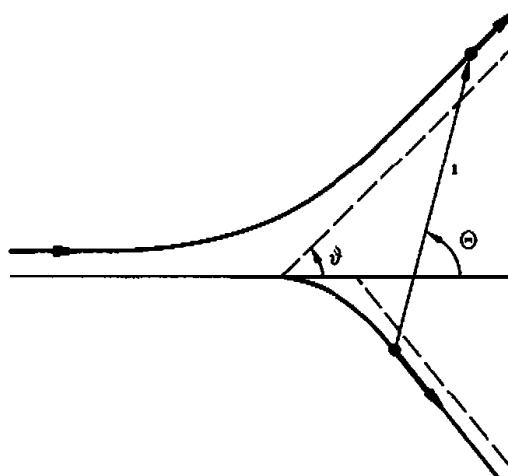


FIGURE 3.24 Scattering of two particles as viewed in the laboratory system.

central force problem can be reduced to a one-body problem, it might be thought that the only change is to replace m by the reduced mass μ . However, the matter is not quite that simple. The scattering angle actually measured in the laboratory, which we shall denote by ϑ , is the angle between the final and incident directions of the scattered particle in laboratory coordinates.[†] On the other hand, the angle Θ calculated from the equivalent one-body problem is the angle between the final and initial directions of the relative vector between the two particles in the center of mass coordinates. These two angles, θ and Θ , would be the same only if the second particle remains stationary through the scattering process. In general, however, the second particle, though initially at rest, is itself set in motion by the mutual force between the two particles, and, as is indicated in Fig. 3.24, the two angles then have different values. The equivalent one-body problem thus does not directly furnish the scattering angle as measured in the laboratory coordinate system.

The relationship between the scattering angles Θ and ϑ can be determined by examining how the scattering takes place in a coordinate system moving with the center of mass of both particles. In such a system the total linear momentum is zero, of course, and the two particles always move with equal and opposite momenta. Figure 3.25 illustrates the appearance of the scattering process to an observer in the center of mass system. Before the scattering, the particles are moving directly toward each other; after, they are moving directly away from each other. The angle between the initial and final directions of the relative vector, Θ , must therefore be the same as the scattering angle of either particle in the center-of-mass system. The connection between the two scattering angles Θ and ϑ can thus be obtained by considering the transformation between the center-of-mass system and the laboratory system.

[†]The scattering angle ϑ must not be confused with the angle coordinate θ of the relative vector, \mathbf{r} , between the two particles

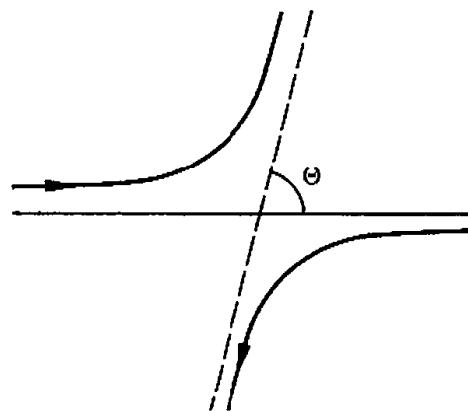


FIGURE 3.25 Scattering of two particles as viewed in the center of mass system.

It is convenient here to use the terminology of Section 3.1, with slight modifications:

- \mathbf{r}_1 and \mathbf{v}_1 are the position and velocity, after scattering, of the incident particle, m_1 , in the laboratory system,
- \mathbf{r}'_1 and \mathbf{v}'_1 are the position and velocity, after scattering, of particle m_1 in the center of mass system, and
- \mathbf{R} and \mathbf{V} are the position and (constant) velocity in the center of mass in the laboratory system.

At any instant, by definition

$$\mathbf{r}_1 = \mathbf{R} + \mathbf{r}'_1,$$

and consequently

$$\mathbf{v}_1 = \mathbf{V} + \mathbf{v}'_1. \quad (3.104)$$

Figure 3.26 graphically portrays this vector relation evaluated *after* the scattering has taken place; at which time \mathbf{v}_1 and \mathbf{v}'_1 make the angles ϑ and Θ , respectively,

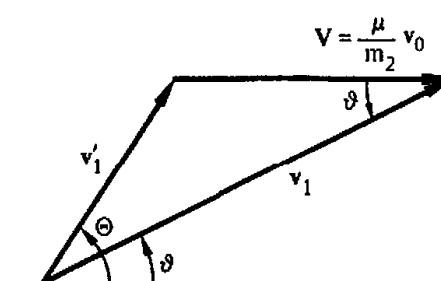


FIGURE 3.26 The relations between the velocities in the center of mass and laboratory coordinates.

with the vector \mathbf{V} lying along the initial direction. Since the target is initially stationary in the laboratory system, the incident velocity of particle 1 in that system, v_0 , is the same as the initial relative velocity of the particles. By conservation of total linear momentum, the constant velocity of the center of mass is therefore given by

$$(m_1 + m_2)\mathbf{V} = m_1\mathbf{v}_0,$$

or

$$\mathbf{V} = \frac{\mu}{m_2}\mathbf{v}_0, \quad (3.105)$$

where $\mu = m_1 m_2 / (m_1 + m_2)$. From Fig. 3.26, it is readily seen that

$$v_1 \sin \vartheta = v'_1 \sin \Theta$$

and

$$v_1 \cos \vartheta = v'_1 \cos \Theta + V. \quad (3.106)$$

The ratio of these two equations gives a relation between ϑ and Θ :

$$\tan \vartheta = \frac{\sin \Theta}{\cos \Theta + \rho}, \quad (3.107)$$

where ρ is defined as

$$\rho \equiv \frac{\mu}{m_2} \frac{v_0}{v'_1}. \quad (3.108)$$

An alternative relation can be obtained by expressing v_1 in terms of the other speeds through the cosine law as applied to the triangle of Fig. 3.26:

$$v_1^2 = v'_1^2 + V^2 + 2v'_1 V \cos \Theta. \quad (3.109)$$

When this is used to eliminate v_1 from Eq. (3.106) and V is expressed in terms of v_0 by Eq. (3.105), we find

$$\cos \vartheta = \frac{\cos \Theta + \rho}{\sqrt{1 + 2\rho \cos \Theta + \rho^2}}. \quad (3.110)$$

Both these relations still involve a ratio of speeds through ρ . By the definition of center of mass, the speed of particle 1 in the center-of-mass system, v'_1 , is connected with the relative speed v after the collision, by the equation (cf. Eq. (3.2)), where $v = |\dot{r}|$:

$$v'_1 = \frac{\mu}{m_1} v.$$

Hence, ρ can also be written as

$$\rho = \frac{m_1}{m_2} \frac{v_0}{v}, \quad (3.108')$$

where v , it should be emphasized, is the relative speed *after* the collision. When the collision is *elastic*, the total kinetic energy of the two particles remains unaltered and v must equal v_0 so that ρ is simply

$$\rho = \frac{m_1}{m_2}, \quad (\text{elastic collision}) \quad (3.111)$$

independent of energies or speeds. If the collision is *inelastic*, the total kinetic energy of the two particles is altered (e.g., some of the kinetic energy goes into the form of internal excitation energy of the target). Since the total energy is conserved and momentum is conserved, the energy change resulting from the collision can be expressed as

$$\frac{\mu v^2}{2} = \frac{\mu v_0^2}{2} + Q. \quad (3.112)$$

The so-called Q value of the inelastic collision is clearly negative in magnitude, but the sign convention is chosen to conform to that used in general for atomic and nuclear reactions. From Eq. (3.112) the ratio of relative speeds before and after collision can be written

$$\frac{v}{v_0} = \sqrt{1 + \frac{m_1 + m_2}{m_2} \frac{Q}{E}}, \quad (3.113)$$

where $E = \frac{1}{2}mv_0^2$ is the energy of the incoming particle (in the laboratory system). Thus, for inelastic scattering ρ becomes

$$\rho = \frac{m_1}{m_2 \sqrt{1 + \frac{m_1 + m_2}{m_2} \frac{Q}{E}}}. \quad (\text{inelastic scattering}) \quad (3.114)$$

Not only are the scattering angles ϑ and Θ in general different in magnitude, but the values of the differential scattering cross section depend upon which of the two angles is used as the argument of σ . The connection between the two functional forms is obtained from the observation that in a particular experiment the number of particles scattered into a given element of solid angle must be the same whether we measure the event in terms of ϑ or Θ . As an equation, this statement can be written

$$2\pi I\sigma(\Theta) \sin \Theta |d\Theta| = 2\pi I\sigma'(\vartheta) \sin \vartheta |d\vartheta|,$$

or

$$\sigma'(\vartheta) = \sigma(\Theta) \frac{\sin \Theta}{\sin \vartheta} \left| \frac{d\Theta}{d\vartheta} \right| = \sigma(\Theta) \left| \frac{d(\cos \Theta)}{d(\cos \vartheta)} \right|, \quad (3.115)$$

where $\sigma'(\vartheta)$ is the differential scattering cross section expressed in terms of the scattering angle in the laboratory system. The derivative can easily be evaluated from Eq. (3.110), leading to the result

$$\sigma'(\vartheta) = \sigma(\Theta) \frac{(1 + 2\rho \cos \Theta + \rho^2)^{3/2}}{1 + \rho \cos \Theta}. \quad (3.116)$$

Note that $\sigma(\Theta)$ is *not* the cross section an observer would measure in the center-of-mass system. Both $\sigma(\Theta)$ and $\sigma'(\vartheta)$ are cross sections measured in the laboratory system; they are merely expressed in terms of different coordinates. An observer fixed in the center-of-mass system would see a different flux density of incident particles from that measured in the laboratory system, and this transformation of flux density would have to be included if (for some reason) we wanted to relate the cross sections as measured in the two different systems.

The two scattering angles have a particularly simple relation for elastic scattering when the two masses of particles are equal. It then follows that $\rho = 1$, and from Eq. (3.110) we have

$$\cos \vartheta = \sqrt{\frac{1 + \cos \Theta}{2}} = \cos \frac{\Theta}{2},$$

or

$$\vartheta = \frac{\Theta}{2}, \quad (\rho = 1).$$

Thus, with equal masses, scattering angles greater than 90° cannot occur in the laboratory system; all the scattering is in the forward hemisphere. Correspondingly, the scattering cross section is given in terms of Θ from Eq. (3.116) as

$$\sigma'(\vartheta) = 4 \cos \vartheta \cdot \sigma(\Theta), \quad \vartheta \leq \frac{\pi}{2}, \quad (\rho = 1).$$

Even when the scattering is isotropic in terms of Θ , i.e., $\sigma(\Theta)$ is constant, independent of Θ , then the cross section in terms of ϑ varies as the cosine of the angle! When, however, the scattering mass m_2 is very large compared to the incident particle mass m_1 and the scattering is elastic, then from Eq. (3.111) $\rho \approx 0$, so $\sigma'(\vartheta) \approx \sigma(\Theta)$ from Eq. (3.116).

We have seen that even in elastic collisions, where the total kinetic energy remains constant, a collision with an initially stationary target results in a transfer of kinetic energy to the target with a corresponding decrease in the kinetic energy of the incident particle. In other words, the collision *slows down* the incident

particle. The degree of slowing down can be obtained from Eq. (3.109) if v'_1 and V are expressed in terms of v_0 by Eqs. (3.108) and (3.105), respectively:

$$\frac{v_1^2}{v_0^2} = \left(\frac{\mu}{m_2 \rho} \right)^2 (1 + 2\rho \cos \Theta + \rho^2) \quad (3.117)$$

For elastic collisions $\rho = m_1/m_2$, and Eq. (3.117) can be simplified to

$$\frac{E_1}{E_0} = \frac{1 + 2\rho \cos \Theta + \rho^2}{(1 + \rho)^2}, \quad (\text{elastic collision}) \quad (3.117')$$

where E_0 is the initial kinetic energy of the incident particle in the laboratory system and E_1 the corresponding energy after scattering. When the particles are of equal mass, this relation becomes

$$\frac{E_1}{E_0} = \frac{1 + \cos \Theta}{2} = \cos \vartheta.$$

Thus, at the maximum scattering angle ($\Theta = \pi$, $\vartheta = \pi/2$), the incident particle loses all its energy and is completely stopped in the laboratory system.

This transfer of kinetic energy by scattering is, of course, the principle behind the “moderator” in a thermal neutron reactor. Fast neutrons produced by fission make successive elastic scatterings until their kinetic energy is reduced to thermal energies, where they are more liable to cause fission than to be captured. Clearly the best moderators will be the light elements, ideally hydrogen ($\rho = 1$). For a nuclear reactor, hydrogen is practical only when contained as part of a mixture or compound, such as water. Other light elements useful for their moderating properties include deuterium, of mass 2, and carbon, of mass 12. Hydrogen, as present in paraffin, water, or plastics, is frequently used in the laboratory to slow down neutrons.

Despite their current useful applications, these calculations of the transformation from laboratory to center of mass coordinates, and of the transfer of kinetic energy, are not particularly “modern” or “quantum” in nature. Nor is the classical mechanics involved particularly advanced or difficult. All that has been used, essentially, is the conservation of momentum and energy. Indeed, similar calculations may be found in freshman textbooks, usually in terms of elastic collisions between, say, billiard balls. But it is their elementary nature that results in the widespread validity of these calculations. So long as momentum is conserved (and this will be true in quantum mechanics) and the Q value is known, the details of the scattering process are irrelevant. In effect, the vicinity of the scattering particle is a “black box,” and we are concerned only with what goes in and what comes out. It matters not at all whether the phenomena occurring inside the box are “classical” or “quantum.” Consequently, the formulae of this section may be used in the experimental analysis of phenomena essentially quantum in nature, as for example, neutron-proton scattering, so long as the energies are low enough that relativistic effects may be neglected. (See Section 7.7 for a discussion of the relativistic treatment of the kinematics of collisions.)

3.12 ■ THE THREE-BODY PROBLEM

Thus far, we have treated integrable problems in which the equations of motion can be integrated to give a closed-form solution. For the two-body case of the inverse-square law, we found solutions involving motion in elliptic, parabolic, and hyperbolic orbits, the former of which constitute closed orbits. Solutions can also be found for some additional power laws of the form $V(r) = ar^n$. Nevertheless, for almost all other possible central force potentials, the equations of motion cannot be integrated. When one more mass is added, the situation becomes much more complex. Even for inverse-square law forces, this three-body Kepler-type problem has no known general solution. In the present section we shall examine some simple examples of what happens when this third mass is added.

The Newtonian three-body problem involves three masses m_1 , m_2 , and m_3 at the respective positions \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 , interacting with each other via gravitational forces. We assume that the position vectors \mathbf{r}_1 , \mathbf{r}_2 , and \mathbf{r}_3 are expressed in the center of mass system. It is easy to write the equation of motion of the first mass since by Newton's second law $m_1\ddot{\mathbf{r}}_1$ equals the gravitational forces that the other two masses exert on m_1 :

$$\ddot{\mathbf{r}}_1 = -Gm_2 \frac{\mathbf{r}_1 - \mathbf{r}_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3} - Gm_3 \frac{\mathbf{r}_1 - \mathbf{r}_3}{|\mathbf{r}_1 - \mathbf{r}_3|^3} \quad (3.118)$$

and analogously for the other two masses. If we make use of the relative-position vectors defined by

$$\mathbf{s}_i = \mathbf{r}_j - \mathbf{r}_k \quad (3.119)$$

in Fig. 3.27, then clearly

$$\mathbf{s}_1 + \mathbf{s}_2 + \mathbf{s}_3 = 0. \quad (3.120)$$

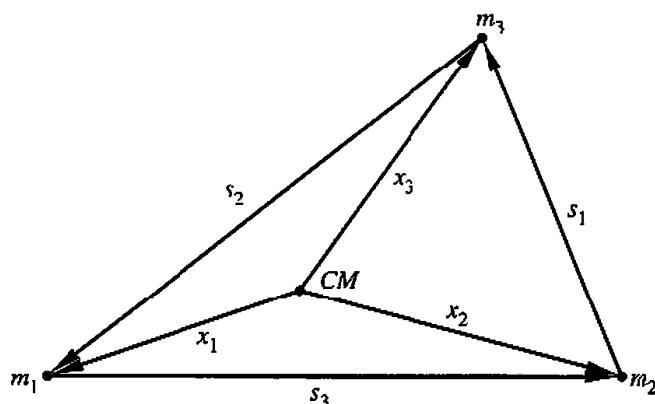


FIGURE 3.27 Position vectors $\mathbf{s}_i = \mathbf{r}_j - \mathbf{r}_k$ for the three-body problem. Adapted from Hestenes, *New Foundations for Classical Mechanics*, 1999, Fig. 5.1.

After a little algebra, the equations of motion assume the symmetrical form

$$\ddot{\mathbf{s}}_i = -mG \frac{\mathbf{s}_i}{s_i^3} + m_i \mathbf{G} \quad (3.121)$$

where $i = 1, 2, 3$, the quantity m is the sum of the three masses

$$m = m_1 + m_2 + m_3 \quad (3.122)$$

and the vector \mathbf{G} is given by

$$\mathbf{G} = G \left(\frac{\mathbf{s}_1}{s_1^3} + \frac{\mathbf{s}_2}{s_2^3} + \frac{\mathbf{s}_3}{s_3^3} \right). \quad (3.123)$$

The three coupled equations in the symmetrical form, (3.121), cannot be solved in general, but they do provide solutions to the three-body problem for some simple cases.

There is a solution due to Euler in which mass m_2 always lies on the straight line between the other two masses so that $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3, \mathbf{s}_1, \mathbf{s}_2, \mathbf{s}_3$, and \mathbf{G} are all collinear. Figure 3.28 shows Euler's negative-energy (i.e., bound-state) solution for the mass ratio $m_1 < m_2 < m_3$ in which the masses move along confocal ellipses with the same period τ . During each period, the masses pass through both a perihelion configuration, in which they lie close together along the axis of the ellipses, and an aphelion configuration, in which they lie along this same axis but far apart. The aphelion positions in the orbits are indicated in Figure 3.28.

If the vector $\mathbf{G} = 0$, the equations of motion decouple, and Eq. (3.121) reduces to the two-body form of the Kepler problem,

$$\ddot{\mathbf{s}}_i = -mG \frac{\mathbf{s}_i}{s_i^3}, \quad (3.124)$$

with each mass moving along an elliptical orbit lying in the same plane with the same focal point and the same period. This decoupling occurs when the three

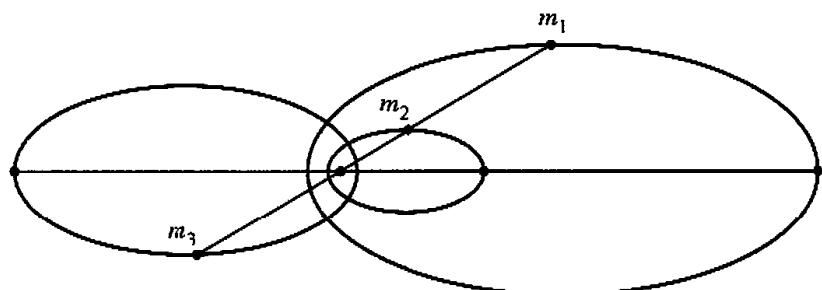


FIGURE 3.28 Euler's collinear solution to the three-body problem for the mass ratio $m_1 < m_2 < m_3$. Three of the dots show aphelion positions. Adapted from Hestenes, *New Foundations for Classical Mechanics*, 1999, Fig. 5.2

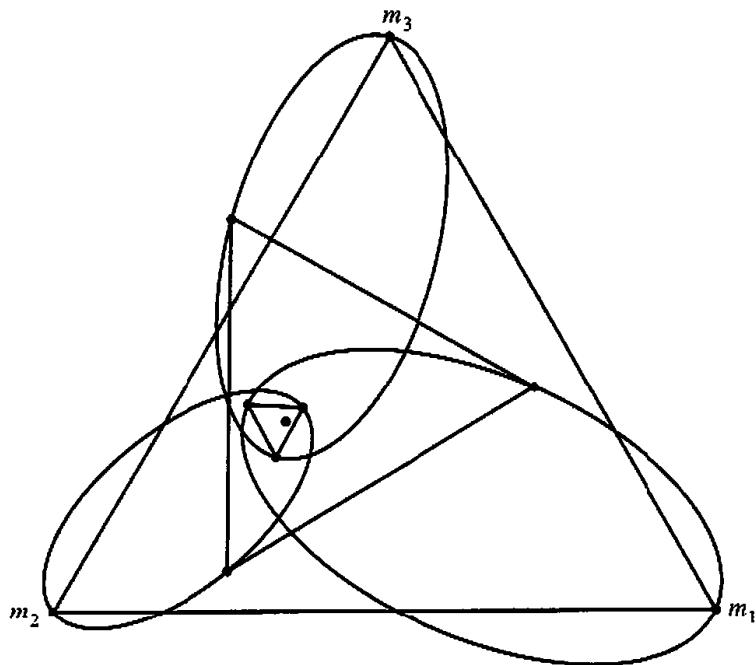


FIGURE 3.29 Lagrange's equilateral triangle solution to the three-body problem for the mass ratio $m_1 < m_2 < m_3$. Adapted from Hestenes, *New Foundations for Classical Mechanics*, 1999, Fig. 5.3.

masses are at the vertices of an equilateral triangle. As the motion proceeds, the equations remain uncoupled so the equilateral triangle condition continues to be satisfied, but the triangle changes in size and orientation. Figure 3.29 presents Lagrange's elliptic solution case with the same mass ratio as before, $m_1 < m_2 < m_3$. The figure shows the configuration when the masses are close together, each at its respective perihelion point, and also indicates the analogous aphelion arrangement.

Various asymptotic solutions have been worked out for the three-body problem. For example, if the total energy is positive, then all three masses can move away from each other, or one can escape, carrying away most of the energy, and leave the other two behind in elliptic orbits. If the energy is negative, one can escape and leave the other two in a bound state, or all three can move in bound orbits.

The restricted three-body problem is one in which two of the masses are large and bound, and the third is small and merely perturbs the motion of the other two. Examples are a spacecraft in orbit between Earth and the Moon, or the perturbation of the Sun on the Moon's orbit. In the spacecraft case, the first approach is to assume that the Earth and Moon move in their unperturbed orbits, and the satellite interacts with them through their respective inverse-square gravitational forces. We should also note that satellites orbiting Earth at altitudes of 90 miles or 150 kilometers have their orbits perturbed by Earth's nonspherical mass distribution.

A complicating factor in the restricted three-body problem is the distribution of gravitational potential energy in the vicinity of the Earth–Moon system. Close to Earth, we experience a gravitational force directed toward Earth, and close to the Moon, the force is directed toward the Moon. This means that the equipotentials, or curves of constant gravitational energy, are closed curves that encircle the Earth, (m_1) and Moon, (m_2), respectively, as shown in Fig. 3.30. In contrast to this, far from the Earth and Moon, the equipotentials encircle the Earth–Moon pair, as shown in the figure. At some point, called Lagrange point L_2 , along the horizontal line in the figure between the Earth and Moon, the attraction to the two bodies is equal in magnitude and opposite in direction so the force experienced by a small mass placed there is zero. In other words, L_2 is a local potential minimum along this line. More precisely, this point is a saddle point because the potential energy is a minimum only along the Earth–Moon axis, and decreases in directions perpendicular to this axis. Two other Lagrange points, L_1 and L_3 , along this same axis between the Earth and Moon are located at the transition points between orbits that encircle the Earth and the Moon individually, and orbits that encircle the

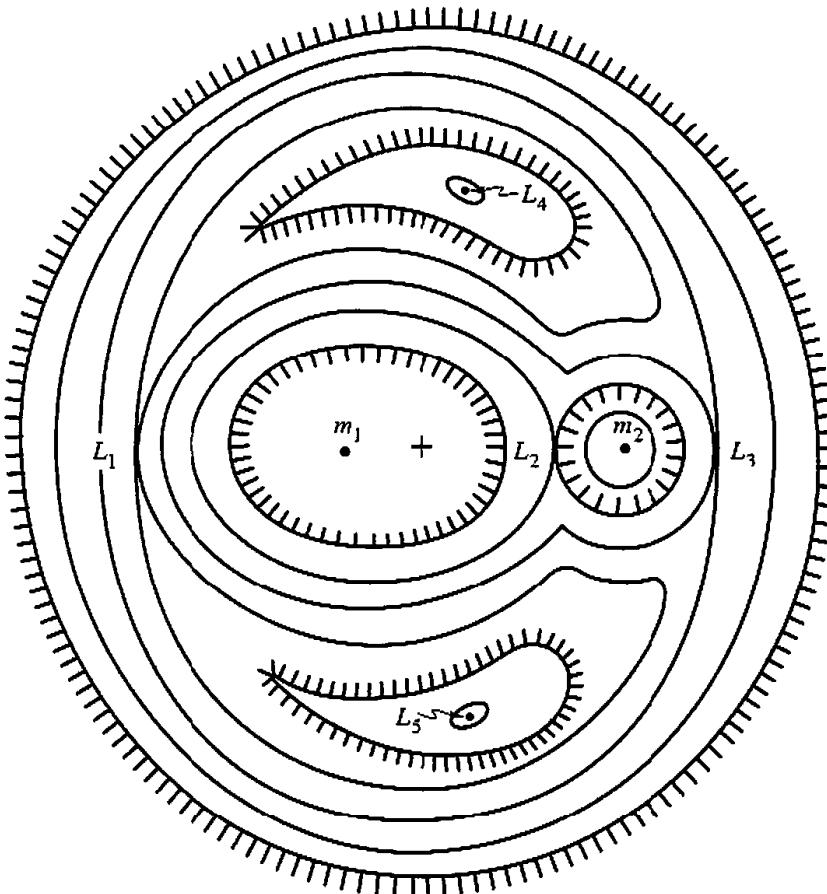


FIGURE 3.30 Contour map of equipotential curves of two masses $m_1 > m_2$ plotted in a reference system rotating with the two masses around each other. From Hestenes, *New Foundations for Classical Mechanics*, 1986, Fig. 5.5.

two together as a pair. These are also saddle points. The fourth and fifth Lagrange points, L_4 and L_5 , which are not collinear with the other three, correspond to local minima in the gravitational potential energy. Masses in the vicinity of these two points experience a force of attraction toward them, and can find themselves in stable elliptical-shaped orbits around them.

We can verify the preceding statements by considering the solutions found in Sections 3.7 and 3.8 for two massive bodies in the center-of-mass frame and asking if there are locations where a small test body will remain at rest relative to the two bodies. By a test body we mean one whose mass is sufficiently small that we can neglect its effect on the motions of the other two bodies. For simplicity, we will limit our attention to the restricted case where the bodies undergo circular motion about the center of mass. The Lagrangian for the motion of the test mass, m , can be written, in general, as

$$L = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - V(r, \theta, t), \quad (3.125)$$

where $V(r, \theta, t)$ is the time-dependent potential due to the two massive bodies.

As a consequence of the circular motion, the radius vector, \mathbf{r} , between the two bodies is of constant length and rotates with a constant frequency, ω , in the inertial frame. If we go to a coordinate system rotating at the frequency, the two massive bodies appear to be at rest and we can write the Lagrangian in terms of the rotating system by using $\theta' = \theta + \omega t$ as the transformation to the rotating frame. Thus, the Lagrangian in the rotating coordinates can be written in terms of the cylindrical coordinates, ρ , $\theta = \theta' - \omega t$, and z , with ρ being the distance from the center of mass and θ the counterclockwise angle from the line joining the two masses shown in Fig. 3.30. So

$$L = \frac{1}{2}m(\dot{\rho}^2 + \rho^2(\dot{\theta}' - \omega)^2 + \dot{z}^2) - V'(\rho, \theta, z), \quad (3.126)$$

or

$$L = \frac{1}{2}m(\dot{\rho}^2 + \rho^2\dot{\theta}'^2 + \dot{z}^2) - \left(m\omega\rho^2\dot{\theta}' - \frac{1}{2}m\rho^2\omega^2 + V'(\rho, \theta, z)\right). \quad (3.127)$$

The fifth and sixth terms are the potentials for the Coriolis effect (cf. Section 4.10) and the centrifugal effect, respectively.

The procedure then is to find the Lagrange equations and look for solutions with the conditions that $\dot{\rho} = \dot{z} = \dot{\theta}' = 0$. The solutions are the five Lagrange points shown in Fig. 3.30. Stability can be determined by investigating the effects of small displacements from these positions using the methods discussed in Chapters 6 and 12. Only L_4 and L_5 are stable.

Even though the L_2 point is not stable against displacements along the line between the masses, it has been useful for studies of the Sun. The L_2 between the Earth and Sun is the approximate location in the 1990s for the solar and heliospheric observatory, SOHO, which orbits the L_2 point in a plane perpendicular to

the Earth–Sun line. The satellite cannot be exactly at the L_2 point, or we could not receive its transmissions against the bright Sun. Small steering rockets correct for the slow drift toward, or away from, L_2 .

DERIVATIONS

1. Consider a system in which the total forces acting on the particles consist of conservative forces \mathbf{F}'_i and frictional forces \mathbf{f}_i , proportional to the velocity. Show that for such a system the virial theorem holds in the form

$$\overline{T} = -\frac{1}{2} \sum_i \mathbf{F}'_i \cdot \mathbf{r}_i,$$

providing the motion reaches a steady state and is not allowed to die down as a result of the frictional forces.

2. By expanding $e \sin \psi$ in a Fourier series in ωt , show that Kepler's equation has the formal solution

$$\psi = \omega t + \sum_{n=1}^{\infty} \frac{2}{n} J_n(ne) \sin \omega t,$$

where J_n is the Bessel function of order n . For small argument, the Bessel function can be approximated in a power series of the argument. Accordingly, from this result derive the first few terms in the expansion of ψ in powers of e .

3. If the difference $\psi - \omega t$ is represented by ρ , Kepler's equation can be written

$$\rho = e \sin(\omega t + \rho).$$

Successive approximations to ρ can be obtained by expanding $\sin \rho$ in a Taylor series in ρ , and then replacing ρ by its expression given by Kepler's equation. Show that the first approximation to ρ is ρ_1 , given by

$$\tan \rho_1 = \frac{e \sin \omega t}{1 - e \cos \omega t},$$

and that the next approximation is found from

$$\sin(\rho_2 - \rho_1) = -e^3 \sin(\omega t + \rho_1)(1 + e \cos \omega t),$$

an expression that is accurate through terms of order e^4 .

4. Show that for repulsive scattering, Eq. (3.96) for the angle of scattering as a function of the impact parameter, s , can be rewritten as

$$\Theta = \pi - 4s \int_0^1 \frac{\rho d\rho}{\sqrt{r_m^2 \left(1 - \frac{V}{E}\right) - s^2(1 - \rho^2)}}.$$

or

$$\Theta = \pi - 4s \int_0^1 \frac{d\rho}{\sqrt{\frac{r_m^2}{\rho^2 E} (V(r_m) - V(r)) + s^2(2 - \rho^2)}},$$

by changing the variable of integration to some function $\rho(r)$. Show that for a repulsive potential the integrand is never singular in the limit $r \rightarrow r_m$. Because of the definite limits of integration, these formulations have advantages for numerical calculations of $\Theta(s)$ and allow naturally for the use of Gauss-Legendre quadrature schemes.

5. Apply the formulation of the preceding exercise to compute numerically $\Theta(s)$ and the differential cross section of $\sigma(\Theta)$ for the repulsive potential

$$V = \frac{V_0}{1 +},$$

and for a total energy $E = 1.2V_0$. It is suggested that 16-point Gauss-Legendre quadrature will give adequate accuracy. Does the scattering exhibit a rainbow?

6. If a repulsive potential drops off monotonically with r , then for energies high compared to $V(r_m)$ the angle of scattering will be small. Under these conditions show that Eq. (3.97) can be manipulated so that the deflection angle is given approximately by

$$\Theta = \frac{1}{E} \int_0^1 \frac{(V(u_m) - V(u)) dy}{(1 - y^2)^{3/2}},$$

where y , obviously, is u/u_m .

Show further, that if $V(u)$ is of the form Cu^n , where n is a positive integer, then in the high-energy limit the cross section is proportional to $\Theta^{-2(1+1/n)}$.

7. (a) Show that the angle of recoil of the target particle relative to the incident direction of the scattered particle is simply $\Phi = \frac{1}{2}(\pi - \Theta)$.
 (b) It is observed that in elastic scattering the scattering cross section is isotropic in terms of Θ . What are the corresponding probability distributions for the scattered energy of the incident particle, E_1 , and for the recoil energy of the target particle, E_2 ?
 8. Show that the angle of scattering in the laboratory system, ϑ , is related to the energy before scattering, E_0 , and the energy after scattering E_1 , according to the equation

$$\cos \vartheta = \frac{m_2 + m_1}{2m_1} \sqrt{\frac{E_1}{E_0}} - \frac{m_2 - m_1}{2m_1} \sqrt{\frac{E_0}{E_1}} + \frac{m_2 Q}{2m_1 \sqrt{E_0 E_1}}.$$

9. Show that the central force problem is soluble in terms of elliptic functions when the force is a power-law function of the distance with the following fractional exponents:

$$n = -\frac{3}{2}, -\frac{5}{2}, -\frac{1}{3}, -\frac{5}{3}, -\frac{7}{3}.$$

EXERCISES

10. A planet of mass M is in an orbit of eccentricity $e = 1 - \alpha$ where $\alpha \ll 1$, about the Sun. Assume the motion of the Sun can be neglected and that only gravitational forces act. When the planet is at its greatest distance from the Sun, it is struck by a comet of mass m , where $m \ll M$ traveling in a tangential direction. Assuming the collision is completely inelastic, find the minimum kinetic energy the comet must have to change the new orbit to a parabola.
11. Two particles move about each other in circular orbits under the influence of gravitational forces, with a period τ . Their motion is suddenly stopped at a given instant of time, and they are then released and allowed to fall into each other. Prove that they collide after a time $\tau/4\sqrt{2}$.
12. Suppose that there are long-range interactions between atoms in a gas in the form of central forces derivable from a potential

$$U(r) = \frac{k}{r^m}.$$

where r is the distance between any pair of atoms and m is a positive integer. Assume further that relative to any given atom the other atoms are distributed in space such that their volume density is given by the Boltzmann factor:

$$\rho(r) = \frac{N}{V} e^{-U(r)/kT},$$

where N is the total number of atoms in a volume V . Find the addition to the virial of Clausius resulting from these forces between pairs of atoms, and compute the resulting correction to Boyle's law. Take N so large that sums may be replaced by integrals. While closed results can be found for any positive m , if desired, the mathematics can be simplified by taking $m = +1$

13. (a) Show that if a particle describes a circular orbit under the influence of an attractive central force directed toward a point on the circle, then the force varies as the inverse-fifth power of the distance.
 (b) Show that for the orbit described the total energy of the particle is zero.
 (c) Find the period of the motion.
 (d) Find \dot{x} , \dot{y} , and v as a function of angle around the circle and show that all three quantities are infinite as the particle goes through the center of force.
14. (a) For circular and parabolic orbits in an attractive $1/r$ potential having the same angular momentum, show that the perihelion distance of the parabola is one-half the radius of the circle.
 (b) Prove that in the same central force as in part (a) the speed of a particle at any point in a parabolic orbit is $\sqrt{2}$ times the speed in a circular orbit passing through the same point.
15. A meteor is observed to strike Earth with a speed v , making an angle ϕ with the zenith. Suppose that far from Earth the meteor's speed was v' and it was proceeding in a direction making a zenith angle ϕ' , the effect of Earth's gravity being to pull it into

a hyperbolic orbit intersecting Earth's surface. Show how v' and ϕ' can be determined from v and ϕ in terms of known constants.

16. Prove that in a Kepler elliptic orbit with small eccentricity e the angular motion of a particle as viewed from the *empty* focus of the ellipse is uniform (the empty focus is the focus that is *not* the center of attraction) to first order in e . It is this theorem that enables the Ptolemaic picture of planetary motion to be a reasonably accurate approximation. On this picture the Sun is assumed to move uniformly on a circle whose center is shifted from Earth by a distance called the *equant*. If the equant is taken as the distance between the two foci of the correct elliptical orbit, then the angular motion is thus described by the Ptolemaic picture accurately to first order in e .
17. One classic theme in science fiction is a twin planet ("Planet X") to Earth that is identical in mass, energy, and momentum but is located on the orbit 90° out of phase with Earth so that it is hidden from the Sun. However, because of the elliptical nature of the orbit, it is not always completely hidden. Assume this twin planet is in the same Keplerian orbit as Earth in such a manner that it is in aphelion when Earth is in perihelion. Calculate to first order in the eccentricity e the maximum angular separation of the twin and the Sun as viewed from the Earth. Could such a twin be visible from Earth? Suppose the twin planet is in an elliptical orbit having the same size and shape as that of Earth, but rotated 180° from Earth's orbit, so that Earth and the twin are in perihelion at the same time. Repeat your calculation and compare the visibility in the two situations.
18. At perigee of an elliptic gravitational orbit a particle experiences an impulse S (cf. Exercise 11, Chapter 2) in the radial direction, sending the particle into another elliptic orbit. Determine the new semimajor axis, eccentricity, and orientation in terms of the old.
19. A particle moves in a force field described by

$$F(r) = -\frac{k}{r^2} \exp\left(-\frac{r}{a}\right),$$

where k and a are positive.

- (a) Write the equations of motion and reduce them to the equivalent one-dimensional problem. Use the effective potential to discuss the qualitative nature of the orbits for different values of the energy and the angular momentum.
- (b) Show that if the orbit is nearly circular, the apsides will advance approximately by $\pi\rho/a$ per revolution, where ρ is the radius of the circular orbit.
20. A uniform distribution of dust in the solar system adds to the gravitational attraction of the Sun on a planet an additional force

$$\mathbf{F} = -mC\mathbf{r},$$

where m is the mass of the planet, C is a constant proportional to the gravitational constant and the density of the dust, and \mathbf{r} is the radius vector from the Sun to the planet (both considered as points). This additional force is very small compared to the direct Sun–planet gravitational force.

Chapter 3 The Central Force Problem

- (a) Calculate the period for a circular orbit of radius r_0 of the planet in this combined field.
- (b) Calculate the period of radial oscillations for slight disturbances from this circular orbit.
- (c) Show that nearly circular orbits can be approximated by a precessing ellipse and find the precession frequency. Is the precession in the same or opposite direction to the orbital angular velocity?
21. Show that the motion of a particle in the potential field

$$V(r) = -\frac{k}{r} + \frac{h}{r^2}$$

is the same as that of the motion under the Kepler potential alone when expressed in terms of a coordinate system rotating or precessing around the center of force.

For negative total energy, show that if the additional potential term is very small compared to the Kepler potential, then the angular speed of precession of the elliptical orbit is

$$\dot{\Omega} = \frac{2\pi m h}{l^2 \tau}.$$

The perihelion of Mercury is observed to precess (after correction for known planetary perturbations) at the rate of about $40''$ of arc per century. Show that this precession could be accounted for classically if the dimensionless quantity

$$\eta = \frac{h}{ka}$$

(which is a measure of the perturbing inverse-square potential relative to the gravitational potential) were as small as 7×10^{-8} . (The eccentricity of Mercury's orbit is 0.206, and its period is 0.24 year.)

22. The additional term in the potential behaving as r^{-2} in Exercise 21 looks very much like the centrifugal barrier term in the equivalent one-dimensional potential. Why is it then that the additional force term causes a precession of the orbit, while an addition to the barrier, through a change in l , does not?
23. Evaluate approximately the ratio of mass of the Sun to that of Earth, using only the lengths of the year and of the lunar month (27.3 days), and the mean radii of Earth's orbit (1.49×10^8 km) and of the Moon's orbit (3.8×10^5 km).
24. Show that for elliptical motion in a gravitational field the radial speed can be written as

$$\dot{r} = \frac{\omega a}{r} \sqrt{a^2 e^2 - (r - a)^2}.$$

Introduce the eccentric anomaly variable ψ in place of r and show that the resulting differential equation in ψ can be integrated immediately to give Kepler's equation.

25. If the eccentricity e is small, Kepler's equation for the eccentric anomaly ψ as a function of ωt , Eq. (3.76), is easily solved on a computer by an iterative technique that treats the $e \sin \psi$ term as of lower order than ψ . Denoting ψ_n by the n th iterative

solution, the obvious iteration relation is

$$\psi_n = \omega t + e \sin \psi_{n-1}.$$

Using this iteration procedure, find the analytic form for an expansion of ψ in powers of e at least through terms in e^3 .

26. Earth's period between successive perihelion transits (the "anomalistic year") is 365.2596 mean solar days, and the eccentricity of its orbit is 0.0167504. Assuming motion in a Keplerian elliptical orbit, how far does the Earth move in angle in the orbit, starting from perihelion, in a time equal to one-quarter of the anomalistic year? Give your result in degrees to an accuracy of one second of arc or better. Any method may be used, including numerical computation with a calculator or computer.
27. In hyperbolic motion in a $1/r$ potential, the analogue of the eccentric anomaly is F defined by

$$r = a(e \cosh F - 1),$$

where $a(e - 1)$ is the distance of closest approach. Find the analogue to Kepler's equation giving t from the time of closest approach as a function of F .

28. A *magnetic monopole* is defined (if one exists) by a magnetic field singularity of the form $\mathbf{B} = b\mathbf{r}/r^3$, where b is a constant (a measure of the magnetic charge, as it were). Suppose a particle of mass m moves in the field of a magnetic monopole and a central force field derived from the potential $V(r) = -k/r$.
- (a) Find the form of Newton's equation of motion, using the Lorentz force given by Eq. (1.60). By looking at the product $\mathbf{r} \times \dot{\mathbf{p}}$ show that while the mechanical angular momentum is not conserved (the field of force is noncentral) there is a conserved vector

$$\mathbf{D} = \mathbf{L} - \frac{qb}{c} \frac{\mathbf{r}}{r}.$$

- (b) By paralleling the steps leading from Eq. (3.79) to Eq. (3.82), show that for some $f(r)$ there is a conserved vector analogous to the Laplace–Runge–Lenz vector in which \mathbf{D} plays the same role as \mathbf{L} in the pure Kepler force problem.

29. If all the momentum vectors of a particle along its trajectory are translated so as to start from the center of force, then the heads of the vectors trace out the particle's *hodograph*, a locus curve of considerable antiquity in the history of mechanics, with something of a revival in connection with space vehicle dynamics. By taking the cross product of \mathbf{L} with the Laplace–Runge–Lenz vector \mathbf{A} , show that the hodograph for elliptical Kepler motion is a circle of radius mk/l with origin on the y axis displaced a distance A/l from the center of force.
30. What changes, if any, would there be in Rutherford scattering if the Coulomb force were attractive, instead of repulsive?
31. Examine the scattering produced by a repulsive central force $f = kr^{-3}$. Show that the differential cross section is given by

$$\sigma(\Theta) d\Theta = \frac{k}{2E} \frac{(1-x) dx}{x^2(2-x)^2 \sin \pi x},$$

where x is the ratio of Θ/π and E is the energy.

Chapter 3 The Central Force Problem

32. A central force potential frequently encountered in nuclear physics is the *rectangular well*, defined by the potential

$$\begin{aligned} V &= 0 & r > a \\ &= -V_0 & r \leq a. \end{aligned}$$

Show that the scattering produced by such a potential in classical mechanics is identical with the refraction of light rays by a sphere of radius a and relative index of refraction

$$n = \sqrt{\frac{E + V_0}{E}}.$$

(This equivalence demonstrates why it was possible to explain refraction phenomena both by Huygen's waves and by Newton's mechanical corpuscles.) Show also that the differential cross section is

$$\sigma(\Theta) = \frac{n^2 a^2}{4 \cos \frac{\Theta}{2}} \frac{(n \cos \frac{\Theta}{2} - 1)(n \cos \frac{\Theta}{2})}{\left(1 + n^2 - 2n \cos \frac{\Theta}{2}\right)^2}$$

What is the total cross section?

33. A particle of mass m is constrained to move under gravity without friction on the inside of a paraboloid of revolution whose axis is vertical. Find the one-dimensional problem equivalent to its motion. What is the condition on the particle's initial velocity to produce circular motion? Find the period of small oscillations about this circular motion.
34. Consider a truncated repulsive Coulomb potential defined as

$$\begin{aligned} V &= \frac{k}{r} & r > a \\ &= \frac{k}{a} & r \leq a. \end{aligned}$$

For a particle of total energy $E > k/a$, obtain expressions for the scattering angle Θ as a function of s/s_0 , where s_0 is the impact parameter for which the periapsis occurs at the point $r = a$. (The formulas can be given in closed form but they are not simple!) Make a numerical plot of Θ versus s/s_0 for the special case $E = 2k/a$. What can you deduce about the angular scattering cross section from the dependence of Θ on s/s_0 for this particular case?

35. Another version of the truncated Coulomb potential has the form

$$\begin{aligned} V &= \frac{k}{r} - \frac{k}{a} & r > a \\ &= 0 & r < a. \end{aligned}$$

Obtain closed-form expressions for the scattering angle and the differential scattering cross section. These are most conveniently expressed in terms of a parameter measuring the distance of closest approach in units of a . What is the total cross section?

- 36.** The restricted three-body problem consists of two masses in circular orbits about each other and a third body of much smaller mass whose effect on the two larger bodies can be neglected.
- Define an effective potential $V(x, y)$ for this problem where the x axis is the line of the two larger masses. Sketch the function $V(x, 0)$ and show that there are two “valleys” (points of stable equilibrium) corresponding to the two masses. Also show that there are three “hills” (three points of unstable equilibrium).
 - Using a computer program, calculate some orbits for the restricted three-body problem. Many orbits will end with ejection of the smaller mass. Start by assuming a position and a vector velocity for the small mass.

CHAPTER

4

The Kinematics of Rigid Body Motion

A rigid body was defined previously as a system of mass points subject to the holonomic constraints that the distances between all pairs of points remain constant throughout the motion. Although something of an idealization, the concept is quite useful, and the mechanics of rigid body motion deserves a full exposition. In this chapter we shall discuss principally the *kinematics* of rigid bodies, i.e., the nature and characteristics of their motions. We devote some time to developing the mathematical techniques involved, which are of considerable interest in themselves, and have many important applications to other fields of physics.

Of essential importance is the rotational motion of a rigid body. These considerations lead directly to the relation between the time rate of change of a vector in an inertial frame and the time rate of change of the same vector in a rotating frame. Since it is appropriate at that point, we leave kinematics and develop the description of the dynamics of motion in a rotating frame. In the next chapter we discuss, using the Lagrangian formulation, how the motion of extended objects is generated by applied forces and torques.

4.1 ■ THE INDEPENDENT COORDINATES OF A RIGID BODY

Before discussing the motion of a rigid body, we must first establish how many independent coordinates are necessary to specify its configuration. From experience, we expect that there should be six independent coordinates. Three external coordinates are needed to specify the position of some reference point in the body and three more to specify how the body is oriented with respect to the external coordinates. In this section we show that these intuitive expectations are correct.

A rigid body with N particles can at most have $3N$ degrees of freedom, but these are greatly reduced by the constraints, which can be expressed as equations of the form

$$r_{ij} = c_{ij}. \quad (4.1)$$

Here r_{ij} is the distance between the i th and j th particles and the c 's are constants. The actual number of degrees of freedom cannot be obtained simply by subtracting the number of constraint equations from $3N$, for there are $\frac{1}{2}N(N - 1)$ possible equations of the form of Eq. (4.1), which is far greater than $3N$ for large N . In truth, the Eqs. (4.1) are not all independent.

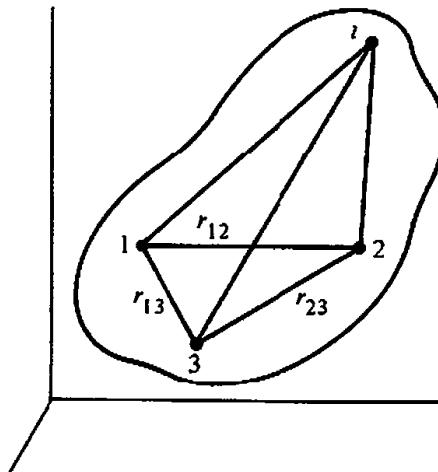


FIGURE 4.1 The location of a point in a rigid body by its distances from three reference points.

To fix a point in the rigid body, it is not necessary to specify its distances to *all* other points in the body; we need only state the distances to any three other noncollinear points (cf. Fig. 4.1). Thus, once the positions of three of the particles of the rigid body are determined, the constraints fix the positions of all remaining particles. The number of degrees of freedom therefore cannot be more than nine. But the three reference points are themselves not independent; there are in fact three equations of rigid constraint imposed on them,

$$r_{12} = c_{12}, \quad r_{23} = c_{23}, \quad r_{13} = c_{13},$$

that reduce the number of degrees of freedom to six. That only six coordinates are needed can also be seen from the following considerations. To establish the position of one of the reference points, three coordinates must be supplied. But once point 1 is fixed, point 2 can be specified by only two coordinates, since it is constrained to move on the surface of a sphere centered at point 1. With these two points determined, point 3 has only one degree of freedom, for it can only rotate about the axis joining the other two points. Hence, a total of six coordinates is sufficient.

A rigid body in space thus needs six independent generalized coordinates to specify its configuration, no matter how many particles it may contain—even in the limit of a continuous body. Of course, there may be additional constraints on the body besides the constraint of rigidity. For example, the body may be constrained to move on a surface, or with one point fixed. In such case, the additional constraints will further reduce the number of degrees of freedom, and hence the number of independent coordinates.

How shall these coordinates be assigned? Note that the set of configuration of a rigid body is completely specified by locating a Cartesian set of coordinates

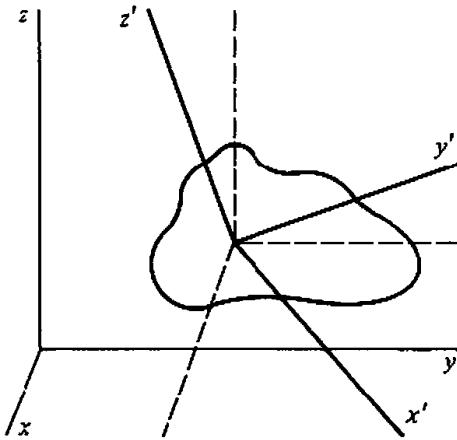


FIGURE 4.2 Unprimed axes represent an external reference set of axes: the primed axes are fixed in the rigid body.

fixed in the rigid body (the primed axes shown in Fig. 4.2) relative to the coordinate axes of the external space. Clearly three of the coordinates are needed to specify the coordinates of the origin of this “body” set of axes. The remaining three coordinates must then specify the orientation of the primed axes relative to a coordinate system parallel to the external axes, but with the same origin as the primed axes.

There are many ways of specifying the orientation of a Cartesian set of axes relative to another set with common origin. One fruitful procedure is to state the direction cosines of the primed axes relative to the unprimed. Thus, the x' axis could be specified by its three direction cosines $\alpha_1, \alpha_2, \alpha_3$, with respect to the x, y, z axes. If, as customary, $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are three unit vectors along x, y, z , and $\mathbf{i}', \mathbf{j}', \mathbf{k}'$ perform the same function in the primed system (cf. Fig. 4.3), then these direction cosines are defined as

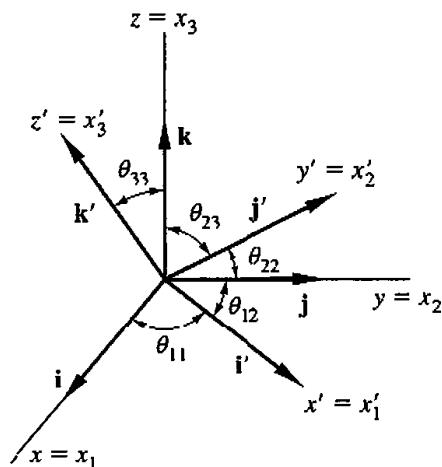


FIGURE 4.3 Direction cosines of the body set of axes relative to an external set of axes.

$$\begin{aligned}
 \cos \theta_{11} &= \cos(\mathbf{i}' \cdot \mathbf{i}) = \mathbf{i}' \cdot \mathbf{i} = \mathbf{i} \cdot \mathbf{i}' \\
 \cos \theta_{12} &= \cos(\mathbf{i}' \cdot \mathbf{j}) = \mathbf{i}' \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{i}' \\
 \cos \theta_{21} &= \cos(\mathbf{j}' \cdot \mathbf{i}) = \mathbf{j}' \cdot \mathbf{i} = \mathbf{i} \cdot \mathbf{j}' \\
 \cos \theta_{22} &= \cos(\mathbf{j}' \cdot \mathbf{j}) = \mathbf{j}' \cdot \mathbf{j} = \mathbf{j} \cdot \mathbf{j}' \tag{4.2}
 \end{aligned}$$

and similarly for $\cos \theta_{13}$, $\cos \theta_{31}$, etc. Note that the angle θ_{ij} is defined so that the first index refers to the primed system and the second index to the unprimed system. These direction cosines can also be used to express the unit vector in the primed system in terms of the unit vectors of the unprimed system giving

$$\begin{aligned}
 \mathbf{i}' &= \cos \theta_{11}\mathbf{i} + \cos \theta_{12}\mathbf{j} + \cos \theta_{13}\mathbf{k} \\
 \mathbf{j}' &= \cos \theta_{21}\mathbf{i} + \cos \theta_{22}\mathbf{j} + \cos \theta_{23}\mathbf{k} \\
 \mathbf{k}' &= \cos \theta_{31}\mathbf{i} + \cos \theta_{32}\mathbf{j} + \cos \theta_{33}\mathbf{k}. \tag{4.3}
 \end{aligned}$$

These sets of nine directions cosines then completely specify the orientation of the x' , y' , z' axes relative to the x , y , z set. We can equally well invert the process, and use the direction cosines to express the \mathbf{i} , \mathbf{j} , \mathbf{k} unit vectors in terms of their components along the primed axes. Thus, we can write

$$\mathbf{r} - xi + y\mathbf{j} + z\mathbf{k} = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}' \tag{4.4}$$

by

$$\begin{aligned}
 x' &= (\mathbf{r} \cdot \mathbf{i}') = \cos \theta_{11}x + \cos \theta_{12}y + \cos \theta_{13}z \\
 y' &= (\mathbf{r} \cdot \mathbf{j}') = \cos \theta_{21}x + \cos \theta_{22}y + \cos \theta_{23}z \\
 z' &= (\mathbf{r} \cdot \mathbf{k}') = \cos \theta_{31}x + \cos \theta_{32}y + \cos \theta_{33}z \tag{4.5}
 \end{aligned}$$

with analogous equations for \mathbf{i} , \mathbf{j} and \mathbf{k} .

The direction cosines also furnish directly the relations between the coordinates of a given point in one system and the coordinates in the other system. Thus, the coordinates of a point in a given reference frame are the components of the position vector, \mathbf{r} , along the primed and unprimed axes of the system, respectively. The primed coordinates are then given in terms of x , y , and z , as shown in Eq. (4.5). What has been done here for the components of the \mathbf{r} vector can obviously be done for any arbitrary vector. If \mathbf{G} is some vector, then the component of \mathbf{G} along the x' axis will be related to its x -, y -, z -components by

$$G_{x'} = \mathbf{G} \cdot \mathbf{i}' = \cos \theta_{11}G_x + \cos \theta_{12}G_y + \cos \theta_{13}G_z, \tag{4.6}$$

and so on. The set of nine direction cosines thus completely spells out the transformation between the two coordinate systems.

If the primed axes are taken as fixed in the body, then the nine direction cosines will be functions of time as the body changes its orientation in the course of the

motion. In this sense, the direction cosines can be considered as coordinates describing the instantaneous orientation of the body, relative to a coordinate system fixed in space but with origin in common with the body system. But, clearly, they are not independent coordinates, for there are nine of them and it has been shown that only three coordinates are needed to specify an orientation.

The connections between the direction cosines arise from the fact that the basis vectors in both coordinate systems are orthogonal to each other and have unit magnitude; in symbols,

$$\begin{aligned} \mathbf{i} \cdot \mathbf{j} &= \mathbf{j} \cdot \mathbf{k} = \mathbf{k} \cdot \mathbf{i} = 0, \\ \text{and} \\ \mathbf{i} \cdot \mathbf{i} &= \mathbf{j} \cdot \mathbf{j} = \mathbf{k} \cdot \mathbf{k} = 1, \end{aligned} \quad (4.7)$$

with similar relations for \mathbf{i}' , \mathbf{j}' , and \mathbf{k}' . We can obtain the conditions satisfied by the nine coefficients by forming all possible dot products among the three equations for \mathbf{i} , \mathbf{j} , and \mathbf{k} in terms of \mathbf{i}' , \mathbf{j}' , and \mathbf{k}' (as in Eq. (4.4)), making use of the Eqs. (4.7):

$$\begin{aligned} \sum_{l=1}^3 \cos \theta_{lm'} \cos \theta_{lm} &= 0 \quad m \neq m' \\ \sum_{l=1}^3 \cos^2 \theta_{lm} &= 1. \end{aligned} \quad (4.8)$$

These two sets of three equations each are exactly sufficient to reduce the number of independent quantities from nine to three. Formally, the six equations can be combined into one by using the Kronecker δ -symbol δ_{lm} , defined by

$$\begin{aligned} \delta_{lm} &= 1 \quad l = m \\ &= 0 \quad l \neq m. \end{aligned}$$

Equations (4.8) can then be written as

$$\sum_{l=1}^3 \cos \theta_{lm'} \cos \theta_{lm} = \delta_{m'm} \quad (4.9)$$

It is therefore not possible to set up a Lagrangian and subsequent equations of motion with the nine direction cosines as generalized coordinates. For this purpose, we must use some set of three independent functions of the direction cosines. A number of such sets of independent variables will be described later, the most important being the Euler angles. The use of direction cosines to describe the connections between two Cartesian coordinate systems nevertheless has a number of important advantages. With their aid, many of the theorems about the motion of rigid bodies can be expressed with great elegance and generality, and in a form naturally leading to the procedures used in special relativity and quantum mechanics. Such a mode of description therefore merits an extended discussion here.

4.2 ■ ORTHOGONAL TRANSFORMATIONS

To study the properties of the nine direction cosines with greater ease, it is convenient to change the notation and denote all coordinates by x , distinguishing the axes by subscripts:

$$\begin{aligned} x &\rightarrow x_1 \\ y &\rightarrow x_2 \\ z &\rightarrow x_3 \end{aligned} \tag{4.10}$$

as shown in Fig. 4.3. We also change the notation for the direction cosines to

$$a_{ij} = \cos \theta_{ij} \tag{4.11}$$

Equations (4.5) and (4.6) constitute a group of transformation equations from a set of coordinates x_1, x_2, x_3 to a new set x'_1, x'_2, x'_3 . In particular, they form an example of a *linear* or *vector* transformation, defined by transformation equations of the form

$$\begin{aligned} x'_1 &= a_{11}x_1 + a_{12}x_2 + a_{13}x_3 \\ x'_2 &= a_{21}x_1 + a_{22}x_2 + a_{23}x_3 \\ x'_3 &= a_{31}x_1 + a_{32}x_2 + a_{33}x_3, \end{aligned} \tag{4.12}$$

where the a_{11}, a_{12}, \dots , are any set of constant (independent of x, x') coefficients.* To simplify the appearance of many of the expressions, we will also make use of the summation convention first introduced by Einstein: Whenever an index occurs two or more times in a term, it is implied, without any further symbols, that the terms are to be summed over all possible values of the index. Thus, Eqs. (4.12) can be written most compactly in accordance with this convention as

$$x'_i = a_{ij}x_j, \quad i = 1, 2, 3. \tag{4.12'}$$

The repeated appearance of the index j indicates that the left-hand side of Eq. (4.12') is a sum over the dummy index j for all possible values (here, $j = 1, 2, 3$). Some ambiguity is possible where powers of an indexed quantity occur, and for that reason, an expression such as

$$\sum_i x_i^2$$

appears under the summation convention as

$$x_i x_i.$$

*Equations (4.12) of course are not the most general set of transformation equations, cf., for example, those from the r 's to the q 's (1-38).

For the rest of the book the summation convention should be automatically assumed in reading the equations unless otherwise explicitly indicated. Where convenient, or to remove ambiguity, the summation sign may be occasionally displayed explicitly, e.g., when certain values of the index are to be excluded from the summation.

The transformation represented by Eqs. (4.11) is only a special case of the general linear transformation, Eqs. (4.12), since the direction cosines are not all independent. The connections between the coefficients, Eqs. (4.8) are rederived here in terms of the newer notation. Since both coordinate systems are Cartesian, the magnitude of a vector is given in terms of the sum of squares of the components. Further, since the actual vector remains unchanged no matter which coordinate system is used, the magnitude of the vector must be the same in both systems. In symbols, we can state the invariance of the magnitude as

$$x'_i x'_i = x_i x_i. \quad (4.13)$$

The left-hand side of Eq. (4.13) is therefore

$$a_{ij} a_{ik} x_j x_k,$$

and it will reduce to the right-hand side of Eq. (4.13), if, and only if

$$\begin{aligned} a_{ij} a_{ik} &= 1 & j &= k \\ &= 0 & j &\neq k, \end{aligned} \quad (4.14)$$

or, in a more compact form, if

$$a_{ij} a_{ik} = \delta_{jk}, \quad j, k = 1, 2, 3. \quad (4.15)$$

When the a_{ij} coefficients are expressed in terms of the direction cosines, the six equations contained in Eq. (4.15) become identical with the Eqs. (4.9).

Any linear transformation, Eq. (4.12), that has the properties required by Eq. (4.15) is called an *orthogonal* transformation, and Eq. (4.15) itself is known as the *orthogonality condition*. Thus, the transition from coordinates fixed in space to coordinates fixed in the rigid body (with common origin) is accomplished by means of an orthogonal transformation. The array of transformation quantities (the direction cosines), written as

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}, \quad (4.16)$$

is called the *matrix of transformation*, and will be denoted by a capital letter **A**. The quantities a_{ij} are correspondingly known as the *matrix elements* of the transformation.

To make these formal considerations more meaningful, consider the simple example of motion in a plane, so that we are restricted to two-dimensional rotations,

and the transformation matrix reduces to the form

$$\begin{bmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The four matrix elements, a_{ij} , are connected by three orthogonality conditions:

$$a_{ij}a_{ik} = \delta_{jk}, \quad j, k = 1, 2,$$

and therefore only one independent parameter is needed to specify the transformation. But this conclusion is not surprising. A two-dimensional transformation from one Cartesian coordinate system to another corresponds to a rotation of the axes in the plane (cf. Fig. 4.4), and such a rotation can be specified completely by only one quantity, the rotation angle ϕ . Expressed in terms of this single parameter, the transformation equations become

$$\begin{aligned} x'_1 &= x_1 \cos \phi + x_2 \sin \phi \\ x'_2 &= -x_1 \sin \phi + x_2 \cos \phi \\ x'_3 &= x_3. \end{aligned}$$

The matrix elements are therefore

$$\begin{array}{lll} a_{11} = \cos \phi & a_{12} = \sin \phi & a_{13} = 0 \\ a_{21} = -\sin \phi & a_{22} = \cos \phi & a_{23} = 0 \\ a_{31} = 0 & a_{32} = 0 & a_{33} = 1, \end{array} \quad (4.17)$$

so that the matrix \mathbf{A} can be written

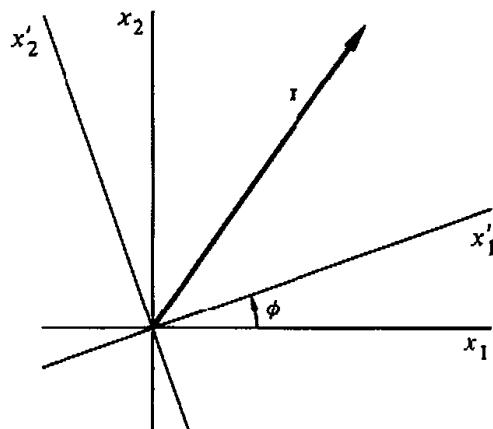


FIGURE 4.4 Rotation of the coordinate axes, as equivalent to two-dimensional orthogonal transformation.

$$\mathbf{A} = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad (4.17')$$

The three nontrivial orthogonality conditions expand into the equations

$$a_{11}a_{11} + a_{21}a_{21} = 1$$

$$a_{12}a_{12} + a_{22}a_{22} = 1$$

$$a_{11}a_{12} + a_{21}a_{22} = 0.$$

These conditions are obviously satisfied by the matrix (4.17'), for in terms of the matrix elements (4.17) they reduce to the identities

$$\cos^2 \phi + \sin^2 \phi = 1$$

$$\sin^2 \phi + \cos^2 \phi = 1$$

$$\cos \phi \sin \phi - \sin \phi \cos \phi = 0.$$

The transformation matrix \mathbf{A} can be thought of as an *operator* that, acting on the unprimed system, transforms it into the primed system. Symbolically, the process might be written

$$(\mathbf{r}') = \mathbf{A}\mathbf{r}, \quad (4.18)$$

which is to be read: The matrix \mathbf{A} operating on the components of a vector in the unprimed system yields the components of the vector in the primed system. Note that in the development of the subject so far, \mathbf{A} acts on the coordinate system only, the vector is unchanged, and we ask merely for its components in two different coordinate frames. Parentheses have therefore been placed around \mathbf{r} on the left in Eq. (4.18) to make clear that the same vector is involved on both sides of the equation. Only the components have changed. In three dimensions, the transformation of coordinates, as shown earlier, is simply a rotation, and \mathbf{A} is then identical with the *rotation operator* in a plane.

Despite this, note that without changing the formal mathematics, \mathbf{A} can also be thought of as an operator acting on the *vector* \mathbf{r} , changing it to a different vector \mathbf{r}' :

$$\mathbf{r}' = \mathbf{A}\mathbf{r}, \quad (4.19)$$

with both vectors expressed in the same coordinate system. Thus, in two dimensions, instead of rotating the coordinate system counterclockwise, we can rotate the vector \mathbf{r} *clockwise* by an angle ϕ to a new vector \mathbf{r}' , as shown in Fig. 4.5. The components of the new vector will then be related to the components of the old by the same Eqs. (4.12) that describe the transformation of coordinates. From a formal standpoint, it is therefore not necessary to use parentheses in Eq. (4.18); rather, it can be written as in Eq. (4.19) and interpreted equally as an operation on the coordinate system or on the vector. The algebra remains the same no matter

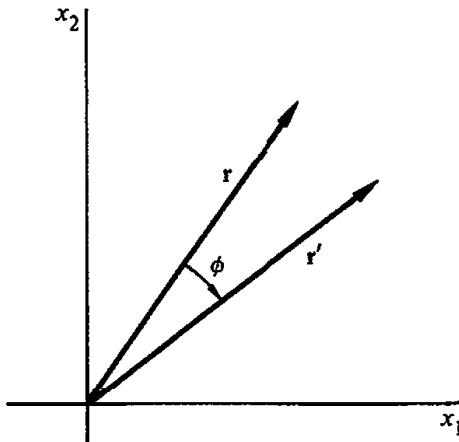


FIGURE 4.5 Interpretation of an orthogonal transformation as a rotation of the vector, leaving the coordinate system unchanged.

which of these two points of view is followed. The interpretation as an operator acting on the coordinates is the more pertinent one when using the orthogonal transformation to specify the orientation of a rigid body. On the other hand, the notion of an operator changing one vector into another has the more widespread application. In the mathematical discussion either interpretation will be freely used, as suits the convenience of the situation. Of course, note that the nature of the operation represented by A will change according to which interpretation is selected. Thus, if A corresponds to a *counterclockwise* rotation by an angle ϕ when applied to the coordinate system, it will correspond to a *clockwise* rotation when applied to the vector.

The same duality of roles often occurs with other types of coordinate transformations that are more general than orthogonal transformations. They may at times be looked on as affecting only the coordinate system, expressing some given quantity or function in terms of a new coordinate system. At other times, they may be considered as operating on the quantity or functions themselves, changing them to new quantities in the same coordinate system. When the transformation is taken as acting only on the coordinate system, we speak of the *passive* role of the transformation. In the *active* sense, the transformation is looked on as changing the vector or other physical quantity. These alternative interpretations of a transformation will be encountered in various formulations of classical mechanics to be considered below (cf. Chapter 9) and indeed occur in many fields of physics.

To develop further the kinematics of rigid body motion about a fixed origin, we shall make much use of the algebra governing the manipulation of the transformation matrix. The following section is therefore a brief summary of the elementary aspects of matrix algebra with specific application to orthogonal matrices. For those unacquainted with this branch of mathematics, the section should provide an introduction adequate for the immediate purpose. The material also details the particular terminology and notation we will employ. Those already thoroughly fa-

miliar with matrix algebra may however omit the section and proceed directly to Section 4.4.

4.3 ■ FORMAL PROPERTIES OF THE TRANSFORMATION MATRIX

Let us consider what happens when two successive transformations are made—corresponding to two successive displacements of the rigid body. Let the first transformation from \mathbf{r} to \mathbf{r}' be denoted by \mathbf{B} :

$$x'_k = b_{kj} x_j, \quad (4.20)$$

and the succeeding transformation from \mathbf{r}' to a third coordinate set \mathbf{r}'' by \mathbf{A} :

$$x''_i = a_{ik} x'_k. \quad (4.21)$$

The relation between x''_i and x_j can then be obtained by combining the two Eqs. (4.20) and (4.21):

$$x''_i = a_{ik} b_{kj} x_j.$$

This may also be written as

$$x''_i = c_{ij} x_j, \quad (4.22)$$

where

$$c_{ij} = a_{ik} b_{kj}. \quad (4.23)$$

The successive application of two orthogonal transformations \mathbf{A} , \mathbf{B} is thus equivalent to a third linear transformation \mathbf{C} . It can be shown that \mathbf{C} is also an orthogonal transformation in consequence of the orthogonality of \mathbf{A} and \mathbf{B} . The detailed proof will be left for the exercises. Symbolically, the resultant operator \mathbf{C} can be considered as the product of the two operators \mathbf{A} and \mathbf{B} :

$$\mathbf{C} = \mathbf{AB},$$

and the matrix elements c_{ij} are by definition the elements of the square matrix obtained by multiplying the two square matrices \mathbf{A} and \mathbf{B} .

Note that this “matrix” or operator multiplication is not commutative,

$$\mathbf{BA} \neq \mathbf{AB},$$

for, by definition, the elements of the transformation $\mathbf{D} = \mathbf{BA}$ are

$$d_{ij} = b_{ik} a_{kj}, \quad (4.24)$$

which generally do not agree with the matrix elements of \mathbf{C} , Eq. (4.23). Thus, the final coordinate system depends upon the order of application of the operators \mathbf{A} and \mathbf{B} , i.e., whether first \mathbf{A} then \mathbf{B} , or first \mathbf{B} and then \mathbf{A} . However, matrix multiplication is associative; in a product of three or more matrices the order of the *multiplications* is unimportant:

$$(\mathbf{AB})\mathbf{C} = \mathbf{A}(\mathbf{BC}). \quad (4.25)$$

In Eq. (4.19) the juxtaposition of \mathbf{A} and \mathbf{r} , to indicate the operation of \mathbf{A} on the coordinate system (or on the vector), was said to be merely symbolic. But, by extending our concept of matrices, it may also be taken as indicating an actual matrix multiplication. Thus far, the matrices used have been square, i.e., with equal number of rows and columns. However, we may also have one-column matrices, such as \mathbf{x} and \mathbf{x}' defined by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, \quad \mathbf{x}' = \begin{bmatrix} x'_1 \\ x'_2 \\ x'_3 \end{bmatrix}. \quad (4.26)$$

The product \mathbf{Ax} , by definition, shall be taken as a one-column matrix, with the elements

$$(\mathbf{Ax})_i = a_{ij}x_j = x'_i.$$

Hence, Eq. (4.19) can also be written as the matrix equation

$$\mathbf{x}' = \mathbf{Ax}.$$

The *addition* of two matrices, while not as important a concept as multiplication, is a frequently used operation. The sum $\mathbf{A} + \mathbf{B}$ is a matrix \mathbf{C} whose elements are the sum of the corresponding elements of \mathbf{A} and \mathbf{B} :

$$c_{ij} = a_{ij} + b_{ij}.$$

Of greater importance is the transformation inverse to \mathbf{A} , the operation that changes r' back to r . This transformation will be called \mathbf{A}^{-1} and its matrix elements designated by a'_{ij} . We then have the set of equations

$$x_i = a'_{ij}x'_j, \quad (4.27)$$

which must be consistent with

$$x'_k = a_{ki}x_i. \quad (4.28)$$

Substituting x_i from (4.27), Eq. (4.28) becomes

$$x'_k = a_{ki}a'_{ij}x'_j. \quad (4.29)$$

Since the components of \mathbf{r}' are independent, Eq. (4.29) is correct only if the summation reduces identically to x'_k . The coefficient of x'_j must therefore be 1 for $j = k$ and 0 for $j \neq k$; in symbols,

$$a_{kl} a'_{ij} = \delta_{kj}. \quad (4.30)$$

The left-hand side of Eq. (4.30) is easily recognized as the matrix element for the product $\mathbf{A}\mathbf{A}^{-1}$, while the right-hand side is the element of the matrix known as the unit matrix $\mathbf{1}$:

$$\mathbf{1} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.31)$$

Equation (4.30) can therefore be written as

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{1}, \quad (4.32)$$

which indicates the reason for the designation of the inverse matrix by \mathbf{A}^{-1} . The transformation corresponding to $\mathbf{1}$ is known as the *identity transformation*, producing no change in the coordinate system:

$$\mathbf{x} = \mathbf{1x}.$$

Similarly multiplying any matrix \mathbf{A} by $\mathbf{1}$, in any order, leaves \mathbf{A} unaffected:

$$\mathbf{1A} = \mathbf{A1} = \mathbf{A}.$$

By slightly changing the order of the proof of Eq. (4.28), it can be shown that \mathbf{A} and \mathbf{A}^{-1} commute. Instead of substituting x_i in Eq. (4.29) in terms of x' , we could equally as well demand consistency by eliminating x' from the two equations, leading in analogous fashion to

$$a'_{ij} a_{jk} = \delta_{ik}.$$

In matrix notation, this reads

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{1}, \quad (4.33)$$

which proves the statement.

Now let us consider the double sum

$$a_{kl} a_{ki} a'_{ij},$$

which can be written either as

$$c_{li} a'_{ij} \quad \text{with } c_{li} = a_{kl} a_{ki}$$

or as

$$a_{kl}d_{kj} \quad \text{with } d_{kj} = a_{ki}a'_{ij}.$$

Applying the orthogonality conditions, Eq. (4.15), the sum in the first form reduces to

$$\delta_{li}a'_{ij} = a'_{lj}.$$

On the other hand, the same sum from the second point of view, and with the help of Eq. (4.30), can be written

$$a_{kl}\delta_{kj} = a_{jl}.$$

Thus, the elements of the direct matrix \mathbf{A} and the reciprocal \mathbf{A}^{-1} are related by

$$a'_{lj} = a_{jl}. \quad (4.34)$$

In general, the matrix obtained from \mathbf{A} by interchanging rows and columns is known as the *transposed matrix*, indicated by the tilde thus $\tilde{\mathbf{A}}$. Equation (4.34) therefore states that for *orthogonal matrices* the reciprocal matrix is to be identified as the transposed matrix; symbolically,

$$\mathbf{A}^{-1} = \tilde{\mathbf{A}}. \quad (4.35)$$

If this result is substituted in Eq. (4.33), we obtain

$$\tilde{\mathbf{A}}\mathbf{A} = \mathbf{1}, \quad (4.36)$$

which is identical with the set of orthogonality conditions, Eq. (4.15), written in abbreviated form, as can be verified by direct expansion. Similarly, an alternative form of the orthogonality conditions can be obtained from Eq. (4.30) by substituting (4.34):

$$a_{ki}a_{ji} = \delta_{kj}. \quad (4.37)$$

In symbolic form, (4.37) can be written

$$\mathbf{A}\tilde{\mathbf{A}} = \mathbf{1}$$

and may be derived directly from (4.36) by multiplying it from the left by \mathbf{A} and from the right by \mathbf{A}^{-1} .

A rectangular matrix is said to be of dimension $m \times n$ if it has m rows and n columns; i.e., if the matrix element is a_{ij} , then i runs from 1 to m , and j from 1 to n . Clearly the transpose of such a matrix has the dimension $n \times m$. If a vector column matrix is considered as a rectangular matrix of dimension $m \times 1$, the transpose of a vector is of dimension $1 \times m$, i.e., a one-row matrix. The product

AB of two rectangular matrices exists only if the number of columns of **A** is the same as the number of rows of **B**. This is an obvious consequence of the definition of the multiplication operation leading to a matrix element

$$c_{ij} = a_{ik} b_{kj}.$$

From this viewpoint, the product of a vector column matrix with a square matrix does not exist. The only product between these quantities that can be formed is that of a square matrix with a single column matrix. But note that a single row matrix, i.e., a vector transpose, can indeed pre-multiply a square matrix. For a vector, however, the distinction between the column matrix and its transpose is often of no consequence. The symbol **x** may therefore be used to denote either a column or a row matrix, as the situation warrants.* Thus in the expression **Ax**, where **A** is a square matrix, the symbol **x** stands for a column matrix, whereas in the expression **xA** it represents the same elements arranged in a single row. Note that the *i*th component of **Ax** can be written as

$$A_{ij} x_j = x_j (\tilde{A})_{ji}.$$

Hence, we have a useful commutation property of the product of a vector and a square matrix that

$$\mathbf{Ax} = \mathbf{x}\tilde{\mathbf{A}}.$$

A square matrix that is the same as its transpose,

$$A_{ij} = A_{ji}, \quad (4.38)$$

is said (for obvious reasons) to be *symmetric*. When the transpose is the negative of the original matrix,

$$A_{ij} = -A_{ji}, \quad (4.39)$$

the matrix is *antisymmetric* or *skew symmetric*. Clearly in an antisymmetric matrix, the diagonal elements are always zero.

The two interpretations of an operator as transforming the vector, or alternatively the coordinate system, are both involved if we find the transformation of an operator under a change of coordinates. Let **A** be considered an operator acting upon a vector **F** (or a single-column matrix **F**) to produce a vector **G**:

$$\mathbf{G} = \mathbf{AF}.$$

If the coordinate system is transformed by a matrix **B**, the components of the vector **G** in the new system will be given by

$$\mathbf{BG} = \mathbf{BAF},$$

*The transpose sign on vector matrices will occasionally be retained where it is useful to emphasize the distinction between column and row matrices.

which can also be written

$$\mathbf{BG} = \mathbf{BAB}^{-1}\mathbf{BF}. \quad (4.40)$$

Equation (4.40) can be stated as the operator \mathbf{BAB}^{-1} acting upon the vector \mathbf{F} , expressed in the new system, produces the vector \mathbf{G} , likewise expressed in the new coordinates. We may therefore consider \mathbf{BAB}^{-1} to be the form taken by the operator \mathbf{A} when transformed to a new set of axes:

$$\mathbf{A}' = \mathbf{BAB}^{-1}. \quad (4.41)$$

Any transformation of a matrix having the form of Eq. (4.41) is known as a *similarity transformation*.

It is appropriate at this point to consider the properties of the determinant formed from the elements of a square matrix. As is customary, we shall denote such a determinant by vertical bars, thus: $|\mathbf{A}|$. Note that the definition of matrix multiplication is identical with that for the multiplication of determinants

$$|\mathbf{AB}| = |\mathbf{A}| \cdot |\mathbf{B}|. \quad (4.41')$$

Since the determinant of the unit matrix is 1, the determinantal form of the orthogonality conditions, Eq. (4.36), can be written

$$|\tilde{\mathbf{A}}| \cdot |\mathbf{A}| = 1.$$

Further, as the value of a determinant is unaffected by interchanging rows and columns, we can write

$$|\mathbf{A}|^2 = 1, \quad (4.42)$$

which implies that the determinant of an orthogonal matrix can only be +1 or -1. (The geometrical significance of these two values will be considered in the next section.)

When the matrix is not orthogonal, the determinant does not have these simple values, of course. It can be shown however that the value of the determinant is invariant under a similarity transformation. Multiplying Eq. (4.41) for the transformed matrix from the right by \mathbf{B} , we obtain the relation

$$\mathbf{A}'\mathbf{B} = \mathbf{BA},$$

or in determinantal form

$$|\mathbf{A}'| \cdot |\mathbf{B}| = |\mathbf{B}| \cdot |\mathbf{A}|.$$

Since the determinant of \mathbf{B} is merely a number, and not zero,* we can divide by

*If it were zero, there could be no inverse operator \mathbf{B}^{-1} (by Cramer's rule), which is required for Eq. (4.41) to make sense.

$|B|$ on both sides to obtain the desired result:

$$|A'| = |A|.$$

In discussing rigid body motion later, all these properties of matrix transformations, especially of orthogonal matrices, will be employed. In addition, other properties are needed, and they will be derived as the occasion requires.

4.4 ■ THE EULER ANGLES

We have noted (cf. p. 137) that the nine elements a_{ij} are not suitable as generalized coordinates because they are not independent quantities. The six relations that express the orthogonality conditions, Eqs. (4.9) or Eqs. (4.15), of course reduce the number of independent elements to three. But in order to characterize the motion of a rigid body, there is an additional requirement the matrix elements must satisfy, beyond those implied by orthogonality. In the previous section we pointed out that the determinant of a real orthogonal matrix could have the value +1 or -1. The following argument shows however that an orthogonal matrix whose determinant is -1 cannot represent a physical displacement of a rigid body.

Consider the simplest 3×3 matrix with the determinant -1:

$$S = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = -1.$$

The transformation S has the effect of changing the sign of each of the components or coordinate axes (cf. Fig. 4.6). Such an operation transforms a right-handed coordinate system into a left-handed one and is known as an *inversion* of the coordinate axes.

One method of performing an inversion is to rotate about a coordinate axis by 180° and then reflect in that coordinate axis direction. For the z -direction, this gives

$$\left(\begin{array}{l} \text{rotate} \\ \text{by } 180^\circ \\ \text{about } z \end{array} \right) \left(\begin{array}{l} \text{reflect} \\ \text{in the} \\ xy \text{ plane} \end{array} \right) = \text{inversion.}$$

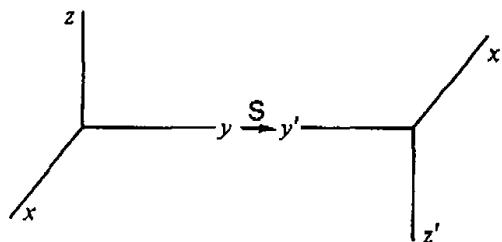


FIGURE 4.6 Inversion of the coordinate axes.

In matrix notation, this has the form

$$\begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

where the 180° rotation is obtained by setting $\phi = 180^\circ$ in Eq. (4.17).

From the nature of this operation, it is clear that an inversion of a right-handed system into a left-handed one cannot be accomplished by any *rigid* change in the orientation of the coordinate axes. An inversion therefore never corresponds to a physical displacement of a rigid body. What is true for the inversion S is equally valid for any matrix whose determinant is -1 , for any such matrix can be written as the product of S with a matrix whose determinant is $+1$, and thus includes the inversion operation. Consequently, it cannot describe a rigid change in orientation. Therefore, the transformations representing rigid body motion must be restricted to matrices having the determinant $+1$. Another method of reaching this conclusion starts from the fact that the matrix of transformation must evolve continuously from the unit matrix, which of course has the determinant $+1$. It would be incompatible with the continuity of the motion to have the matrix determinant suddenly change from its initial value $+1$ to -1 at some given time. Orthogonal transformations with determinant $+1$ are said to be *proper*, and those with the determinant -1 are called *improper*.

In order to describe the motion of rigid bodies in the Lagrangian formulation of mechanics, it will therefore be necessary to seek three independent parameters that specify the orientation of a rigid body in such a manner that the corresponding orthogonal matrix of transformation has the determinant $+1$. Only when such generalized coordinates have been found can we write a Lagrangian for the system and obtain the Lagrangian equations of motion. A number of such sets of parameters have been described in the literature, but the most common and useful are the *Euler* or *Eulerian angles*. We shall therefore define these angles at this point, and show how the elements of the orthogonal transformation matrix can be expressed in terms of them.

We can carry out the transformation from a given Cartesian coordinate system to another by means of three successive rotations performed in a specific sequence. The Euler angles are then defined as the three successive angles of rotation. Within limits, the choice of rotation angles is arbitrary. The main convention that will be followed here is used widely in celestial mechanics, applied mechanics, and frequently in molecular and solid-state physics. Other conventions will be described below and in Appendix A.

The sequence employed here is started by rotating the initial system of axes, xyz , by an angle ϕ counterclockwise about the z axis, and the resultant coordinate system is labeled the $\xi\eta\zeta$ axes. In the second stage, the intermediate axes, $\xi\eta\zeta$, are rotated about the ξ axis counterclockwise by an angle θ to produce another intermediate set, the $\xi'\eta'\zeta'$ axes. The ξ' axis is at the intersection of the xy and $\xi'\eta'$ planes and is known as the *line of nodes*. Finally, the $\xi'\eta'\zeta'$ axes are rotated coun-

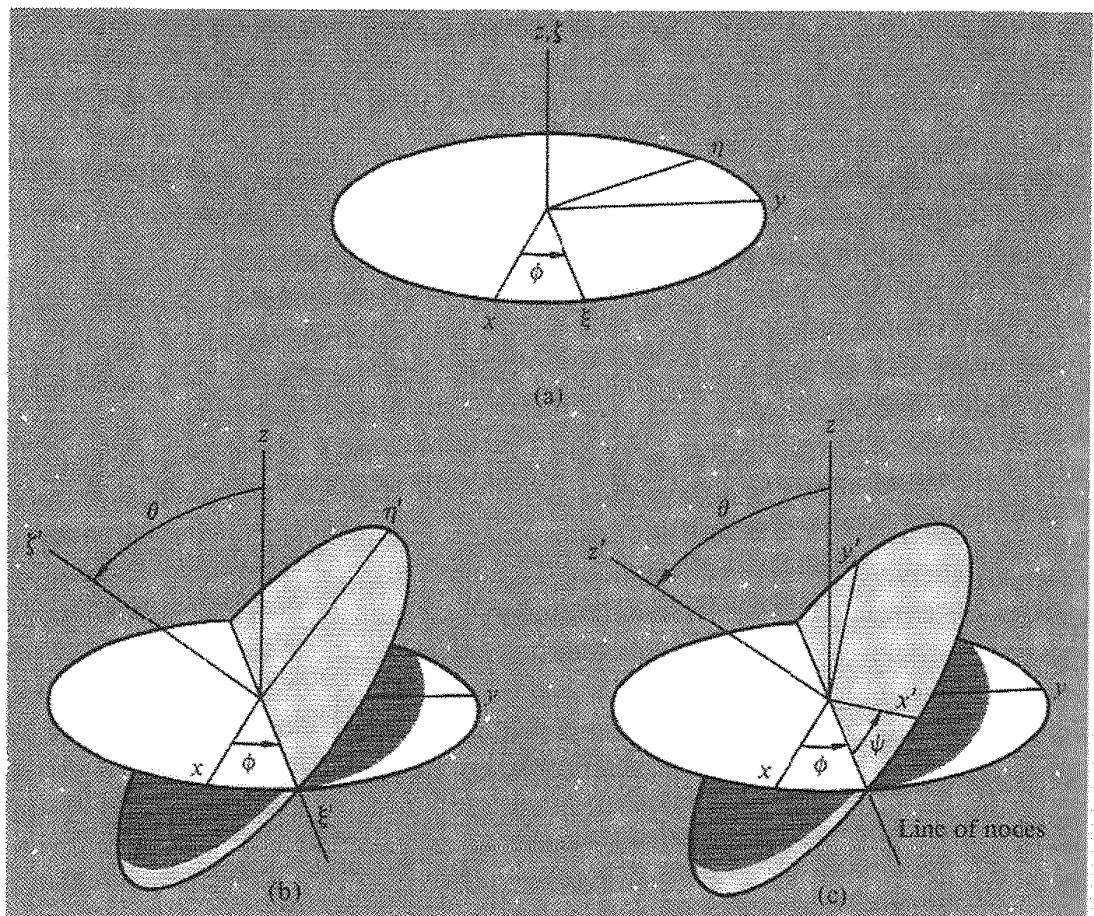


FIGURE 4.7 The rotations defining the Eulerian angles.

terclockwise by an angle ψ about the ζ' axis to produce the desired $x'y'z'$ system of axes. Figure 4.7 illustrates the various stages of the sequence. The Euler angles θ , ϕ , and ψ thus completely specify the orientation of the $x'y'z'$ system relative to the xyz and can therefore act as the three needed generalized coordinates.*

The elements of the complete transformation A can be obtained by writing the matrix as the triple product of the separate rotations, each of which has a relatively simple matrix form. Thus, the initial rotation about z can be described by a matrix D :

$$\xi = Dx,$$

where ξ and x stand for column matrices. Similarly, the transformation from $\xi\eta\zeta$ to $\xi'\eta'\zeta'$ can be described by a matrix C ,

*A number of minor variations will be found in the literature even within this convention. The differences are not very great, but they are often sufficient to frustrate easy comparison of the end formulae, such as the matrix elements. Greatest confusion, perhaps, arises from the occasional use of left-handed coordinate systems.

$$\xi' = \mathbf{C}\xi,$$

and the last rotation to $x'y'z'$ by a matrix \mathbf{B} ,

$$x' = \mathbf{B}\xi'.$$

Hence, the matrix of the complete transformation,

$$x' = \mathbf{A}x,$$

is the product of the successive matrices,

$$\mathbf{A} = \mathbf{BCD}.$$

Now the \mathbf{D} transformation is a rotation about z , and hence has a matrix of the form (cf. Eq. (4.17))

$$\mathbf{D} = \begin{bmatrix} \cos \phi & \sin \phi & 0 \\ -\sin \phi & \cos \phi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.43)$$

The \mathbf{C} transformation corresponds to a rotation about ξ , with the matrix

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \theta & \sin \theta \\ 0 & -\sin \theta & \cos \theta \end{bmatrix}, \quad (4.44)$$

and finally \mathbf{B} is a rotation about ζ' and therefore has the same form as \mathbf{D} :

$$\mathbf{B} = \begin{bmatrix} \cos \psi & \sin \psi & 0 \\ -\sin \psi & \cos \psi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (4.45)$$

The product matrix $\mathbf{A} = \mathbf{BCD}$ then follows as

$$\mathbf{A} = \begin{bmatrix} \cos \psi \cos \phi - \cos \theta \sin \phi \sin \psi & \cos \psi \sin \phi + \cos \theta \cos \phi \sin \psi & \sin \psi \sin \theta \\ -\sin \psi \cos \phi - \cos \theta \sin \phi \cos \psi & -\sin \psi \sin \phi + \cos \theta \cos \phi \cos \psi & \cos \psi \sin \theta \\ \sin \theta \sin \phi & -\sin \theta \cos \phi & \cos \theta \end{bmatrix}. \quad (4.46)$$

The inverse transformation from body coordinates to space axes

$$x = \mathbf{A}^{-1}x'$$

is then given immediately by the transposed matrix $\tilde{\mathbf{A}}$:

$$\mathbf{A}^{-1} =$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} \cos \psi \cos \phi - \cos \theta \sin \phi \sin \psi & -\sin \psi \cos \phi - \cos \theta \sin \phi \cos \psi & \sin \theta \sin \phi \\ \cos \psi \sin \phi + \cos \theta \cos \phi \sin \psi & -\sin \psi \sin \phi + \cos \theta \cos \phi \cos \psi & -\sin \theta \cos \phi \\ \sin \theta \sin \phi & \sin \theta \cos \phi & \cos \theta \end{bmatrix}. \quad (4.47)$$

Verification of the multiplication, and demonstration that \mathbf{A} represents a proper, orthogonal matrix will be left to the exercises.

Note that the sequence of rotations used to define the final orientation of the coordinate system is to some extent arbitrary. The initial rotation could be taken about any of the three Cartesian axes. In the subsequent two rotations, the only limitation is that no two successive rotations can be about the same axis. A total of 12 conventions is therefore possible in defining the Euler angles (in a right-handed coordinate system). The two most frequently used conventions differ only in the choice of axis for the second rotation. In the Euler's angle definitions described above, and used throughout the book, the second rotation is about the intermediate x axis. We will refer to this choice as the *x-convention*. In quantum mechanics, nuclear physics, and particle physics, we often take the second defining rotation about the intermediate y axis; this form will be denoted as the *y-convention*.

A third convention is commonly used in engineering applications relating to the orientation of moving vehicles such as aircraft and satellites. Both the *x*- and *y*-conventions have the drawback that when the primed coordinate system is only slightly different from the unprimed system, the angles ϕ and ψ become indistinguishable, as their respective axes of rotation, z and z' are then nearly coincident. To get around this problem, all three rotations are taken around different axes. The first rotation is about the vertical axis and gives the *heading* or *yaw* angle. The second is around a perpendicular axis fixed in the vehicle and normal to the figure axis; it is measured by the *pitch* or *attitude* angle. Finally, the third angle is one of rotation about the figure axis of the vehicle and is the *roll* or *bank* angle. Because all three axes are involved in the rotations, it will be designated as the *xyz-convention* (although the order of axes chosen may actually be different). This last convention is sometimes referred to as the *Tait-Bryan* angles.

While only the *x*-convention will be used in the text, for reference purposes Appendix A lists formulae involving Euler's angles, such as rotation matrices, in both the *y*- and *xyz*-conventions.

4.5 ■ THE CAYLEY-KLEIN PARAMETERS AND RELATED QUANTITIES

We have seen that only three independent quantities are needed to specify the orientation of a rigid body. Nonetheless, there are occasions when it is desirable to use sets of variables containing more than the minimum number of quantities to describe a rotation, even though they are not suitable as generalized coordinates. Thus, Felix Klein introduced the set of four parameters bearing his name to facilitate the integration of complicated gyroscopic problems. The Euler angles are difficult to use in numerical computation because of the large number of trigonometric functions involved, and the four-parameter representations are much better adapted for use on computers. Further, the four-parameter sets are of great theoretical interest in branches of physics beyond the scope of this book, wherever

rotations or rotational symmetry are involved. It therefore seems worthwhile to briefly describe these parameters, leaving the details to Appendix A.

The four *Cayley–Klein parameters* are complex numbers denoted by α , β , γ , and δ with the constraints that $\beta = \gamma^*$ and $\delta = \alpha^*$. In terms of these numbers, the transformation matrix of a rotated body is given by

$$\mathbf{A} = \begin{bmatrix} \frac{1}{2}(\alpha^2 - \gamma^2 + \delta^2 - \beta^2) & \frac{i}{2}(\gamma^2 - \alpha^2 + \delta^2 - \beta^2) & \gamma\delta - \alpha\beta \\ \frac{i}{2}(\alpha^2 + \gamma^2 - \beta^2 - \delta^2) & \frac{1}{2}(\alpha^2 + \gamma^2 + \beta^2 + \delta^2) & -i(\alpha\beta + \gamma\delta) \\ \beta\delta - \alpha\gamma & i(\alpha\gamma + \beta\delta) & \alpha\delta + \beta\gamma \end{bmatrix}.$$

The matrix \mathbf{A} is real in spite of its appearance, as we can see by writing

$$\alpha = e_0 + ie_3$$

$$\beta = e_2 + ie_1,$$

where the four real quantities e_0 , e_1 , e_2 , and e_3 are often referred to as the Cayley–Klein parameters but should be called the *Euler parameters* to be correct. They satisfy the relation

$$e_0^2 + e_1^2 + e_2^2 + e_3^2 = 1.$$

A bit of algebraic manipulation then shows that the matrix \mathbf{A} can be written in terms of the four real parameters in the form

$$\mathbf{A} = \begin{bmatrix} e_0^2 + e_1^2 - e_2^2 - e_3^2 & 2(e_1e_2 + e_0e_3) & 2(e_1e_3 - e_0e_2) \\ 2(e_1e_2 - e_0e_3) & e_0^2 - e_1^2 + e_2^2 - e_3^2 & 2(e_2e_3 + e_0e_1) \\ 2(e_1e_3 + e_0e_2) & 2(e_2e_3 - e_0e_1) & e_0^2 - e_1^2 - e_2^2 + e_3^2 \end{bmatrix}. \quad (4.47')$$

The reality of the matrix elements is now manifest. It can also be easily demonstrated that the matrix \mathbf{A} in terms of these parameters cannot be put in the form of the inversion transformation \mathbf{S} . An examination of the off-diagonal elements and their transposes shows that they all vanish only if at least three of the parameters are zero. We cannot then choose the remaining nonzero parameter such that all three of the diagonal elements (or only one of them) are -1 .

4.6 ■ EULER'S THEOREM ON THE MOTION OF A RIGID BODY

The discussions of the previous sections provide a complete mathematical technique for describing the motions of a rigid body. At any instant, the orientation of the body can be specified by an orthogonal transformation, the elements of which may be expressed in terms of some suitable set of parameters. As time progresses, the orientation will change, and hence the matrix of transformation will be a func-

tion of time and may be written $\mathbf{A}(t)$. If the body axes are chosen coincident with the space axes at the time $t = 0$, then the transformation is initially simply the identity transformation:

$$\mathbf{A}(0) = \mathbf{1}.$$

At any later time, $\mathbf{A}(t)$ will in general differ from the identity transformation, but since the physical motion must be continuous, $\mathbf{A}(t)$ must be a continuous function of time. The transformation may thus be said to evolve *continuously from the identity transformation*.

With this method of describing the motion, and using only the mathematical apparatus already introduced, we are now in a position to obtain the important characteristics of rigid body motion. Of basic importance is:

Euler's Theorem: The general displacement of a rigid body with one point fixed is a rotation about some axis.

The theorem means that for every such rotation it is always possible to find an axis through the fixed point oriented at particular polar angles θ and ϕ such that a rotation by the particular angle ψ about this axis duplicates the general rotation. Thus, three parameters (angles) characterize the general rotation. It is also possible to find three Euler angles to produce the same rotation.

If the fixed point (not necessarily at the center of mass of the object) is taken as the origin of the body set of axes, then the displacement of the rigid body involves no translation of the body axes; the only change is in orientation. The theorem then states that the body set of axes at any time t can always be obtained by a single rotation of the initial set of axes (taken as coincident with the space set). In other words, the *operation* implied in the matrix \mathbf{A} describing the physical motion of the rigid body is a *rotation*. Now it is characteristic of a rotation that one direction, namely, the axis of rotation, is left unaffected by the operation. Thus, any vector lying along the axis of rotation must have the same components in both the initial and final axes.

The other necessary condition for a rotation, that the magnitude of the vectors be unaffected, is automatically provided by the orthogonality conditions. Hence, Euler's theorem will be proven if it can be shown that there exists a vector \mathbf{R} having the same components in both systems. Using matrix notation for the vector,

$$\mathbf{R}' = \mathbf{AR} = \mathbf{R}. \quad (4.48)$$

Equation (4.48) constitutes a special case of the more general equation:

$$\mathbf{R}' = \mathbf{AR} = \lambda \mathbf{R}, \quad (4.49)$$

where λ is some constant, which may be complex. The values of λ for which Eq. (4.49) is soluble are known as the characteristic values, or *eigenvalues*,* of

*This term is derived from the German *Eigenwerte* literally "proper values."

the matrix. Since equations of the form of (4.49) are of general interest and will be used in Chapter 6, we shall examine Eq. (4.49) and then specialize the discussion to Eq. (4.48).

The problem of finding vectors that satisfy Eq. (4.49) is therefore called the *eigenvalue problem* for the given matrix, and Eq. (4.49) itself is referred to as the *eigenvalue equation*. Correspondingly, the vector solutions are the *eigenvectors* of \mathbf{A} . Euler's theorem can now be restated in the following language:

The real orthogonal matrix specifying the physical motion of a rigid body with one point fixed always has the eigenvalue +1.

The eigenvalue equations (4.49) may be written

$$(\mathbf{A} - \lambda \mathbf{1})\mathbf{R} = 0, \quad (4.50)$$

or, in expanded form,

$$\begin{aligned} (a_{11} - \lambda)X + a_{12}Y - a_{13}Z &= 0 \\ a_{21}X + (a_{22} - \lambda)Y - a_{23}Z &= 0 \\ a_{31}X + a_{32}Y + (a_{33} - \lambda)Z &= 0. \end{aligned} \quad (4.51)$$

Equations (4.51) comprise a set of three homogeneous simultaneous equations for the components X, Y, Z of the eigenvector \mathbf{R} . As such, they can never furnish definite values for the three components, but only ratios of components. Physically, this corresponds to the circumstance that only the *direction* of the eigenvector can be fixed; the magnitude remains undetermined. The product of a constant with an eigenvector is also an eigenvector. In any case, being homogeneous, Eqs. (4.51) can have a nontrivial solution only when the determinant of the coefficients vanishes.

$$|\mathbf{A} - \lambda \mathbf{1}| = \begin{vmatrix} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{vmatrix} = 0. \quad (4.52)$$

Equation (4.52) is known as the *characteristic* or *secular* equation of the matrix, and the values of λ for which the equation is satisfied are the desired eigenvalues. Euler's theorem reduces to the statement that, for the real orthogonal matrices under consideration, the secular equation must have the root $\lambda = +1$.

In general, the secular equation will have three roots with three corresponding eigenvectors. For convenience, the notation X_1, X_2, X_3 will often be used instead of X, Y, Z . In such a notation, the components of the eigenvectors might be labeled as X_{ik} , the first subscript indicating the particular component, the second denoting which of the three eigenvectors is involved. A typical member of the group of Eqs. (4.51) would then be written (with explicit summation) as

$$\sum_j a_{ij} X_{jk} = \lambda_k X_{ik}$$

or, alternatively, as

$$\sum_j a_{ij} X_{jk} = \sum_j X_{ij} \delta_{jk} \lambda_k. \quad (4.53)$$

Both sides of Eq. (4.53) then have the form of a matrix product element; the left side as the product of \mathbf{A} with a matrix \mathbf{X} having the elements X_{jk} , the right side as the product of \mathbf{X} with a matrix whose jk th element is $\delta_{jk} \lambda_k$. The last matrix is diagonal, and its diagonal elements are the eigenvalues of \mathbf{A} . We shall therefore designate the matrix by $\boldsymbol{\lambda}$:

$$\boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}. \quad (4.54)$$

Equation (4.53) thus implies the matrix equation

$$\mathbf{AX} = \mathbf{X}\boldsymbol{\lambda},$$

or, multiplying from the left by \mathbf{X}^{-1} ,

$$\mathbf{X}^{-1}\mathbf{AX} = \boldsymbol{\lambda}. \quad (4.55)$$

Now, the left side is in the form of a similarity transformation operating on \mathbf{A} . (We have only to denote \mathbf{X}^{-1} by the symbol \mathbf{Y} to reduce it to the form Eq. (4.41).) Thus, Eq. (4.55) provides the following alternative approach to the eigenvalue problem: We seek to diagonalize \mathbf{A} by a similarity transformation. Each column of the matrix used to carry out the similarity transformation consists of the components of an eigenvector. The elements of the diagonalized form of \mathbf{A} are the corresponding eigenvalues.

Euler's theorem can be proven directly by using the orthogonality property of $\tilde{\mathbf{A}}$. Consider the expression

$$(\mathbf{A} - \mathbf{1})\tilde{\mathbf{A}} = \mathbf{1} - \tilde{\mathbf{A}}.$$

If we take the determinant of the matrices forming both sides (cf. Eq. (4.41')), we can write the equality

$$|\mathbf{A} - \mathbf{1}| |\tilde{\mathbf{A}}| = |\mathbf{1} - \tilde{\mathbf{A}}|. \quad (4.56)$$

To describe the motion of a rigid body, the matrix $\mathbf{A}(t)$ must correspond to a proper rotation; therefore the determinant of \mathbf{A} , and of its transpose, must be +1. Further, since in general the determinant of the transpose of a matrix is the same as that of the matrix, the transpose signs in Eq. (4.56) can be removed:

$$|\mathbf{A} - \mathbf{1}| = |\mathbf{1} - \mathbf{A}|. \quad (4.57)$$

Equation (4.57) says that the determinant of a particular matrix is the same as the determinant of the negative of the matrix. Suppose \mathbf{B} is some $n \times n$ matrix. Then it is a well-known property of determinants that

$$| -\mathbf{B} | = (-1)^n |\mathbf{B}|.$$

Since we are working in a three-dimensional space ($n = 3$), it is clear that Eq. (4.57) can hold for any arbitrary proper rotation only if

$$|\mathbf{A} - \mathbf{1}| = 0. \quad (4.58)$$

Comparing Eq. (4.58) with the secular equation (4.52), we can see that one of the eigenvalues satisfying Eq. (4.52) must always be $\lambda = +1$, which is the desired result of Euler's theorem.

Note how the proof of Euler's theorem emphasizes the importance of the number of dimensions in the space considered. In spaces with an even number of dimensions, Eq. (4.57) is an identity for all matrices and Euler's theorem doesn't hold. Thus, for two dimensions there is no vector *in the space* that is left unaltered by a rotation—the axis of rotation is perpendicular to the plane and therefore out of the space.

It is now a simple matter to determine the properties of the other eigenvalues in three dimensions. Designate the $+1$ eigenvalue as λ_3 . The determinant of any matrix is unaffected by a similarity transformation (cf. Section 4.3). Hence, by Eqs. (4.54) and (4.55) and the properties of \mathbf{A} as a proper rotation,

$$|\mathbf{A}| = \lambda_1 \lambda_2 \lambda_3 = \lambda_1 \lambda_2 = 1. \quad (4.59)$$

Further, since \mathbf{A} is a real matrix, then if λ is a solution of the secular equation (4.52), the complex conjugate λ^* must also be a solution.

If a given eigenvalue λ_i is complex, then the corresponding eigenvector, \mathbf{R}_i , that satisfies Eq. (4.59) will in general also be complex. We have not previously dealt with the properties of complex vectors under (real) orthogonal transformations, and there are some modifications to previous definitions. The square of the length or magnitude of a complex vector \mathbf{R} is $\mathbf{R} \cdot \mathbf{R}^*$, or in matrix notation $\tilde{\mathbf{R}} \mathbf{R}^*$, where the transpose sign on the left-hand vector indicates it is represented by a row matrix. Under a real orthogonal transformation, the square of the magnitude is invariant

$$\tilde{\mathbf{R}}' \mathbf{R}'^* = (\widetilde{\mathbf{A}} \mathbf{R}) \mathbf{A} \mathbf{R}^* = \tilde{\mathbf{R}} \mathbf{A} \mathbf{A} \mathbf{R}^* = \tilde{\mathbf{R}} \mathbf{R}^*.$$

Suppose now that \mathbf{R} is a complex eigenvector corresponding to a complex eigenvalue λ . Hence, by Eq. (4.49), we have

$$\tilde{\mathbf{R}}' \mathbf{R}'^* = \lambda \lambda^* \tilde{\mathbf{R}} \mathbf{R}^*,$$

which leads to the conclusion that *all eigenvalues have unit magnitude*:

$$\lambda\lambda^* = 1. \quad (4.60)$$

From these properties it may be concluded that there are three possible distributions of eigenvalues. If all of the eigenvalues are real, then only two situations are possible:

1. All eigenvalues are +1. The transformation matrix is then just 1, a case we may justly call trivial.
2. One eigenvalue is +1 and the other two are both -1. Such a transformation may be characterized as an inversion in two coordinate axes with the third unchanged. Equally it is a rotation through the angle π about the direction of the unchanged axis.

If not all of the eigenvalues are real, there is only one additional possibility:

3. One eigenvalue is +1, and the other two are complex conjugates of each other of the form $e^{i\Phi}$ and $e^{-i\Phi}$.

A more complete statement of Euler's theorem thus is that any nontrivial real orthogonal matrix has one, *and only one*, eigenvalue +1.

The direction cosines of the axis of rotation can then be obtained by setting $\lambda = 1$ in the eigenvalue equations (4.51) and solving for X , Y , and Z .* The angle of rotation can likewise be obtained without difficulty. By means of some similarity transformation, it is always possible to transform the matrix \mathbf{A} to a system of coordinates where the z axis lies along the axis of rotation. In such a system of coordinates, \mathbf{A}' represents a rotation about the z axis through an angle Φ , and therefore has the form

$$\mathbf{A}' = \begin{bmatrix} \cos \Phi & \sin \Phi & 0 \\ -\sin \Phi & \cos \Phi & 0 \\ 0 & 0 & 1 \end{bmatrix}.$$

The trace of \mathbf{A}' is simply

$$1 + 2 \cos \Phi.$$

Since the trace is always invariant under a similarity transformation, the trace of \mathbf{A} with respect to any initial coordinate system must have the same form,

$$\text{Tr}\mathbf{A} = a_{11} = 1 + 2 \cos \Phi, \quad (4.61)$$

*If there are multiple roots to the secular equation, then the corresponding eigenvectors cannot be found as simply (cf. Sections 5.4 and 6.2). Indeed, it is not always possible to completely diagonalize a general matrix if the eigenvalues are not all distinct. These exceptions are of no importance for the present considerations, as Euler's theorem shows that for all nontrivial orthogonal matrices +1 is a single root.

which gives the value of Φ in terms of the matrix elements. The rotation angle Φ is to be identified also with the phase angle of the complex eigenvalues λ , as the sum of the eigenvalues is just the trace of \mathbf{A} in its diagonal form, Eq. (4.54). By Euler's theorem and the properties of the eigenvalues, this sum is

$$\text{Tr}\mathbf{A} = \sum_i \lambda_i = 1 + e^{i\Phi} + e^{-i\Phi} = 1 + 2 \cos \Phi.$$

We see that the situations in which the eigenvalues are all real are actually special cases of \mathbf{A} having complex eigenvalues. All the $\lambda_i = +1$ corresponds to a rotation angle $\Phi = 0$ (the identity transformation), while the case with a double eigenvalue -1 corresponds to $\Phi = \pi$, as previously noted.

The prescriptions for the direction of the rotation axis and for the rotation angle are not unambiguous. Clearly if \mathbf{R} is an eigenvector, so is $-\mathbf{R}$; hence the sense of the direction of the rotation axis is not specified. Further, $-\Phi$ satisfies Eq. (4.61) if Φ does. Indeed, it is clear that the eigenvalue solution does not uniquely fix the orthogonal transformation matrix \mathbf{A} . From the determinantal secular equation (4.52), it follows that the inverse matrix $\mathbf{A}^{-1} = \tilde{\mathbf{A}}$ has the same eigenvalues and eigenvectors as \mathbf{A} . However, the ambiguities can at least be ameliorated by assigning Φ to \mathbf{A} and $-\Phi$ to \mathbf{A}^{-1} , and fixing the sense of the axes of rotation by the right-hand screw rule.

Finally, note should be made of an immediate corollary of Euler's theorem, sometimes called

Chasles' Theorem: The most general displacement of a rigid body is a translation plus a rotation.

Detailed proof is hardly necessary. Simply stated, removing the constraint of motion with one point fixed introduces three translatory degrees of freedom for the origin of the body system of axes.*

4.7 ■ FINITE ROTATIONS

The relative orientation of two Cartesian coordinate systems with common origin has been described by various representations, including the three successive Euler angles of rotation that transform one coordinate system to the other. In the previous section it was shown that the coordinate transformation can be carried through by a single rotation about a suitable direction. It is therefore natural to seek a representation of the coordinate transformation in terms of the parame-

*M Chasles (1793–1881) also proved a stronger form of the theorem, namely, that it is possible to choose the origin of the body set of coordinates so that the translation is in the same direction as the axis of rotation. Such a combination of translation and rotation is called a *screw motion*.

This formalism has some use in crystallographic studies of crystals with a screw axis of symmetry. Such symmetry produces strange optical properties. Aside from that application, there seems to be little present use for this version of Chasles' theorem, nor for the elaborate mathematics of screw motions developed in the nineteenth century.

ters of the rotation—the angle of rotation and the direction cosines of the axis of rotation.

With the help of some simple vector algebra, we can derive such a representation. For this purpose, it is convenient to treat the transformation in its active sense, i.e., as one that rotates the vector in a fixed coordinate system (cf. Section 4.2). Recall that a counterclockwise rotation of the coordinate system then appears as a *clockwise* rotation of the vector. In Fig. 4.8(a) the initial position of the vector \mathbf{r} is denoted by \overrightarrow{OP} and the final position \mathbf{r}' by \overrightarrow{OQ} , while the unit vector along the axis of rotation is denoted by \mathbf{n} . The distance between O and N has the magnitude $\mathbf{n} \cdot \mathbf{r}$, so that the vector \overrightarrow{ON} can be written as $\mathbf{n}(\mathbf{n} \cdot \mathbf{r})$. Figure 4.8(b) sketches the vectors in the plane normal to the axis of rotation. The vector \overrightarrow{NP} can be described also as $\mathbf{r} - \mathbf{n}(\mathbf{n} \cdot \mathbf{r})$, but its magnitude is the same as that of the vectors \overrightarrow{NQ} and $\mathbf{r} \times \mathbf{n}$. To obtain the desired relation between \mathbf{r}' and \mathbf{r} , we construct \mathbf{r}' as the sum of three vectors:

$$\mathbf{r}' = \overrightarrow{ON} + \overrightarrow{NV} + \overrightarrow{VQ}$$

or

$$\mathbf{r}' = \mathbf{n}(\mathbf{n} \cdot \mathbf{r}) + [\mathbf{r} - \mathbf{n}(\mathbf{n} \cdot \mathbf{r})] \cos \Phi + (\mathbf{r} \times \mathbf{n}) \sin \Phi.$$

A slight rearrangement of terms leads to the final result:

$$\mathbf{r}' = \mathbf{r} \cos \Phi + \mathbf{n}(\mathbf{n} \cdot \mathbf{r})(1 - \cos \Phi) + (\mathbf{r} \times \mathbf{n}) \sin \Phi. \quad (4.62)$$

Equation (4.62) will be referred to as the *rotation formula*. Note that Eq. (4.62) holds for any rotation, no matter what its magnitude, and thus is a finite-rotation version (in a clockwise sense) of the description given in Section 2.6, for the change of a vector under infinitesimal rotation (cf. also Section 4.8).

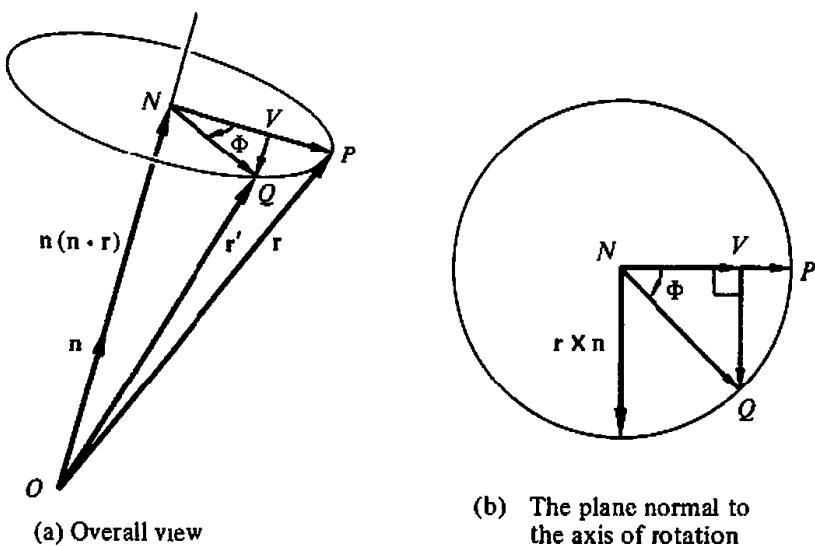


FIGURE 4.8 Vector diagrams for derivation of the rotation formula.

It is straightforward to express the rotation angle, Φ , in terms of the Euler angles. Equation (4.61) gives the trace of the rotation matrix in the plane perpendicular to the axis of rotation. Since the trace of a matrix is invariant, this expression must equal the trace of \mathbf{A} as given in Eq. (4.46). If we use this equality, add one (1) to both sides, and use trigonometric identities, we get an equation whose square root is

$$\cos \frac{\Phi}{2} = \cos \frac{\phi + \psi}{2} \cos \frac{\theta}{2}, \quad (4.63)$$

where the sign of the square root is fixed by the physical requirement that $\Phi \rightarrow 0$ as ϕ , ψ , and $\theta \rightarrow 0$.

4.8 ■ INFINITESIMAL ROTATIONS

In the previous sections various matrices have been associated with the description of the rigid body orientation. However, the number of matrix elements has always been larger than the number of independent variables, and various subsidiary conditions have had to be tagged on. Now that we have established that any given orientation can be obtained by a single rotation about some axis, it is tempting to try to associate a vector, characterized by three independent quantities, with the finite displacement of a rigid body about a fixed point. Certainly a direction suggests itself obviously—that of the axis of rotation—and any function of the rotation angle would seem suitable as the magnitude. But it soon becomes evident that such a correspondence cannot be made successfully. Suppose \mathbf{A} and \mathbf{B} are two such “vectors” associated with transformations \mathbf{A} and \mathbf{B} . Then to qualify as vectors they must be commutative in addition:

$$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}.$$

But the addition of two rotations, i.e., one rotation performed after another, it has been seen, corresponds to the product \mathbf{AB} of the two matrices. However, matrix multiplication is not commutative, $\mathbf{AB} \neq \mathbf{BA}$, and hence \mathbf{A} , \mathbf{B} are not commutative in addition and cannot be accepted as vectors. This conclusion, that the sum of finite rotations depends upon the order of the rotations, is strikingly demonstrated by a simple experiment. Thus, Fig. 4.9 illustrates the sequence of events in rotating a block first through 90° about the z' axis fixed in the block, and then 90° about the y' axis, while Fig. 4.10 presents the same rotations in reverse order. The final position is markedly different in the two sequences.

While a finite rotation thus cannot be represented by a single vector, the same objections do not hold if only *infinitesimal rotations* are considered. An infinitesimal rotation is an orthogonal transformation of coordinate axes in which the components of a vector are almost the same in both sets of axes—the change is infinitesimal. Thus, the x'_1 component of some vector \mathbf{r} (on the passive interpretation of the transformation) would be practically the same as x_1 , the difference

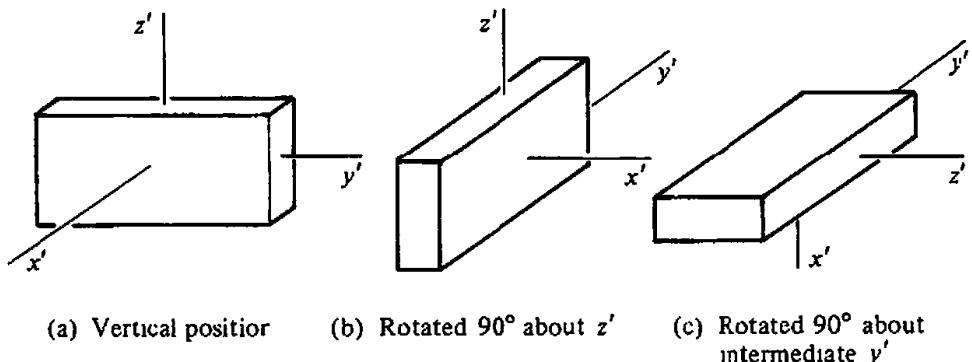


FIGURE 4.9 The effect of two rotations performed in a given order.

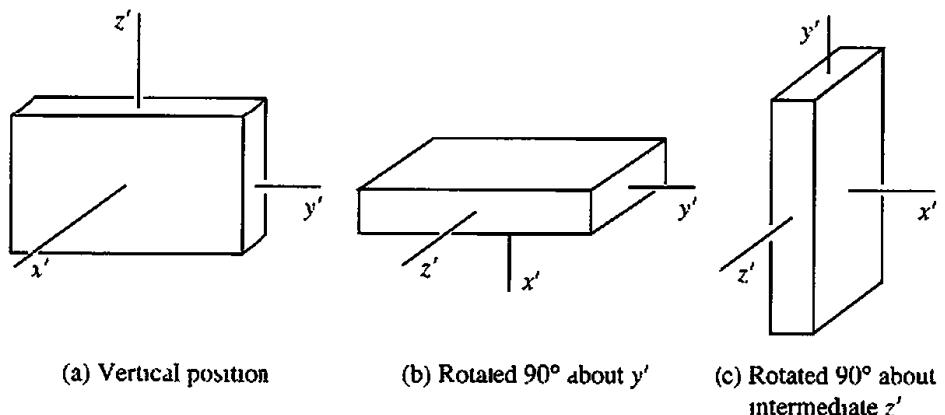


FIGURE 4.10 The two rotations shown in Fig. 4.9, but performed in reverse order.

being extremely small:

$$x'_i = x_i + \epsilon_{11}x_1 + \epsilon_{12}x_2 + \epsilon_{13}x_3. \quad (4.64)$$

The matrix elements ϵ_{11} , ϵ_{12} , etc., are to be considered as infinitesimals, so that in subsequent calculations only the first nonvanishing order in ϵ_{ij} need be retained. For any general component x'_i , the equations of infinitesimal transformation can be written as

$$x'_i = x_i + \epsilon_{ij}x_j$$

or

$$x'_i = (\delta_{ij} + \epsilon_{ij})x_j. \quad (4.65)$$

The quantity δ_{ij} will be recognized as the element of the unit matrix, and Eq. (4.65) appears in matrix notation as

$$\mathbf{x}' = (\mathbf{1} + \boldsymbol{\epsilon})\mathbf{x}. \quad (4.66)$$

Equation (4.66) states that the typical form for the matrix of an infinitesimal transformation is $\mathbf{1} + \boldsymbol{\epsilon}$; i.e., it is almost the identity transformation, differing at most by an infinitesimal operator.

It can now be seen that the sequence of operations is unimportant for infinitesimal transformations; in other words, they *commute*. If $\mathbf{1} + \boldsymbol{\epsilon}_1$ and $\mathbf{1} + \boldsymbol{\epsilon}_2$ are two infinitesimal transformations, then one of the possible products is

$$\begin{aligned} (\mathbf{1} + \boldsymbol{\epsilon}_1)(\mathbf{1} + \boldsymbol{\epsilon}_2) &= \mathbf{1}^2 + \boldsymbol{\epsilon}_1\mathbf{1} + \mathbf{1}\boldsymbol{\epsilon}_2 + \boldsymbol{\epsilon}_1\boldsymbol{\epsilon}_2 \\ &= \mathbf{1} + \boldsymbol{\epsilon}_1 + \boldsymbol{\epsilon}_2, \end{aligned} \quad (4.67)$$

neglecting higher-order infinitesimals. The product in reverse order merely interchanges $\boldsymbol{\epsilon}_1$ and $\boldsymbol{\epsilon}_2$; this has no effect on the result, as matrix addition is always commutative. The commutative property of infinitesimal transformations overcomes the objection to their representation by vectors. For example, the rotation matrix (4.46) for infinitesimal Euler rotation angles is given by

$$\mathbf{A} = \begin{bmatrix} 1 & (d\phi + d\psi) & 0 \\ -(d\phi + d\psi) & 1 & d\theta \\ 0 & -d\theta & 1 \end{bmatrix}$$

and

$$d\Omega = \mathbf{i} d\theta + \mathbf{k} (d\phi + d\psi),$$

where \mathbf{i} and \mathbf{k} are the unit vectors in the x - and z -directions, respectively.

The inverse matrix for an infinitesimal transformation is readily obtained. If $\mathbf{A} = \mathbf{1} + \boldsymbol{\epsilon}$ is the matrix of the transformation, then the inverse is

$$\mathbf{A}^{-1} = \mathbf{1} - \boldsymbol{\epsilon}. \quad (4.68)$$

As proof, note that the product \mathbf{AA}^{-1} reduces to the unit matrix,

$$\mathbf{AA}^{-1} = (\mathbf{1} + \boldsymbol{\epsilon})(\mathbf{1} - \boldsymbol{\epsilon}) = \mathbf{1},$$

in agreement with the definition for the inverse matrix, Eq. (4.32). Further, the orthogonality of \mathbf{A} implies that $\tilde{\mathbf{A}} \equiv (\mathbf{1} + \tilde{\boldsymbol{\epsilon}})$ must be equal to \mathbf{A}^{-1} as given by Eq. (4.68). Hence, the infinitesimal matrix is antisymmetric* (cf. Eq. (4.39)):

$$\tilde{\boldsymbol{\epsilon}} = -\boldsymbol{\epsilon}.$$

Since the diagonal elements of an antisymmetric matrix are necessarily zero, there can be only three distinct elements in any 3×3 antisymmetric matrix. Hence,

*In this section we have assumed implicitly that an infinitesimal orthogonal transformation corresponds to a rotation. In a sense this assumption is obvious; an “infinitesimal inversion” is a contradiction in terms. Formally, the statement follows from the antisymmetry of $\boldsymbol{\epsilon}$. All the diagonal elements of $\mathbf{1} + \boldsymbol{\epsilon}$ are then unity, and to first order in small quantities, the determinant of the transformation is always +, which is the mark of a proper rotation.

there is no loss of generality in writing ϵ in the form

$$\epsilon = \begin{bmatrix} 0 & d\Omega_3 & -d\Omega_2 \\ -d\Omega_3 & 0 & d\Omega_1 \\ d\Omega_2 & -d\Omega_1 & 0 \end{bmatrix} \quad (4.69)$$

The three quantities $d\Omega_1, d\Omega_2, d\Omega_3$ are clearly to be identified with the three independent parameters specifying the rotation. We will now show that these three quantities also form the components of a particular kind of vector. By Eq. (4.66) the *change* in the components of a vector under the infinitesimal transformation of the coordinate system can be expressed by the matrix equation

$$\mathbf{r}' - \mathbf{r} \equiv d\mathbf{r}' = \boldsymbol{\epsilon}\mathbf{r}, \quad (4.70)$$

which in expanded form, with ϵ given by (4.69), becomes

$$\begin{aligned} dx_1 &= x_2 d\Omega_3 - x_3 d\Omega_2 \\ dx_2 &= x_3 d\Omega_1 - x_1 d\Omega_3 \\ dx_3 &= x_1 d\Omega_2 - x_2 d\Omega_1. \end{aligned} \quad (4.71)$$

The right-hand side of each of Eqs. (4.71) is in the form of a component of the cross product of two vectors, namely, the cross product of \mathbf{r} with a vector $d\boldsymbol{\Omega}$ having components* $d\Omega_1, d\Omega_2, d\Omega_3$. We can therefore write Eq. (4.71) equivalently as

$$d\mathbf{r} = \mathbf{r} \times d\boldsymbol{\Omega}. \quad (4.72)$$

The vector \mathbf{r} transforms under an orthogonal matrix \mathbf{B} according to the relations (cf. Eq. (4.20))

$$x'_i = b_{ij} x_j. \quad (4.73)$$

If $d\boldsymbol{\Omega}$ is to be a vector in the same sense as \mathbf{r} , it must transform under \mathbf{B} in the same way. As we shall see, $d\boldsymbol{\Omega}$ passes most of this test for a vector, although in one respect it fails to make the grade. One way of examining the transformation properties of $d\boldsymbol{\Omega}$ is to find how the matrix ϵ transforms under a coordinate transformation. As was shown in Section 4.3, the transformed matrix ϵ' is obtained by a similarity transformation:

$$\epsilon' = \mathbf{B}\epsilon\mathbf{B}^{-1}.$$

*It cannot be emphasized too strongly that $d\boldsymbol{\Omega}$ is *not* the differential of a vector. The combination $d\boldsymbol{\Omega}$ stands for a differential vector, that is, a vector of differential magnitude. Unfortunately, notational convention results in having the vector characteristic applied only to $\boldsymbol{\Omega}$, but it should be clear to the reader there is no vector of which $d\boldsymbol{\Omega}$ represents a differential. As we have seen, a finite rotation *cannot* be represented by a single vector.

As the antisymmetry property of a matrix is preserved under an orthogonal similarity transformation (see Derivation 3), ϵ' can also be put in the form of Eq. (4.69) with nonvanishing elements $d\Omega'_i$. A detailed study of these elements shows that ϵ transforms under the similarity transformation such that

$$d\Omega'_i = |\mathbf{B}| b_{ij} d\Omega_j. \quad (4.74)$$

The transformation of $d\Omega$ is thus almost the same as for \mathbf{r} , but differs by the factor $|\mathbf{B}|$, the determinant of the transformation matrix.

There is however a simpler way to uncover the vector characteristics of $d\Omega$, and indeed to verify its transformation properties as given by Eq. (4.74). In the previous section a vector formula was derived for the change in the components of \mathbf{r} under a finite rotation Φ of the coordinate system. By letting Φ go to the limit of an infinitesimal angle $d\Phi$, the corresponding formula for an infinitesimal rotation can be obtained. In this limit, $\cos \Phi$ in Eq. (4.62) approaches unity, and $\sin \Phi$ goes to Φ ; the resultant expression for the infinitesimal change in \mathbf{r} is then

$$\mathbf{r}' - \mathbf{r} \equiv d\mathbf{r} = \mathbf{r} \times \mathbf{n} d\Phi. \quad (4.75)$$

Comparison with Eq. (4.72) indicates that $d\Omega$ is indeed a vector and is determined by

$$d\Omega = \mathbf{n} d\Phi. \quad (4.76)$$

Equation (4.75) can of course be derived directly without recourse to the finite rotation formula. Considered in its active sense, the infinitesimal coordinate transformation corresponds to a rotation of a vector \mathbf{r} *clockwise* through an angle $d\Phi$ about the axis of rotation, a situation that is depicted in Fig. 4.11.* The magnitude of $d\mathbf{r}$, to first order in $d\Phi$ is, from the figure,

$$dr = r \sin \theta d\Phi,$$

and the direction $d\mathbf{r}$ is, in this limit, perpendicular to both \mathbf{r} and $d\Omega = \mathbf{n} d\Phi$. Finally, the sense of $d\mathbf{r}$ is in the direction a right-hand screw advances as \mathbf{r} is turned into $d\Omega$. Figure 4.11 thus shows that in magnitude, direction, and sense $d\mathbf{r}$ is the same as that predicted by Eq. (4.75).

The transformation properties of $d\Omega$, as defined by Eq. (4.76), are still to be discussed. As is well known from elementary vector algebra, there are two kinds of vectors in regard to transformation properties under an inversion. Vectors that transform according to Eq. (4.72) are known as *polar vectors*. Under a three-dimensional inversion,

$$\mathbf{s} = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

*Figure 4.11 is the clockwise-rotation version of Fig. 2.8.

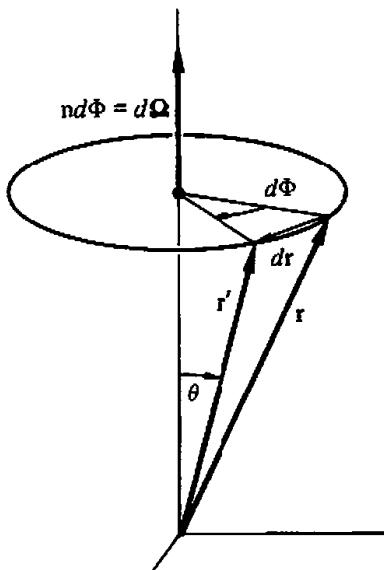


FIGURE 4.11 Change in a vector produced by an infinitesimal clockwise rotation of the vector.

whose components are

$$S_{ij} = -\delta_{ij},$$

all components of a polar vector change sign.

On the other hand, the components of *axial vectors* or *pseudovectors* do not change sign under inversion. The simplest example of an axial vector is a cross product of two polar vectors,

$$\mathbf{V}^* = \mathbf{D} \times \mathbf{F},$$

where the components of the cross product are given, as customary, by the definitions:

$$\mathbf{V}_i^* = D_j F_k - F_j D_k. \quad i, j, k \text{ in cyclic order.} \quad (4.77)$$

The components of \mathbf{D} and \mathbf{F} change sign under inversion; hence those of \mathbf{C} do not. Many familiar physical quantities are axial vectors, such as the angular momentum $\mathbf{L} = \mathbf{r} \times \mathbf{p}$, and the magnetic field intensity. The transformation law for an axial vector is of the form of Eq. (4.74). For proper orthogonal transformations, axial and polar vectors are indistinguishable, but for improper transformations, i.e., involving inversion, the determinant $|\mathbf{V}^*|$ is -1 , and the two types of vectors behave differently.

Another way to explain this property is to define a parity operator \mathbf{P} . The operator \mathbf{P} performs the inversion $x \rightarrow -x$, $y \rightarrow -y$, $z \rightarrow -z$. Then if S is scalar, \mathbf{V} a polar vector, and \mathbf{V}^* an axial vector,

$$\mathbf{P}S = S$$

$$\mathbf{P}\mathbf{V} = -\mathbf{V}$$

$$\mathbf{P}\mathbf{V}^* = \mathbf{V}^*,$$

and, obviously,

$$\mathbf{P}(\mathbf{V} \cdot \mathbf{V}^*) = -(\mathbf{V} \cdot \mathbf{V}^*).$$

Thus, $\mathbf{V} \cdot \mathbf{V}^*$ is a pseudoscalar S^* with the property $\mathbf{P}S^* = -S^*$ and of course $\mathbf{P}(SS^*) = -SS^*$, $\mathbf{P}(S\mathbf{V}) = -S\mathbf{V}$, $\mathbf{P}(S\mathbf{V}^*) = S\mathbf{V}^*$.

On the passive interpretation of the transformation, it is easy to see why polar vectors behave as they do under inversion. The vector remains unaffected by the transformation, but the coordinate axes, and therefore the components, change sign. What then is different for an axial vector? It appears that an axial vector always carries with it a “handedness” convention, as implied, e.g., by the definition, Eq. (4.77), of a cross product. Under inversion a right-handed coordinate system changes to a left-handed system, and the cyclic order requirement of Eq. (4.77) implies a similar change from the right-hand screw convention to a left-hand convention. Hence, even on the passive interpretation, there is an actual change in the direction of the cross product upon inversion.

It is clear now why $d\Omega$ transforms as an axial vector according to Eq. (4.74). Algebraically, we see that since both \mathbf{r} and $d\mathbf{r}$ in Eq. (4.75) are polar vectors, then \mathbf{n} , and therefore $d\Omega$, must be axial vectors. Geometrically, the inversion of the coordinates corresponds to the switch from a right-hand screw law to a left-hand screw to define the sense of \mathbf{n} .

The discussion of the cross product provides an opportunity to introduce a notation that will be most useful on future occasions. The *permutation symbol* or *Levi-Civita density** ϵ_{ijk} is defined to be zero if any two of the indices ijk are equal, and otherwise either +1 or -1 according as ijk is an even or odd permutation of 1, 2, 3. Thus, in terms of the permutation symbol, Eq. (4.77) for the components of a cross product can be written

$$C_i = \epsilon_{ijk} D_j F_k, \quad (4.77')$$

where the usual summation convention has been employed.

The descriptions of rotation presented so far in this chapter have been developed so that we can represent the orientation of a rigid body. Note that the transformations primarily involve rotation of the *coordinate system* (cf. Fig. 4.12a). The corresponding “active” interpretation of rotation of a vector in a fixed coordinate system therefore implies a rotation in the opposite direction, i.e., in a clockwise sense. But there are many areas of mechanics, or of physics in general for that matter, where we are concerned with the effects of rotating the *physical system* and associated vectors (cf. Fig. 4.12b). The connection between invariance of the system under rotation and conservation of angular momentum has already

*Also known interchangeably as the *alternating tensor* or *isotropic tensor of rank 3*.

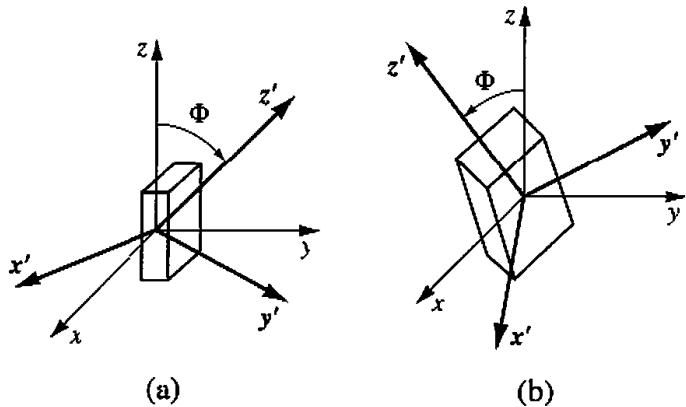


FIGURE 4.12 (a) Transformation from the coordinate system (x, y, z) to a new coordinate system (x', y', z') . By convention, this transformation is considered positive in the clockwise sense. We refer to this as a passive transformation. (b) The rotation of a body through an angle Φ' . By convention, the rotation is positive in a counterclockwise sense. Before the rotation, the coordinates of points of the body were given by (x, y, z) ; after the rotation, they are given by (x', y', z') . This is called an active transformation because the physical body moves.

been pointed out (cf. Section 2.6). In such applications it is necessary to consider the consequences of rotation of vectors in the usual counterclockwise sense. For reference purposes, a number of rotation formulae given above will be listed here, but for counterclockwise rotation of vectors. *All equations and statements from here to the end of this section apply only for such counterclockwise rotations.*

The rotation formula, Eq. (4.62), becomes

$$\mathbf{r}' = \mathbf{r} \cos \Phi + \mathbf{n}(\mathbf{n} \cdot \mathbf{r})(1 - \cos \Phi) + (\mathbf{n} \times \mathbf{r}) \sin \Phi, \quad (4.62')$$

and the corresponding infinitesimal rotation, Eq. (4.75), appears as

$$d\mathbf{r}' = d\Omega \times \mathbf{r} = (\mathbf{n} \times \mathbf{r})d\Phi = -(\mathbf{r} \times \mathbf{n})d\Phi. \quad (4.75')$$

The antisymmetric matrix of the infinitesimal rotation, Eq. (4.69), becomes

$$\epsilon = \begin{bmatrix} 0 & -d\Omega_3 & d\Omega_2 \\ d\Omega_3 & 0 & -d\Omega_1 \\ -d\Omega_2 & d\Omega_1 & 0 \end{bmatrix} = \begin{bmatrix} 0 & -n_3 & n_2 \\ n_3 & 0 & -n_1 \\ -n_2 & n_1 & 0 \end{bmatrix} d\Phi, \quad (4.69')$$

where n_i are the components of the unit vector $\hat{\mathbf{n}}$ along the axis of rotation. Letting $d\mathbf{r}$ stand for the infinitesimal change $\mathbf{r}' - \mathbf{r}$, Eq. (4.66) can then take the form of a matrix differential equation with respect to the rotation angle:

$$\frac{d\mathbf{r}}{d\Phi} = -\mathbf{Nr}, \quad (4.78)$$

where \mathbf{N} is the transpose of the matrix on right in Eq. (4.69') with elements $N_{ij} = \epsilon_{ijk} n_k$.

Another useful representation is to write ϵ in Eq. (4.69') as

$$\epsilon = n_i \mathbf{M}_i d\Phi$$

where \mathbf{M}_i are the three matrices:

$$\mathbf{M}_1 = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \mathbf{M}_2 = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{bmatrix}, \quad \mathbf{M}_3 = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \quad (4.79)$$

The matrices \mathbf{M}_i are known as the *infinitesimal rotation generators* and have the property that their products are

$$\mathbf{M}_i \mathbf{M}_j - \mathbf{M}_j \mathbf{M}_i \equiv [\mathbf{M}_i, \mathbf{M}_j] = \epsilon_{ijk} \mathbf{M}_k. \quad (4.80)$$

The difference between the two matrix products, or *commutator*, is also called the *Lie bracket* or $[\mathbf{M}_i]$, and Eq. (4.80) defines the *Lie algebra* of the rotation group parametrized in terms of the rotation angle. To go further into the group theory of rotation would take us too far afield, but we shall have occasion to refer to these properties of the rotation operation. (cf. Section 9.5 and Appendix B)

4.9 ■ RATE OF CHANGE OF A VECTOR

The concept of an infinitesimal rotation provides a powerful tool for describing the motion of a rigid body in time. Let us consider some arbitrary vector or pseudovector \mathbf{G} involved in the mechanical problem, such as the position vector of a point in the body, or the total angular momentum. Usually such a vector will vary in time as the body moves, but the change will often depend upon the coordinate system to which the observations are referred. For example, if the vector happens to be the radius vector from the origin of the body set of axes to a point in the rigid body, then clearly such a vector appears constant when measured by the body set of axes. However, to an observer fixed in the space set of axes, the components of the vector (as measured on the space axes) will vary in time if the body is in motion.

The change in a time dt of the components of a general vector \mathbf{G} as seen by an observer in the body system of axes will differ from the corresponding change as seen by an observer in the space system. A relation between the two differential changes in \mathbf{G} can be derived on the basis of physical arguments. We can write that the only difference between the two is the effect of rotation of the body axes:

$$(d\mathbf{G})_{\text{space}} = (d\mathbf{G})_{\text{body}} + (d\mathbf{G})_{\text{rot}}.$$

Now consider a vector fixed in the rigid body. As the body rotates, there is of course no change in the components of this vector as seen by the body observer,

i.e., relative to body axes. The only contribution to $(d\mathbf{G})_{\text{space}}$ is then the effect of the rotation of the body. But since the vector is fixed in the body system, it rotates with it *counterclockwise*, and the change in the vector as observed in space is that given by Eq. (4.75'), and hence $(d\mathbf{G})_{\text{rot}}$ is given by

$$(d\mathbf{G})_{\text{rot}} = d\boldsymbol{\Omega} \times \mathbf{G}.$$

For an arbitrary vector, the change relative to the space axes is the sum of the two effects:

$$(d\mathbf{G})_{\text{space}} = (d\mathbf{G})_{\text{body}} + d\boldsymbol{\Omega} \times \mathbf{G}. \quad (4.81)$$

The time *rate of change* of the vector \mathbf{G} as seen by the two observers is then obtained by dividing the terms in Eq. (4.81) by the differential time element dt under consideration;

$$\left(\frac{d\mathbf{G}}{dt} \right)_{\text{space}} = \left(\frac{d\mathbf{G}}{dt} \right)_{\text{body}} + \boldsymbol{\omega} \times \mathbf{G}. \quad (4.82)$$

Here $\boldsymbol{\omega}$ is the instantaneous *angular velocity* of the body defined by the relation*

$$\boldsymbol{\omega} dt = d\boldsymbol{\Omega}. \quad (4.83)$$

The vector $\boldsymbol{\omega}$ lies along the axis of the infinitesimal rotation occurring between t and $t + dt$, a direction known as the *instantaneous axis of rotation*. In magnitude, $\boldsymbol{\omega}$ measures the instantaneous rate of rotation of the body.

A more formal derivation of the basic Eq. (4.82) can be given in terms of the orthogonal matrix of transformation between the space and body coordinates. The component of \mathbf{G} along the i th space axis is related to the components along the body axes:

$$G_i = \tilde{a}_{ij} G'_j = a_{ji} G'_j.$$

As the body moves in time, the components G'_j will change as will the elements a_{ij} of the transformation matrix. Hence, the change in G_i in a differential time element dt is

$$dG_i = a_{ji} dG'_j + da_{ji} G'_j. \quad (4.84)$$

It is no loss of generality to take the space and body axes as instantaneously coincident at the time t . Components in the two systems will then be the same instantaneously, but differentials will *not* be the same, since the two systems are moving relative to each other. Thus, $G'_j = G_j$ but $a_{ji} dG'_j = dG'_j$, the prime emphasizing the differential is measured in the body axis system. The change in the matrix \mathbf{A} in the time dt is thus a change from the unit matrix and therefore

*Note that $\boldsymbol{\omega}$ is *not* the derivative of any vector.

corresponds to the matrix ϵ of the infinitesimal rotation. Hence,

$$da_{ji} = (\tilde{\epsilon})_{ij} = -\epsilon_{ij},$$

using the antisymmetry property of ϵ . In terms of the permutation symbol ϵ_{ijk} , the elements of ϵ are such that (cf. Eq. (4.69))

$$-\epsilon_{ij} = -\epsilon_{ijk} d\Omega_k = \epsilon_{ikj} d\Omega_k.$$

Equation (4.84) can now be written

$$dG_i = dG'_i + \epsilon_{ikj} d\Omega_k G_j.$$

The last term on the right will be recognized as the expression for the i th component of a cross product, so that the final expression for the relation between differentials in the two systems is

$$dG_i = dG'_i + (d\Omega \times \mathbf{G})_i, \quad (4.85)$$

which is the same as the i th component of Eq. (4.81).

Equation (4.81) is not so much an equation about a particular vector \mathbf{G} as it is a statement of the transformation of the time derivative between the two coordinate systems. The arbitrary nature of the vector \mathbf{G} made use of in the derivation can be emphasized by writing Eq. (4.82) as an operator equation acting on some given vector:

$$\left(\frac{d}{dt} \right)_s = \left(\frac{d}{dt} \right)_r + \boldsymbol{\omega} \times . \quad (4.86)$$

Here the subscripts s and r indicate the time derivatives observed in the space and body (rotating) system of axes, respectively. The resultant vector equation can then of course be resolved along any desired set of axes, fixed or moving. But again note that the time rate of change is only relative to the specified coordinate system. When a time derivative of a vector is with respect to one coordinate system, components may be taken along another set of coordinate axes only *after* the differentiation has been carried out.

It is often convenient to express the angular velocity vector in terms of the Euler angles and their time derivatives. The general infinitesimal rotation associated with $\boldsymbol{\omega}$ can be considered as consisting of three successive infinitesimal rotations with angular velocities $\omega_\phi = \dot{\phi}$, $\omega_\theta = \dot{\theta}$, $\omega_\psi = \dot{\psi}$. In consequence of the vector property of infinitesimal rotations, the vector $\boldsymbol{\omega}$ can be obtained as the sum of the three separate angular velocity vectors. Unfortunately, the directions ω_ϕ , ω_θ , and ω_ψ are not symmetrically placed: ω_ϕ is along the space z axis, ω_θ is along the line of nodes, while ω_ψ alone is along the body z' axis. However, the orthogonal transformations \mathbf{B} , \mathbf{C} , \mathbf{D} of Section 4.4 may be used to furnish the components of these vectors along any desired set of axes.

The body set of axes proves most useful for discussing the equations of motion, and we shall therefore obtain the components of ω for such a coordinate system. Since ω_ϕ is parallel to the space z axis, its components along the body axes are given by applying the complete orthogonal transformation $\mathbf{A} = \mathbf{BCD}$, Eq. (4.46):

$$(\omega_\phi)_{x'} = \dot{\phi} \sin \theta \sin \psi, \quad (\omega_\phi)_{y'} = \dot{\phi} \sin \theta \cos \psi, \quad (\omega_\phi)_{z'} = \dot{\phi} \cos \theta.$$

Note that $\dot{\phi}$ has the projection $\dot{\phi} \sin \theta$ in the x' , y' plane, and it is perpendicular to the line of nodes.

The line of nodes, which is the direction of ω_θ , coincides with the ξ' axis, so that the components of ω_θ with respect to the body axes are furnished by applying only the final orthogonal transformation \mathbf{B} , Eq. (4.45):

$$(\omega_\theta)_{x'} = \dot{\theta} \cos \psi, \quad (\omega_\theta)_{y'} = -\dot{\theta} \sin \psi, \quad (\omega_\theta)_{z'} = 0.$$

No transformation is necessary for the components of ω_ψ , which lies along the z' axis. Adding these components of the separate angular velocities, the components of ω with respect to the body axes are

$$\begin{aligned}\omega_{x'} &= \dot{\phi} \sin \theta \sin \psi + \dot{\theta} \cos \psi \\ \omega_{y'} &= \dot{\phi} \sin \theta \cos \psi - \dot{\theta} \sin \psi \\ \omega_{z'} &= \dot{\phi} \cos \theta + \dot{\psi}.\end{aligned}\tag{4.87}$$

Similar techniques may be used to express the components of ω along the space set of axes in terms of the Euler angles.

4.10 ■ THE CORIOLIS EFFECT

Equation (4.86) is the basic kinematical law upon which the dynamical equations of motion for a rigid body are founded. But its validity is not restricted solely to rigid body motion. It may be used whenever we wish to discuss the motion of a particle, or system of particles, relative to a rotating coordinate system.

A particularly important problem in this latter category is the description of particle motion relative to coordinate axes rotating with Earth. Recall that in Section 1.1 an inertial system was defined as one in which Newton's laws of motion are valid. For many purposes, a system of coordinates fixed in the rotating Earth is a sufficient approximation to an inertial system. However, the system of coordinates in which the local stars are fixed comes still closer to the ideal inertial system. Detailed examination shows there are observable effects arising from Earth's rotation relative to this nearly inertial system. Equation (4.86) provides the needed modifications of the equations of motion relative to the *noninertial* system fixed in the rotating Earth.

The initial step is to apply Eq. (4.86) to the radius vector, \mathbf{r} , from the origin of the terrestrial system to the given particle:

$$\mathbf{v}_s = \mathbf{v}_r + \boldsymbol{\omega} \times \mathbf{r}, \quad (4.88)$$

where \mathbf{v}_s and \mathbf{v}_r are the velocities of the particle relative to the space and rotating set of axes, respectively, and $\boldsymbol{\omega}$ is the (constant) angular velocity of Earth relative to the inertial system. In the second step, Eq. (4.86) is used to obtain the time rate of change of \mathbf{v}_s :

$$\begin{aligned} \left(\frac{d\mathbf{v}_s}{dt} \right)_s &= \mathbf{a}_s = \left(\frac{d\mathbf{v}_s}{dt} \right)_r + \boldsymbol{\omega} \times \mathbf{v}_s \\ &= \mathbf{a}_r + 2(\boldsymbol{\omega} \times \mathbf{v}_r) + \boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}), \end{aligned} \quad (4.89)$$

where \mathbf{v}_s has been substituted from Eq. (4.88), and where \mathbf{a}_s and \mathbf{a}_r are the accelerations of the particle in the two systems. Finally, the equation of motion, which in the inertial system is simply

$$\mathbf{F} = m\mathbf{a}_s,$$

expands, when expressed in the rotating coordinates, into the equation

$$\mathbf{F} - 2m(\boldsymbol{\omega} \times \mathbf{v}_r) - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}) = m\mathbf{a}_r. \quad (4.90)$$

To an observer in the rotating system, it therefore appears as if the particle is moving under the influence of an effective force \mathbf{F}_{eff} :

$$\mathbf{F}_{\text{eff}} = \mathbf{F} - 2m(\boldsymbol{\omega} \times \mathbf{v}_r) - m\boldsymbol{\omega} \times (\boldsymbol{\omega} \times \mathbf{r}). \quad (4.91)$$

Let us examine the nature of the terms appearing in Eq. (4.91). The last term is a vector normal to $\boldsymbol{\omega}$ and pointing outward. Further, its magnitude is $m\boldsymbol{\omega}^2 r \sin \theta$. It will therefore be recognized that this term provides the familiar centrifugal force. When the particle is stationary in the moving system, the centrifugal force is the only added term in the effective force. However, when the particle is moving, the middle term known as the *Coriolis effect** comes into play. The order of magnitude of both of these quantities may easily be calculated for a particle on Earth's surface. Earth rotates counterclockwise about the north pole with an angular velocity relative to the fixed stars.

$$\boldsymbol{\omega} = \left(\frac{2\pi}{24 \times 3600} \right) \left(\frac{366.5}{365.5} \right) = 7.292 \times 10^{-5} \text{ s}^{-1}.$$

Here the first set of parentheses gives the angular velocity relative to the radius vector to the Sun. The quantity in the second parentheses, the ratio of the number of sidereal days in a year to the corresponding number of solar days, is the correction factor to give the angular velocity relative to the fixed stars. With this value

*The term *Coriolis effect* is used instead of the older term, Coriolis force, to remind us that this effect exists because we are using a noninertial frame. In a proper inertial frame, the effect does not exist. You can always visualize the Coriolis effect by asking what is happening in an inertial frame.

for ω , and with r equal to Earth's equatorial radius, the maximum centripetal acceleration is

$$\omega^2 r = 3.38 \text{ cm/s}^2,$$

or about 0.3% of the acceleration of gravity. While small, this acceleration is by no means negligible. However, the measured effects of gravity represent the combination of the gravitational field of the mass distribution of Earth and the effects of centripetal acceleration. It has become customary to speak of the sum of the two as Earth's *gravity* field, as distinguished from its *gravitational* field.

The situation is further complicated by the effect of the centripetal acceleration in flattening the rotating Earth. If Earth were completely fluid, the effect of rotation would be to deform it into the shape of an ellipsoid whose surface would be an equipotential surface of the combined gravity field. The mean level of Earth's seas conforms very closely to this equilibrium ellipsoid (except for local variations of wind and tide) and defines what is called the *geoid*.

Except for effects of local perturbations, the force of gravity will be perpendicular to the equipotential surface of the geoid. Accordingly, the local vertical is defined as the direction perpendicular to the geoid at the given point on the surface. For phenomena that occur in the vicinity of a particular spot on Earth, the centripetal acceleration terms in Eq. (4.91) can be considered as swallowed up in the gravitational acceleration g , which will be oriented in the local vertical direction. The magnitude of g of course varies with the latitude on Earth. The effects of centripetal acceleration and the flattening of Earth combine to make g about 0.53% less at the equator than at the poles.

Incidentally, the centrifugal force on a particle arising from Earth's revolution around the Sun is appreciable compared to gravity, but it is almost exactly balanced by the gravitational attraction to the Sun. If we analyze the motion of the Sun-Earth system from a frame rotating with Earth, it is of course just the balance between the centrifugal effect and the gravitational attraction that keeps the Earth (and all that are on it) and Sun separated. An analysis in a Newtonian inertial frame gives a different picture. As was described in Section 3.3, the angular momentum contributes to the effective potential energy to keep the Earth in orbit.

The Coriolis effect on a moving particle is perpendicular to both ω and v .* In the northern hemisphere, where ω points out of the ground, the Coriolis effect $2m(v \times \omega)$ tends to deflect a projectile shot along Earth's surface, to the right of its direction of travel (cf. Fig. 4.13). The Coriolis deflection reverses direction in the southern hemisphere and is zero at the equator, where ω is horizontal. The magnitude of the Coriolis acceleration is always less than

$$2\omega v \simeq 1.5 \times 10^{-4} v,$$

*From here on, the subscript r will be dropped from v as all velocities will be taken with respect to the rotating coordinate axes only

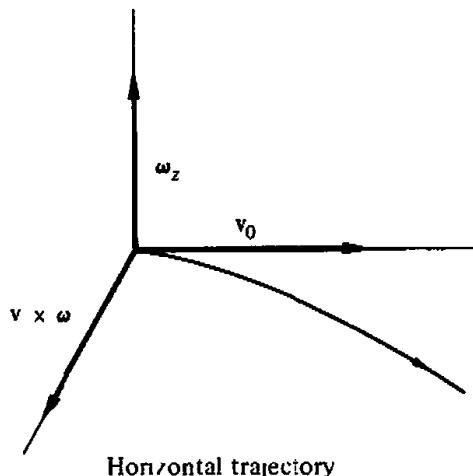


FIGURE 4.13 Direction of Coriolis deflection in the northern hemisphere.

which for a velocity of 10^5 cm/s (roughly 2000 mi/h) is 15 cm/s 2 , or about $0.015g$. Normally, such an acceleration is extremely small, but there are instances when it becomes important. To take an artificial illustration, suppose a projectile were fired horizontally at the north pole. The Coriolis acceleration would then have the magnitude $2\omega v$, so that the linear deflection after a time t is ωvt^2 . The angular deflection would be the linear deflection divided by the distance of travel:

$$\theta = \frac{\omega vt^2}{vt} = \omega t, \quad (4.92)$$

which is the angle Earth rotates in the time t . Physically, this result means that a projectile shot off at the north pole has no initial rotational motion and hence its trajectory in the inertial space is a straight line, the apparent deflection being due to Earth rotating beneath it. Some idea of the magnitude of the effect can be obtained by substituting a time of flight of 100 s—not unusual for large projectiles—in Eq. (4.92). The angular deflection is then of the order of 7×10^{-3} radians, about 0.4° , which is not inconsiderable. Clearly the effect is even more important for long-range missiles, which have a much longer time of flight.

The Coriolis effect also plays a significant role in many oceanographic and meteorological phenomena involving displacements of masses of matter over long distances, such as the circulation pattern of the trade winds and the course of the Gulf stream. A full description of these phenomena requires the solution of complex hydrodynamic problems in which the Coriolis acceleration is only one among many terms involved. It is possible however to give some indication of the contribution of Coriolis effects by considering a highly simplified picture of one particular meteorological problem—the large-scale horizontal wind circulation. Masses of air tend to move, other things being equal, from regions of high pressure to regions of low pressure—the so-called pressure-gradient flow. In the vertical direction the pressure gradient is roughly balanced by gravitational forces so that

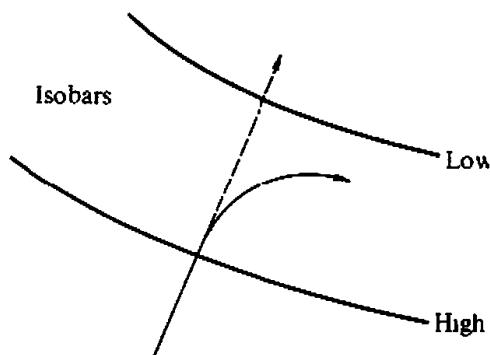


FIGURE 4.14 Deflection of wind from the direction of the pressure gradient by the Coriolis effect (shown for the northern hemisphere).

it is only in the horizontal plane that there are persistent long-range motions of air masses—which we perceive as winds. The pressure gradient forces are quite modest, and comparable in magnitude to the Coriolis effects acting on air masses moving at usual speeds. In the absence of Coriolis effects, the wind directions would ideally be perpendicular to the isobars, as shown in Fig. 4.14. However, the Coriolis effects deflect the wind to the right of this direction in the sense indicated in the figure. The deflection to the right continues until the wind vector is parallel to the isobars and the Coriolis effect is in the opposite direction to, and ideally just balances, the pressure-gradient force. The wind then continues parallel to the isobars, circulating in the northern hemisphere in a counterclockwise direction about a center of low pressure. In the southern hemisphere, the Coriolis effect acts in the opposite direction, and the cyclonic direction (i.e., the flow around a low-pressure center) is clockwise. (Such a wind flow, deflected parallel to the isobars, is known as a *geostrophic wind*.) In this simplified picture, the effect of friction has been neglected. At atmospheric altitudes below several kilometers, the friction effects of eddy viscosity become important, and the equilibrium wind direction never becomes quite parallel to the isobars, as indicated in Fig. 4.15.

Another classical instance where Coriolis effect produces a measurable effect is in the deflection from the vertical of a freely falling particle. Since the particle velocity is almost vertical and ω lies in the north-south vertical plane, the

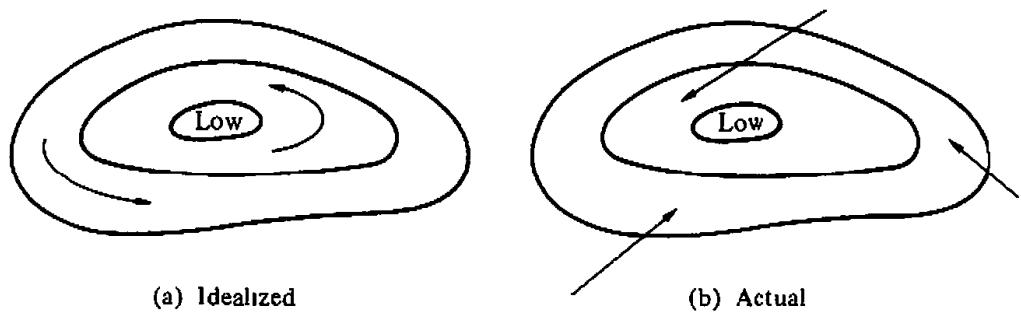


FIGURE 4.15 Cyclone pattern in the northern hemisphere.

deflecting force $2m(\mathbf{v} \times \boldsymbol{\omega})$ is in the east-west direction. Thus, in the northern hemisphere, a body falling freely will be deflected to the East. Calculation of the deflection is greatly simplified by choosing the z axis of the terrestrial coordinate system to be along the direction of the upward vertical as previously defined. If the y axis is taken as pointing North, and the frictional effect of the atmosphere is neglected, then the equation of motion in the x (East) direction is

$$\begin{aligned} m \frac{d^2x}{dt^2} &= -2m(\boldsymbol{\omega} \times \mathbf{v})_x \\ &= -2m\omega v_z \sin \theta, \end{aligned} \quad (4.93)$$

where θ is the co-latitude. The effect of the Coriolis effect on v_z would constitute a small correction to the deflection, which itself is very small. Hence, the vertical velocity appearing in (4.93) may be computed as if Coriolis effects were absent.

$$v_z = -gt.$$

The integral of this is

$$t = \sqrt{\frac{2z}{g}}.$$

With these values, Eq. (4.93) may be easily integrated to give the deflection* as

$$x = \frac{\omega g}{3} t^3 \sin \theta$$

or

$$x = \frac{\omega}{3} \sqrt{\frac{(2z)^3}{g}} \sin \theta.$$

An order of magnitude of the deflection can be obtained by assuming $\theta = \pi/2$ (corresponding to the equator) and $z = 100$ m. The deflection is then, roughly,

$$x \simeq 2.2 \text{ cm.}$$

The actual experiment is difficult to perform, as the small deflection may often be masked by the effects of wind currents, viscosity, or other disturbing influences.[†]

More easily observable is the well-known experiment of the Foucault pendulum. If a pendulum is set swinging at the north pole in a given plane in space, then its linear momentum perpendicular to the plane is zero, and it will continue to swing in this invariable plane while Earth rotates beneath it. To an observer on Earth, the plane of oscillation appears to rotate once a day. At other latitudes the result is more complicated, but the phenomenon is qualitatively the same and detailed calculation will be left as an exercise.

* Again, we neglect the frictional effects of the atmosphere

[†] It is easy to show, using Eq. (4.93), that a particle projected upward will fall back to the ground westward of the original launching spot.

Effects due to the Coriolis terms also appear in atomic physics. Thus, two types of motion may occur simultaneously in polyatomic molecules: The molecule *rotates* as a rigid whole, and the atoms *vibrate* about their equilibrium positions. As a result of the vibrations, the atoms are in motion relative to the rotating coordinate system of the molecule. The Coriolis term will then be different from zero and will cause the atoms to move in a direction perpendicular to the original oscillations. Perturbations in molecular spectra due to Coriolis effects thus appear as interactions between the rotational and vibrational motions of the molecule.

DERIVATIONS

1. Prove that matrix multiplication is associative. Show that the product of two orthogonal matrices is also orthogonal.
2. Prove the following properties of the transposed and adjoint matrices:

$$\tilde{AB} = \tilde{B}\tilde{A},$$

$$(AB)^\dagger = B^\dagger A^\dagger.$$

3. Show that the trace of a matrix is invariant under any similarity transformation. Show also that the antisymmetry property of a matrix is preserved under an orthogonal similarity transformation.
4. (a) By examining the eigenvalues of an antisymmetric 3×3 real matrix A , show that $\mathbf{1} \pm A$ is nonsingular.
 (b) Show then that under the same conditions the matrix

$$B = (1 + A)(1 - A)^{-1}$$

is orthogonal.

5. Obtain the matrix elements of the general rotation matrix in terms of the Euler angles, Eq. (4.46), by performing the multiplications of the successive component rotation matrices. Verify directly that the matrix elements obey the orthogonality conditions.
6. The body set of axes can be related to the space set in terms of Euler's angles by the following set of rotations:
 - (a) Rotation about the x axis by an angle θ
 - (b) Rotation about the z' axis by an angle ψ .
 - (c) Rotation about the *old* z axis by an angle ϕ .
 Show that this sequence leads to the same elements of the matrix of transformation as the sequence of rotations given in the book. [Hint: It is not necessary to carry out the explicit multiplication of the rotation matrices.]
7. If A is the matrix of a rotation through 180° about any axis, show that if

$$P_\pm = \frac{1}{2}(1 \pm A),$$

then $\mathbf{P}_{\pm}^2 = \mathbf{P}_{\pm}$. Obtain the elements of \mathbf{P}_{\pm} in any suitable system, and find a geometric interpretation of the operation \mathbf{P}_+ and \mathbf{P}_- on any vector \mathbf{F} .

8. (a) Show that the rotation matrix in the form of Eq. (4.47') cannot be put in the form of the matrix of the inversion transformation \mathbf{S} .
- (b) Verify by direct multiplication that the matrix in Eq. (4.47') is orthogonal.
9. Show that any rotation can be represented by successive reflection in two planes, both passing through the axis of rotation with the planar angle $\Phi/2$ between them.
10. If \mathbf{B} is a square matrix and \mathbf{A} is the exponential of \mathbf{B} , defined by the infinite series expansion of the exponential,

$$\mathbf{A} \equiv e^{\mathbf{B}} = 1 + \mathbf{B} + \frac{1}{2}\mathbf{B}^2 + \cdots + \frac{\mathbf{B}^n}{n!} + \cdots,$$

then prove the following properties:

- (a) $e^{\mathbf{B}}e^{\mathbf{C}} = e^{\mathbf{B}+\mathbf{C}}$, providing \mathbf{B} and \mathbf{C} commute
- (b) $\mathbf{A}^{-1} = e^{-\mathbf{B}}$
- (c) $e^{\mathbf{C}\mathbf{B}\mathbf{C}^{-1}} = \mathbf{C}\mathbf{A}\mathbf{C}^{-1}$
- (d) \mathbf{A} is orthogonal if \mathbf{B} is antisymmetric.

11. Verify the relation

$$| - \mathbf{B}| = (-1)^n |\mathbf{B}|$$

for the determinant of an $n \times n$ matrix \mathbf{B} .

12. In a set of axes where the z axis is the axis of rotation of a finite rotation, the rotation matrix is given by Eq. (4.43) with θ replaced by the angle of finite rotation Φ . Derive the rotation formula, Eq. (4.62), by transforming to an arbitrary coordinate system, expressing the orthogonal matrix of transformation in terms of the direction cosines of the axis of the finite rotation.
13. (a) Suppose two successive coordinate rotations through angles Φ_1 and Φ_2 are carried out, equivalent to a single rotation through an angle Φ . Show that Φ_1 , Φ_2 , and Φ can be considered as the sides of a spherical triangle with the angle opposite to Φ given by the angle between the two axes of rotation.
- (b) Show that a rotation about any given axis can be obtained as the product of two successive rotations, each through 180° .
14. (a) Verify that the permutation symbol satisfies the following identity in terms of Kronecker delta symbols:

$$\epsilon_{ijp}\epsilon_{rmp} = \delta_{ir}\delta_{jm} - \delta_{im}\delta_{jr}.$$

- (b) Show that

$$\epsilon_{ijp}\epsilon_{ijk} = 2\delta_{pk}.$$

15. Show that the components of the angular velocity along the space set of axes are given in terms of the Euler angles by

$$\begin{aligned}\omega_x &= \dot{\theta} \cos \phi + \dot{\psi} \sin \theta \sin \phi, \\ \omega_y &= \dot{\theta} \sin \phi - \dot{\psi} \sin \theta \cos \phi, \\ \omega_z &= \dot{\psi} \cos \theta + \dot{\phi}.\end{aligned}$$

16. Show that the Euler parameter e_0 has the equation of motion

$$-2\dot{e}_0 = e_1 \omega_{x'} + e_2 \omega_{y'} + e_3 \omega_{z'},$$

where the prime denotes the body set of axes. Find the corresponding equations for the other three Euler parameters and for the complex Cayley–Klein parameters α and β .

17. Verify directly that the matrix generators of infinitesimal rotation, M_i , as given by Eq. (4.79) obey the commutation relations

$$[M_i, M_j] = \epsilon_{ijk} M_k.$$

18. (a) Find the vector equation describing the reflection of \mathbf{r} in a plane whose unit normal is \mathbf{n} .
 (b) Show that if l_i , $i = 1, 2, 3$, are the direction cosines of \mathbf{n} , then the matrix of transformation has the elements

$$A_{ij} = \delta_{ij} - 2l_i l_j,$$

and verify that A is an improper orthogonal matrix.

19. Figures 4.9 and 4.10 show that the order of finite rotations leads to different results. Use the notation that $\mathbf{A}(\alpha, \mathbf{l}_n)$ where \mathbf{A} is a rotation in the direction of \mathbf{l}_n through an angle α . Let n_1 and n_2 be two orthogonal directions.

- (a) If \mathbf{x} is the position vector of a point on a rigid body, which is then rotated by an angle θ around the origin, show that the new value of \mathbf{x} is

$$\mathbf{x}' = (\mathbf{l}_n \cdot \mathbf{x})\mathbf{l}_n + [\mathbf{x} - \mathbf{l}_n(\mathbf{l}_n \cdot \mathbf{x})]\cos \theta - \mathbf{l}_n \times \mathbf{x} \sin \theta.$$

From this, obtain the formula for $\mathbf{A}(\pi/2, \mathbf{l}_n)$ and derive the two rotations in the figures.

- (b) Discuss these two rotations. [Hint: The answer will involve a rotation by the angle $\frac{2}{3}\pi$ in a direction $(1/\sqrt{3})(1, 1, 1)$.]

20. Express the “rolling” constraint of a sphere on a plane surface in terms of the Euler angles. Show that the conditions are nonintegrable and that the constraint is therefore nonholonomic.

EXERCISES

21. A particle is thrown up vertically with initial speed v_0 , reaches a maximum height and falls back to ground. Show that the Coriolis deflection when it again reaches the ground is opposite in direction, and four times greater in magnitude, than the Coriolis deflection when it is dropped at rest from the same maximum height.

22. A projectile is fired horizontally along Earth's surface. Show that to a first approximation the angular deviation from the direction of fire resulting from the Coriolis effect varies linearly with time at a rate

$$\omega \cos \theta,$$

where ω is the angular frequency of Earth's rotation and θ is the co-latitude, the direction of deviation being to the right in the northern hemisphere.

23. The Foucault pendulum experiment consists in setting a long pendulum in motion at a point on the surface of the rotating Earth with its momentum originally in the vertical plane containing the pendulum bob and the point of suspension. Show that the pendulum's subsequent motion may be described by saying that the plane of oscillation rotates uniformly $2\pi \cos \theta$ radians per day, where θ is the co-latitude. What is the direction of rotation? The approximation of small oscillations may be used, if desired.
24. A wagon wheel with spokes is mounted on a vertical axis so it is free to rotate in the horizontal plane. The wheel is rotating with an angular speed of $\omega = 3.0$ radian/s. A bug crawls out on one of the spokes of the wheel with a velocity of 0.5 cm/s holding on to the spoke with a coefficient of friction $\mu = 0.30$. How far can the bug crawl along the spoke before it starts to slip?
25. A carousel (counter-clockwise merry-go-round) starts from rest and accelerates at a constant angular acceleration of 0.02 revolutions/s². A girl sitting on a bench on the platform 7.0 m from the center is holding a 3.0 kg ball. Calculate the magnitude and direction of the force she must exert to hold the ball 6.0 s after the carousel starts to move. Give the direction with respect to the line from the center of rotation to the girl.

CHAPTER

5

The Rigid Body Equations of Motion

Chapter 4 presents all the kinematical tools needed in the discussion of rigid body motion. In the Euler angles we have a set of three coordinates, defined rather unsymmetrically it is true, yet suitable for use as the generalized coordinates describing the orientation of the rigid body. In addition, the method of orthogonal transformations, and the associated matrix algebra, furnish a powerful and elegant technique for investigating the characteristics of rigid body motion. We have already had one application of the technique in deriving Eq. (4.86), the relation between the states of change of a vector as viewed in the space system and in the body system. These tools will now be applied to obtain the Euler dynamical equations of motion of the rigid body in their most convenient form. With the help of the equations of motion, some simple but highly important problems of rigid body motion can be discussed.

5.1 ■ ANGULAR MOMENTUM AND KINETIC ENERGY OF MOTION ABOUT A POINT

Chasles' theorem states that any general displacement of a rigid body can be represented by a translation plus a rotation. The theorem suggests that it ought to be possible to split the problem of rigid body motion into two separate phases, one concerned solely with the translational motion of the body, the other, with its rotational motion. Of course, if one point of the body is fixed, the separation is obvious, for then there is only a rotational motion about the fixed point, without any translation. But even for a general type of motion such a separation is often possible. The six coordinates needed to describe the motion have already been formed into two sets in accordance with such a division: the three Cartesian coordinates of a point fixed in the rigid body to describe the translational motion and, say, the three Euler angles for the motion about the point. If, further, the origin of the body system is chosen to be the center of mass, then by Eq. (1.28) the total angular momentum divides naturally into contributions from the translation of the center of mass and from the rotation about the center of mass. The former term will involve only the Cartesian coordinates of the center of mass, the latter only the angle coordinates. By Eq. (1.31), a similar division holds for the total kinetic energy T , which can be written in the form

$$T = \frac{1}{2} M v^2 + T'(\phi, \theta, \psi),$$

as the sum of the kinetic energy of the entire body as if concentrated at the center of mass, plus the kinetic energy of motion about the center of mass.

Often the potential energy can be similarly divided, each term involving only one of the coordinate sets, either the translational or rotational. Thus, the potential energy in a uniform gravitational field will depend only upon the Cartesian vertical coordinate of the center of gravity.* Or if the force on a body is due to a uniform magnetic field, \mathbf{B} , acting on its magnetic dipole moment, \mathbf{M} , then the potential is proportional to $\mathbf{M} \cdot \mathbf{B}$, which involves only the orientation of the body. Certainly, almost all problems soluble in practice will allow for such a separation. In such a case, the entire mechanical problem does indeed split into two. The Lagrangian, $L = T - V$, divides into two parts, one involving only the translational coordinates, the other only the angle coordinates. These two groups of coordinates will then be completely separated, and the translational and rotational problems can be solved independently of each other.

It is of obvious importance therefore to obtain expressions for the angular momentum and kinetic energy of the motion about some point fixed in the body. To do so, we will make abundant use of Eq. (4.86) linking derivatives relative to a coordinate system fixed at some point in the rigid body. It is intuitively obvious that the rotation angle of a rigid body displacement, as also the instantaneous angular velocity vector, is independent of the choice of origin of the body system of axes. The essence of the rigid body constraint is that all particles of the body move and rotate together. However, a formal proof is easily constructed.

Let \mathbf{R}_1 and \mathbf{R}_2 be the position vectors, relative to a fixed set of coordinates, of the origins of two sets of body coordinates (cf. Fig. 5.1). The difference vector is denoted by \mathbf{R} :

$$\mathbf{R}_2 = \mathbf{R}_1 + \mathbf{R}.$$

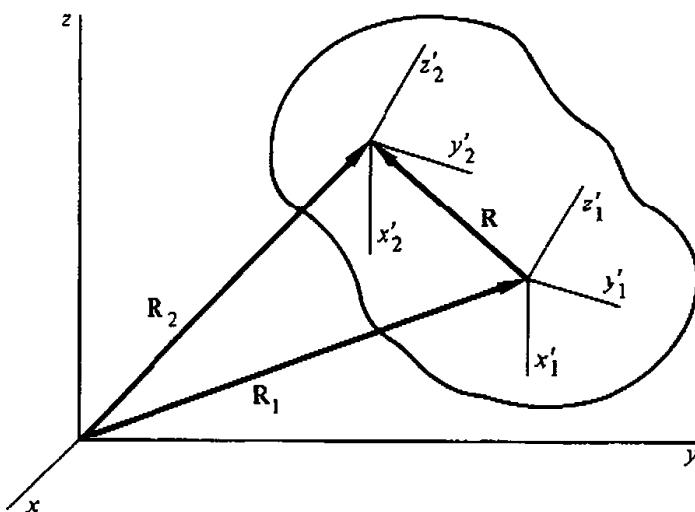


FIGURE 5.1 Vectorial relation between sets of rigid body coordinates with different origins.

*The center of gravity of course coincides with the center of mass in a uniform gravitational field.

Chapter 5 The Rigid Body Equations of Motion

If the origin of the second set of axes is considered as a point defined relative to the first, then the time derivative of \mathbf{R}_2 relative to the space axes is given by

$$\left(\frac{d\mathbf{R}_2}{dt} \right)_s - \left(\frac{d\mathbf{R}_1}{dt} \right)_s + \left(\frac{d\mathbf{R}}{dt} \right)_s - \left(\frac{d\mathbf{R}_1}{dt} \right)_s + \boldsymbol{\omega}_1 \times \mathbf{R}.$$

The last step follows from Eq. (4.86), recalling that the derivatives of \mathbf{R} relative to any rigid body axes must vanish, and with $\boldsymbol{\omega}_1$ as being the angular velocity vector appropriate to the first coordinate system. Alternatively, the origin of the first coordinate system can be considered as fixed in the second system with the position vector $-\mathbf{R}$. In the same manner, then, the derivative of the position vector \mathbf{R}_1 to this origin relative to the fixed-space axes can be written as

$$\left(\frac{d\mathbf{R}_1}{dt} \right)_s = \left(\frac{d\mathbf{R}_2}{dt} \right)_s - \left(\frac{d\mathbf{R}}{dt} \right)_s = \left(\frac{d\mathbf{R}_2}{dt} \right)_s - \boldsymbol{\omega}_2 \times \mathbf{R}.$$

A comparison of these two expressions shows $(\boldsymbol{\omega}_1 - \boldsymbol{\omega}_2) \times \mathbf{R} = 0$. Any difference in the angular velocity vectors at two arbitrary points must lie along the line joining the two points. Assuming the $\boldsymbol{\omega}$ vector field is continuous, the only possible solution for all pairs of points is that the two angular velocity vectors must be equal:

$$\boldsymbol{\omega}_1 = \boldsymbol{\omega}_2.*$$

The angular velocity vector is the same for all coordinate systems fixed in the rigid body.

When a rigid body moves with one point stationary, the total angular momentum about that point is

$$\mathbf{L} = m_i (\mathbf{r}_i \times \mathbf{v}_i), \quad (5.1)$$

(employing the summation convention) where \mathbf{r}_i and \mathbf{v}_i are the radius vector and velocity, respectively, of the i th particle relative to the given point. Since \mathbf{r}_i is a fixed vector relative to the body, the velocity \mathbf{v}_i with respect to the space set of axes arises solely from the rotational motion of the rigid body about the fixed point. From Eq. (4.86), \mathbf{v}_i is then

$$\mathbf{v}_i = \boldsymbol{\omega} \times \mathbf{r}_i. \quad (5.2)$$

Hence, Eq. (5.1) can be written as

$$\mathbf{L} = m_i [\mathbf{r}_i \times (\boldsymbol{\omega} \times \mathbf{r}_i)],$$

or, expanding the triple cross product,

$$\mathbf{L} = m_i [\boldsymbol{\omega} r_i^2 - \mathbf{r}_i (\mathbf{r}_i \cdot \boldsymbol{\omega})]. \quad (5.3)$$

*See also N. A. Lemos, *Am. Jr. Phys.*, **68**(7) 2000, p. 668–669.

Again expanding, the x -component of the angular momentum becomes

$$L_x = \omega_x m_i (r_i^2 - x_i^2) - \omega_y m_i x_i y_i - \omega_z m_i x_i z_i, \quad (5.4)$$

with similar equations for the other components of \mathbf{L} . Thus, each component of the angular momentum is a linear function of all the components of the angular velocity. *The angular momentum vector is related to the angular velocity by a linear transformation.* To emphasize the similarity of (5.4) with the equations of a linear transformation, (4.12), we may write L_x as

$$L_x = I_{xx} \omega_x + I_{xy} \omega_y + I_{xz} \omega_z.$$

Analogously, for L_y and L_z we have

$$L_y = I_{yx} \omega_x + I_{yy} \omega_y + I_{yz} \omega_z, \quad (5.5)$$

$$L_z = I_{zx} \omega_x + I_{zy} \omega_y + I_{zz} \omega_z.$$

The nine coefficients I_{xx} , I_{xy} , etc., are the nine elements of the transformation matrix. The diagonal elements are known as *moment of inertia coefficients*, and have the following form

$$I_{xx} = m_i (r_i^2 - x_i^2), \quad (5.6)$$

while the off-diagonal elements are designated as *products of inertia*, a typical one being

$$I_{xy} = -m_i x_i y_i. \quad (5.7)$$

In Eqs. (5.6) and (5.7), the matrix elements appear in the form suitable if the rigid body is composed of discrete particles. For continuous bodies the summation is replaced by a volume integration, with the particle mass becoming a mass density. Thus, the diagonal element I_{xx} appears as

$$I_{xx} = \int_V \rho(\mathbf{r}) (r^2 - x^2) dV. \quad (5.6')$$

With a slight change in notation, an expression for all matrix elements can be stated for continuous bodies. If the coordinate axes are denoted by x_j , $j = 1, 2, 3$, then the matrix element I_{jk} can be written

$$I_{jk} = \int_V \rho(\mathbf{r}) (r^2 \delta_{jk} - x_j x_k) dV. \quad (5.8)$$

Thus far, the coordinate system used in resolving the components of \mathbf{L} has not been specified. From now on, we will take it to be a system fixed in the body.* The various distances x_i , y_i , z_i are then constant in time, so that the matrix elements are likewise constants, peculiar to the body involved, and dependent on the origin and orientation of the particular body set of axes in which they are expressed.

Equations (5.5) relating the components of \mathbf{L} and $\boldsymbol{\omega}$ can be summarized by a single operator equation,

$$\mathbf{L} = \mathbf{I}\boldsymbol{\omega}, \quad (5.9)$$

where the symbol \mathbf{I} stands for the operator whose matrix elements are the inertia coefficients appearing in (5.5), and $\boldsymbol{\omega}$ and \mathbf{L} are column matrices. Of the two interpretations that have been given to the operator of a linear transformation (cf. Section 4.2), it is clear that here \mathbf{I} must be thought of as acting upon the vector $\boldsymbol{\omega}$, and not upon the coordinate system. The vectors \mathbf{L} and $\boldsymbol{\omega}$ are two physically different vectors, having different dimensions, and are not merely the same vector expressed in two different coordinate systems. Unlike the operator of rotation, \mathbf{I} will have dimensions—mass times length squared—and it is not restricted by any orthogonality conditions. Equation (5.9) is to be read as the operator \mathbf{I} acting upon the vector $\boldsymbol{\omega}$ results in the physically new vector \mathbf{L} .

While full use will be made of the matrix algebra techniques developed in the discussion of the rotation operator, more attention must be paid here to the nature and physical character of the operator per se. However, a certain amount of preliminary mathematical formalism needs first to be discussed. Those already familiar with tensors can proceed immediately to Section 5.3.

5.2 ■ TENSORS

The quantity \mathbf{I} may be considered as defining the quotient of \mathbf{L} and $\boldsymbol{\omega}$ for the product of \mathbf{I} and $\boldsymbol{\omega}$ gives \mathbf{L} . Now, the quotient of two quantities is often not a member of the same class as the dividing factors, but may belong to a more complicated class. Thus, the quotient of two integers is in general not an integer but rather a rational number. Similarly, the quotient of two vectors, as is well known, cannot be defined consistently within the class of vectors. It is not surprising, therefore to find that \mathbf{I} is a new type of quantity, a *tensor of the second rank*.

In a Cartesian three-dimensional space, a tensor \mathbf{T} of the N th rank may be defined for our purposes as a quantity having 3^N components T_{ijk} (with N indices) that transform under an orthogonal transformation of coordinates, \mathbf{A} , according to

*In Chapter 4, such a system was denoted by primes. As components along spatial axes are rarely used here, this convention will be dropped from now on to simplify the notation. Unless otherwise specified, all coordinates used for the rest of the chapter refer to systems fixed in the rigid body.

the following scheme:^{*}

$$T'_{ijk}(\mathbf{x}') = a_{il}a_{jm}a_{kn} \dots T_{lmn}(\mathbf{x}). \quad (5.10)$$

By this definition, a tensor of the zero rank has one component, which is invariant under an orthogonal transformation. Hence, a *scalar is a tensor of zero rank*. A tensor of the first rank has three components transforming as

$$T'_i = a_{ij}T_j.$$

Comparison with the transformation equations for a vector, (4.12'), shows that a *tensor of the first rank is completely equivalent to a vector*.[†] Finally, the nine components of a tensor of the second rank transform as

$$T'_{ij} = a_{ik}a_{jl}T_{kl}. \quad (5.11)$$

Rigorously speaking, we must distinguish between a second-rank tensor \mathbf{T} and the square matrix formed from its components. A tensor is defined only in terms of its transformation properties under orthogonal coordinate transformations. On the other hand, a matrix is in no way restricted in the types of transformations it may undergo and indeed may be considered entirely independently of its properties under some particular class of transformations. Nevertheless, the distinction must not be stressed unduly. Within the restricted domain of orthogonal transformations, there is a practical identity. The tensor components and the matrix elements are manipulated in the same fashion; for every tensor equation there will be a corresponding matrix equation, and vice versa. By Eq. (4.41), the components of a square matrix \mathbf{T} transform under a linear change of coordinates defined by the matrix \mathbf{A} according to a similarity transformation:

$$\mathbf{T}' = \mathbf{ATA}^{-1}.$$

For an orthogonal transformation, we therefore have

$$\mathbf{T}' = \mathbf{ATA}' \quad (5.12)$$

^{*}In a Cartesian space (that is, with orthogonal straight-line axes) there is no distinction between "covariant" and "contravariant" indices, and the terminology will not be needed. Indeed, strictly speaking the tensors defined here should be denoted as "Cartesian tensors." As this is the only type of tensor that will be used in this book (except in Chapters 7 and 13), the adjective will be omitted in subsequent discussions.

[†]A *pseudotensor* in three dimensions transforms as a tensor except under inversion. In general, the transformation equation for a pseudotensor \mathbf{T}^* of the N th rank is (cf. Eq. (4.74))

$$T'^*_{ijk} = |\mathbf{A}|a_{il}a_{jm}a_{kn} T^*_{lmn},$$

and the parity operation P gives

$$PT^* = (-1)^{N+1}\mathbf{T}^*$$

As rigid body motion involves only proper rotations no further use will be made here of the general pseudotensor

or

$$T'_{ij} = a_{ik} T_{kl} a_{jl}. \quad (5.13)$$

Comparison with Eq. (5.11) thus shows that the matrix components transform identically, under an orthogonal transformation, with the components of a tensor of the second rank. All the terminology and operations of matrix algebra, such as “transpose” and “antisymmetrical” can be applied to tensors without change. The equivalence between the tensor and the matrix is not restricted to tensors of the second rank. For example, we already know that the components of a vector, which is a tensor of the first rank, form a column or row matrix and vector manipulation may be treated completely in terms of these associated matrices.

Two vectors can be used to construct a second-rank tensor, \mathbf{T} . Let \mathbf{A} and \mathbf{B} be vectors with components A_i and B_i and construct the tensor \mathbf{T} , by

$$\mathbf{T}_{ij} = A_i B_j. \quad (5.14)$$

For example, if \mathbf{A} and \mathbf{B} are two-dimensional vectors,*

$$\mathbf{T} = \begin{pmatrix} T_{xx} & T_{xy} \\ T_{yx} & T_{yy} \end{pmatrix} = \begin{pmatrix} A_x B_x & A_x B_y \\ A_y B_x & A_y B_y \end{pmatrix}.$$

Since each individual vector transforms as a vector under a Cartesian transformation, each component of \mathbf{T} will transform as required by Eq. (5.10). For example,

$$\mathbf{T}'_{xy} = \sum_{i=1}^3 \sum_{j=1}^3 a_{xi} a_{yj} T_{ij} = a_{xi} a_{yj} A_i B_j = a_{xi} A_i a_{yj} B_j = A'_x B'_y,$$

so \mathbf{T} is a tensor.

The types of operations performed with vectors can be combined with tensors in an obvious way. There is a unit tensor, $\mathbf{1}$, whose components are

$$\mathbf{1}_{ij} = \delta_{ij} \quad (5.15)$$

where δ_{ij} is the delta function (also called the Kronecker delta), $\delta_{ij} = 1$ if $i = j$, and zero otherwise. The dot product on the right of a tensor \mathbf{T} with a vector \mathbf{C} is defined as the vector \mathbf{D} by

$$\mathbf{D} = \mathbf{T} \cdot \mathbf{C} \quad \text{where } D_i = \sum_{j=1}^3 T_{ij} C_j = T_{ij} C_j,$$

*To distinguish between matrices which are transformations and tensors which are physical quantities we use [] for matrices and () for tensors.

and the dot product on the left with a vector \mathbf{F} is defined as the vector \mathbf{E} by

$$\mathbf{E} = \mathbf{F} \cdot \mathbf{T} \quad \text{where } E_i = \sum_{j=1}^3 F_j T_{ji} = F_j T_{ji}.$$

A scalar S can be constructed by a double dot product

$$S = \mathbf{F} \cdot \mathbf{T} \cdot \mathbf{C} \quad \text{where } S = \sum_{i=1}^3 \sum_{j=1}^3 F_i T_{ij} C_j = F_i T_{ij} C_j.$$

These processes are termed contraction. If the tensor \mathbf{T} is constructed of two vectors \mathbf{A} and \mathbf{B} as in Eq. (5.14), then

$$\mathbf{T} \cdot \mathbf{C} = \mathbf{A}(\mathbf{B} \cdot \mathbf{C}) = (\mathbf{B} \cdot \mathbf{C})\mathbf{A}, \quad \text{and} \quad \mathbf{F} \cdot \mathbf{T} = (\mathbf{F} \cdot \mathbf{A})\mathbf{B} = (\mathbf{A} \cdot \mathbf{F})\mathbf{B}.$$

5.3 ■ THE INERTIA TENSOR AND THE MOMENT OF INERTIA

Considered as a linear operator that transforms $\boldsymbol{\omega}$ into \mathbf{L} , the matrix \mathbf{I} has elements that behave as the elements of a second-rank tensor. The quantity \mathbf{I} is therefore identified as a second-rank tensor and is usually called the *moment of inertia tensor* or briefly the *inertia tensor*.

The kinetic energy of motion about a point is

$$T = \frac{1}{2} m_i v_i^2,$$

where \mathbf{v}_i is the velocity of the i th particle relative to the fixed point as measured in the space axes. By Eq. (5.2), T may also be written as

$$T = \frac{1}{2} m_i \mathbf{v}_i \cdot (\boldsymbol{\omega} \times \mathbf{r}_i),$$

which, upon permuting the vectors in the triple dot product, becomes

$$T = \frac{\boldsymbol{\omega}}{2} \cdot m_i (\mathbf{r}_i \times \mathbf{v}_i).$$

The quantity summed over i will be recognized as the angular momentum of the body about the origin, and in consequence the kinetic energy can be written in the form

$$T = \frac{\boldsymbol{\omega} \cdot \mathbf{L}}{2} = \frac{\boldsymbol{\omega} \cdot \mathbf{I} \cdot \boldsymbol{\omega}}{2}. \quad (5.16)$$

Let \mathbf{n} be a unit vector in the direction of $\boldsymbol{\omega}$ so that $\boldsymbol{\omega} = \omega \mathbf{n}$. Then an alternative form for the kinetic energy is

$$T = \frac{\omega^2}{2} \mathbf{n} \cdot \mathbf{l} \cdot \mathbf{n} = \frac{1}{2} I \omega^2, \quad (5.17)$$

where I is a scalar, defined by

$$I = \mathbf{n} \cdot \mathbf{l} \cdot \mathbf{n} = m_i [r_i^2 - (\mathbf{r}_i \cdot \mathbf{n})^2], \quad (5.18)$$

and known as the *moment of inertia about the axis of rotation*.

In the usual elementary discussions, the moment of inertia about an axis is defined as the sum, over the particles of the body, of the product of the particle mass and the square of the perpendicular distance from the axis. It must be shown that this definition is in accord with the expression given in Eq. (5.18). The perpendicular distance is equal to the magnitude of the vector $\mathbf{r}_i \times \mathbf{n}$ (cf. Fig. 5.2). Therefore, the customary definition of I may be written as

$$I = m_i (\mathbf{r}_i \times \mathbf{n}) \cdot (\mathbf{r}_i \times \mathbf{n}). \quad (5.19)$$

Multiplying and dividing by ω^2 , this definition of I may also be written as

$$I = \frac{m_i}{\omega^2} (\boldsymbol{\omega} \times \mathbf{r}_i) \cdot (\boldsymbol{\omega} \times \mathbf{r}_i).$$

But each vector in the dot product is exactly the relative velocity \mathbf{v}_i as measured in the space system of axes. Hence, I so defined is related to the kinetic energy by

$$I = \frac{2T}{\omega^2},$$

which is the same as Eq. (5.17), and therefore I must be identical with the scalar defined by Eq. (5.19).

The value of the moment of inertia depends upon the direction of the axis of rotation. As $\boldsymbol{\omega}$ usually changes its direction with respect to the body in the course

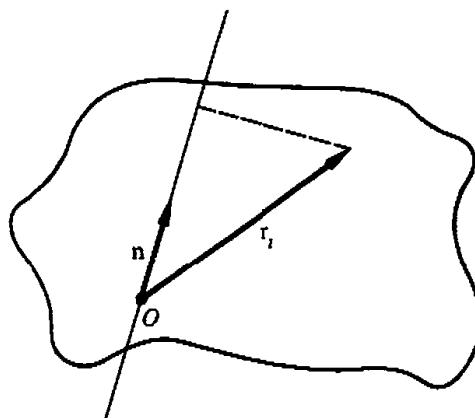


FIGURE 5.2 The definition of the moment of inertia.

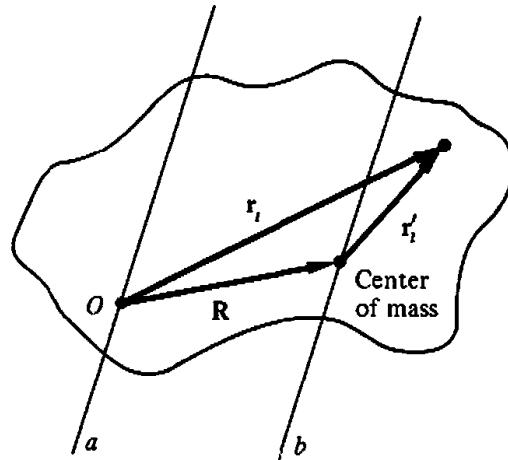


FIGURE 5.3 The vectors involved in the relation between moments of inertia about parallel axes.

of time, the moment of inertia must also be considered a function of time. When the body is constrained so as to rotate only about a fixed axis, then the moment of inertia is a constant. In such a case, the kinetic energy (5.16) is almost in the form required to fashion the Lagrangian and the equations of motion. The one further step needed is to express ω as the time derivative of some angle, which can usually be done without difficulty.

Along with the inertia tensor, the moment of inertia also depends upon the choice of origin of the body set of axes. However, the moment of inertia about some given axis is related simply to the moment about a parallel axis through the center of mass. Let the vector from the given origin O to the center of mass be \mathbf{R} , and let the radii vectors from O and the center of mass to the i th particle be \mathbf{r}_i and \mathbf{r}'_i , respectively. The three vectors so defined are connected by the relation (cf. Fig. 5.3)

$$\mathbf{r}_i = \mathbf{R} + \mathbf{r}'_i. \quad (5.20)$$

The moment of inertia about the axis a is therefore

$$I_a = m_i (\mathbf{r}_i \times \mathbf{n})^2 = m_i [(\mathbf{r}'_i + \mathbf{R}) \times \mathbf{n}]^2$$

or

$$I_a = M (\mathbf{R} \times \mathbf{n})^2 + m_i (\mathbf{r}'_i \times \mathbf{n})^2 + 2m_i (\mathbf{R} \times \mathbf{n}) \cdot (\mathbf{r}'_i \times \mathbf{n}),$$

where M is the total mass of the body. The last term in this expression can be rearranged as

$$-2(\mathbf{R} \times \mathbf{n}) \cdot (\mathbf{n} \times m_i \mathbf{r}'_i).$$

Chapter 5 The Rigid Body Equations of Motion

By the definition of center of mass, the summation $m_i \mathbf{r}'_i$ vanishes. Hence, I_a can be expressed in terms of the moment about the parallel axis b as

$$\begin{aligned} I_a &= I_b + M(\mathbf{R} \times \mathbf{n})^2 \\ &= I_b + MR^2 \sin^2 \theta. \end{aligned} \quad (5.21)$$

The magnitude of $\mathbf{R} \times \mathbf{n}$, which has the value $R \sin \theta$, where θ is the angle between \mathbf{R} and \mathbf{n} , is the perpendicular distance of the center of mass from the axis passing through O . Consequently, the moment of inertia about a given axis is equal to the moment of inertia about a parallel axis through the center of mass plus the moment of inertia of the body, as if concentrated at the center of mass, with respect to the original axis.

The inertia tensor is defined in general from the kinetic energy of rotation about an axis, and is written as

$$T_{\text{rotation}} = \frac{1}{2} m_i (\boldsymbol{\omega} \times \mathbf{r}_i)^2 = \frac{1}{2} \omega_\alpha \omega_\beta m_i (\delta_{\alpha\beta} r_i^2 - r_{i\alpha} r_{i\beta}),$$

where Greek letters indicate the components of $\boldsymbol{\omega}$ and \mathbf{r}_i . In an inertial frame, the sum is over the particles in the body, and $r_{i\alpha}$ is the α th component of the position of the i th particle. Because T_{rotation} is a bilinear form in the components of $\boldsymbol{\omega}$, it can be written as

$$T_{\text{rotation}} = \frac{1}{2} I_{\alpha\beta} \omega_\alpha \omega_\beta,$$

where

$$I_{\alpha\beta} = m_i (\delta_{\alpha\beta} r_i^2 - r_{i\alpha} r_{i\beta}) \quad (5.22)$$

is the moment of inertia tensor. To get the moment of inertia about an axis through the center of mass, choose the rotation about this axis. For a body with a continuous distribution of density $\rho(r)$, the sums in the components of the moment of inertia tensor in Eq. (5.22) reduce to

$$I_{\alpha\beta} = \int_V \rho(r) (\delta_{\alpha\beta} r^2 - r_\alpha r_\beta) dV. \quad (5.23)$$

As an example, let us consider a homogeneous cube of density ρ , mass M , and side a . Choose the origin to be at one corner and the three edges adjacent to that corner to lie on the $+x$, $+y$, and $+z$ axes. If we define $b = Ma^2$, then straightforward integration of Eq. (5.23) gives

$$\mathbf{I} = \begin{pmatrix} \frac{2}{3}b & -\frac{1}{4}b & -\frac{1}{4}b \\ -\frac{1}{4}b & \frac{2}{3}b & -\frac{1}{4}b \\ -\frac{1}{4}b & -\frac{1}{4}b & \frac{2}{3}b \end{pmatrix}.$$

Thus, both the moment of inertia and the inertia tensor possess a type of revolution, relative to the center of mass, very similar to that found for the linear and angular momentum and the kinetic energy in Section (1.2).

5.4 ■ THE EIGENVALUES OF THE INERTIA TENSOR AND THE PRINCIPAL AXIS TRANSFORMATION

The preceding discussion emphasizes the important role the inertia tensor plays in the discussion of the motion of rigid bodies. An examination, at this point, of the properties of this tensor and its associated matrix will therefore prove of considerable interest. From the defining equation, (5.7), it is seen that the components of the tensor are symmetrical; that is

$$I_{xy} = I_{yx}. \quad (5.24)$$

This means that, while the inertia tensor will in general have nine components, only six of them will be independent—the three along the diagonal plus three of the off-diagonal elements.

The inertia coefficients depend both upon the location of the origin of the body set of axes and upon the orientation of these axes with respect to the body. This symmetry suggests that there exists a set of coordinates in which the tensor is diagonal with the three principal values I_1 , I_2 , and I_3 . In this system, the components of \mathbf{L} would involve only the corresponding component of $\boldsymbol{\omega}$, thus*

$$L_1 = I_1\omega_1, \quad L_2 = I_2\omega_2, \quad L_3 = I_3\omega_3. \quad (5.25)$$

A similar simplification would also occur in the form of the kinetic energy:

$$T = \frac{\boldsymbol{\omega} \cdot \mathbf{I} \cdot \boldsymbol{\omega}}{2} = \frac{1}{2}I_1\omega_1^2 + \frac{1}{2}I_2\omega_2^2 + \frac{1}{2}I_3\omega_3^2. \quad (5.26)$$

We can show that it is always possible to find such axes, and the proof is based essentially on the symmetric nature of the inertia tensor.

There are several ways to understand vectors and tensors. For example, a vector is a quantity defined by its transformation properties. In any set of coordinates, a vector is specified by its three components, e.g.,

$$\mathbf{V} = V_x\mathbf{i} + V_y\mathbf{j} + V_z\mathbf{k}, \quad (5.27)$$

or by its magnitude and direction. In any frame, the magnitude is given by $\sqrt{V_x^2 + V_y^2 + V_z^2}$, and the direction is given by the polar angles θ and ϕ . An alternative is to use the first two Euler angles to specify a new z axis chosen such that the vector's direction is along that axis. Since the vector lies along that z axis, the third Euler angle is not needed.

An approach similar to this latter method can be used for the symmetric moment of inertia tensor. Consider the moment of inertia of a body about an axis passing through the center of mass of the body. A similarity transformation per-

*With an eye to future applications, components relative to these axes will be denoted by subscripts 1, 2, 3.

formed by a rotation matrix \mathbf{R} can be chosen such that

$$\mathbf{I}_D = \mathbf{R}\tilde{\mathbf{I}}\mathbf{R}. \quad (5.28)$$

This rotation can be expressed in terms of the Euler angles ϕ , θ , and ψ as shown in Eqs. (4.46) and (4.47). A proper choice of these angles will transform \mathbf{I} into its diagonal form

$$\mathbf{I}_D = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix} \quad (5.29)$$

where I_1 , I_2 , and I_3 , which are the eigenvalues of \mathbf{I} , are referred to as the components of the *principal moment of inertia* tensor. The directions of x' , y' , and z' defined by the rotation matrix in Eq. (5.28) are called the *principal axes*, or *eigenvectors* of the inertia tensor. These eigenvectors lie along the directions x' , y' , and z' .

Once the principal moments and their directions relative to the surface of a body are known, the inertia tensor relative to any other set of axis through the center of mass can be found by a similarity transformation defined by the Euler angles relating the two coordinate systems. If \mathbf{S} is that transformation, then

$$\mathbf{I} = \mathbf{S}\mathbf{I}_D\tilde{\mathbf{S}}, \quad (5.30)$$

gives the moment of inertia in that frame. Equation (5.21) can then be used to transform the rotation center to any desired location. The principal values of \mathbf{I} can be determined by the methods of matrix algebra.

The three principal values of the moment of inertia tensor in Eq. (5.29) can be found by solving the cubic equation for \mathbf{I} that arises from the determinant

$$\begin{vmatrix} I_{xx} - I & I_{xy} & I_{zx} \\ I_{xy} & I_{yy} - I & I_{yz} \\ I_{zx} & I_{yz} & I_{zz} - I \end{vmatrix} = 0, \quad (5.31)$$

where the symmetry of \mathbf{I} has been displayed explicitly. Equation (5.31) is the secular equation, whose three roots are the desired principal moments. For each of these roots, Eqs. (5.28) can be solved to obtain the direction of the corresponding principal axis. In most of the easily soluble problems in rigid dynamics, the principal axes can be determined by inspection. For example, we often have to deal with rigid bodies that are solids of revolution about some axis, with the origin of the body system on the symmetry axis. All directions perpendicular to the axis of symmetry are then alike, which is the mark of a double root to the secular equation. The principal axes are then the symmetry axis and any two perpendicular axes in the plane normal to the symmetry axis.

The principal moments of inertia cannot be negative, because as the diagonal elements in the principal axes system they have the form of sums of squares. Thus,

I_{xx} is given by (cf. Eq. (5.6))

$$I_{xx} = m_i(y_i^2 + z_i^2).$$

For one of the principal moments to vanish, all points of the body must be such that two coordinates of each particle are zero. Clearly this can happen only if all points of the body are collinear with the principal axis corresponding to the zero principal moment. Any two axes perpendicular to the line of the body will then be the other principal axes. Indeed, this is clearly a limiting case of a body with an axis of symmetry passing through the origin.

We can also understand the concept of principal axes through some geometrical considerations that historically formed the first approach to the subject. The moment of inertia about a given axis has been defined as $I = \mathbf{n} \cdot \mathbf{I} \cdot \mathbf{n}$. Let the direction cosines of the axis be α , β , and γ so that

$$\mathbf{n} = \alpha \mathbf{i} + \beta \mathbf{j} + \gamma \mathbf{k};$$

I then can be written as

$$I = I_{xx}\alpha^2 + I_{yy}\beta^2 + I_{zz}\gamma^2 + 2I_{xy}\alpha\beta + 2I_{yz}\beta\gamma + 2I_{zx}\gamma\alpha, \quad (5.32)$$

using the symmetry of \mathbf{I} explicitly. It is convenient to define a vector ρ by the equation

$$\rho = \frac{\mathbf{n}}{\sqrt{I}}. \quad (5.33)$$

The magnitude of ρ is thus related to the moment of inertia about the axis whose direction is given by \mathbf{n} . In terms of the components of this new vector, Eq. (5.32) takes on the form

$$1 = I_{xx}\rho_1^2 + I_{yy}\rho_2^2 + I_{zz}\rho_3^2 + 2I_{xy}\rho_1\rho_2 + 2I_{yz}\rho_2\rho_3 - 2I_{zx}\rho_3\rho_1. \quad (5.34)$$

Considered as a function of the three variables ρ_1 , ρ_2 , ρ_3 , Eq. (5.34) is the equation of some surface in ρ space. In particular, Eq. (5.34) is the equation of an ellipsoid designated as the *inertial ellipsoid*. We can always transform to a set of Cartesian axes in which the equation of an ellipsoid takes on its normal form:

$$1 = I_1\rho'_1^2 + I_2\rho'_2^2 + I_3\rho'_3^2, \quad (5.35)$$

with the principal axes of the ellipsoid along the new coordinate axes. But (5.35) is simply the form Eq. (5.34) has in a system of coordinates in which the inertia tensor \mathbf{I} is diagonal. Hence, the coordinate transformation that puts the equation of ellipsoid into its normal form is exactly the principal axis transformation previously discussed. The principal moments of inertia determine the lengths of the axes of the inertia ellipsoid. If two of the roots of the secular equation are equal, the inertia ellipsoid thus has two equal axes and is an ellipsoid of revolution. If all three principal moments are equal, the inertia ellipsoid is a sphere.

A quantity closely related to the moment of inertia is the *radius of gyration*, R_0 , defined by the equation

$$I = MR_0^2. \quad (5.36)$$

In terms of the radius of gyration, the vector ρ can be written as

$$\rho = \frac{\mathbf{n}}{R_0\sqrt{M}}.$$

The radius vector to a point on the inertia ellipsoid is thus inversely proportional to the radius of gyration about the direction of the vector.

It is worth reemphasizing that the inertia tensor \mathbf{I} and all the quantities associated with it—principal axes, principal moments, inertia ellipsoid, etc.—are only relative to some particular point fixed in the body. If the point is shifted elsewhere in the body, all the quantities will in general be changed. Thus, Eq. (5.21) gives the effect of moving the reference point from the center of mass to some other point. The principal axis transformation that diagonalizes \mathbf{I}' at the center of mass will not necessarily diagonalize \mathbf{I} about another axis, and hence is not in general the principal axis transformation for the shifted tensor \mathbf{I} . Only if the shift vector \mathbf{R} is along one of the principal axes relative to the center of mass will the difference tensor be diagonal in that system. The new inertia tensor \mathbf{I} will in that special case have the same principal axes as at the center of mass. However, the principal moments of inertia are changed, except for that corresponding to the shift axis, where the diagonal element of the difference tensor is clearly zero. The “parallel axis” theorem for the diagonalized form of the inertia tensor thus has a rather specialized and restricted form.

5.5 ■ SOLVING RIGID BODY PROBLEMS AND THE EULER EQUATIONS OF MOTION

Practically all the tools necessary for setting up and solving problems in rigid body dynamics have by now been assembled. If nonholonomic constraints are present, then special means must be taken to include the effects of these constraints in the equations of motion. For example, if there are “rolling constraints,” these must be introduced into the equations of motion by the method of Lagrange undetermined multipliers, as in Section 2.4. As discussed in Section 5.1, we usually seek a particular reference point in the body such that the problem can be split into two separate parts, one purely translational and the other purely rotational about the reference point. Of course, if one point of the rigid body is fixed in an inertial system, then that is the obvious reference point. All that has to be considered then is the rotational problem about the fixed point.

For bodies without a fixed point, the most useful reference point is almost always the center of mass. We have already seen that the total kinetic energy and angular momentum then split neatly into one term relating to the translational

motion of the center of mass and another involving rotation *about* the center of mass. Thus, Eq. (1.31) can now be written

$$T = \frac{1}{2}Mv^2 + \frac{1}{2}I\omega^2.$$

For many problems (certainly all those that will be considered here), a similar sort of division can be made for the potential energy. We can then solve individually for the translational motion of the center of mass and for the rotational motion about the center of mass. For example, the Newtonian equations of motion can be used directly: Eq. (1.22) for the motion of the center of mass and Eq. (1.26) for the motion about that point.

With holonomic conservative systems, the Lagrangian formulation is available, with the Lagrangian taking the form

$$L(q, \dot{q}) = L_c(q_c, \dot{q}_c) + L_b(q_b, \dot{q}_b).$$

Here L_c is that part of the Lagrangian involving the generalized coordinates q_c (and velocities \dot{q}_c) of the center of mass, and L_b the part relating to the orientation of the body about the center of mass, as described by q_b, \dot{q}_b . In effect then, there are two distinct problems, one with Lagrangian L_c and the other with Lagrangian L_b .

In both the Newtonian and Lagrangian formulations, it is convenient to work in terms of the principal axes system of the point of reference, so that the kinetic energy of rotation takes the simple form given in Eq. (5.26). So far, the only suitable generalized coordinates we have for the rotational motion of the rigid body are the Euler angles. Of course, the motion is often effectively confined to two dimensions, as in the motion of a rigid lamina in a plane. The axis of rotation is then fixed in the direction perpendicular to the plane; only one angle of rotation is necessary and we may dispense with the cumbersome machinery of the Euler angles.

For the rotational motion about a fixed point or the center of mass, the direct Newtonian approach leads to a set of equations known as Euler's equations of motion. We consider either an inertial frame whose origin is at the fixed point of the rigid body, or a system of space axes with origin at the center of mass. In these two situations, Eq. (1.26) holds, which here appears simply as

$$\left(\frac{d\mathbf{L}}{dt} \right)_s = \mathbf{N}.$$

The subscript s is used because the time derivative is with respect to axes that do not share the rotation of the body. However, Eq. (4.86) can be used to obtain the derivatives with respect to axes fixed in the body:

$$\left(\frac{d\mathbf{L}}{dt} \right)_s = \left(\frac{d\mathbf{L}}{dt} \right)_b + \boldsymbol{\omega} \times \mathbf{L}.$$

or, by dropping the “body” subscript:

$$\frac{d\mathbf{L}}{dt} + \boldsymbol{\omega} \times \mathbf{L} = \mathbf{N}. \quad (5.37)$$

Equation (5.37) is thus the appropriate form of the Newtonian equation of motion relative to body axes. The i th component of Eq. (5.37) can be written

$$\frac{dL_i}{dt} + \epsilon_{ijk}\omega_j L_k = N_i. \quad (5.38)$$

If now the body axes are taken as the principal axes relative to the reference point, then the angular momentum components are $L_i = I_i\omega_i$. By Eq. (5.25), Eq. (5.38) takes the form (no summation on i *)

$$I_i \frac{d\omega_i}{dt} + \epsilon_{ijk}\omega_j \omega_k I_k = N_i, \quad (5.39)$$

since the principal moments of inertia are of course time independent. In expanded form, the three equations making up Eq. (5.39) look like

$$\begin{aligned} I_1 \dot{\omega}_1 - \omega_2 \omega_3 (I_2 - I_3) &= N_1 \\ I_2 \dot{\omega}_2 - \omega_3 \omega_1 (I_3 - I_1) &= N_2 \\ I_3 \dot{\omega}_3 - \omega_1 \omega_2 (I_1 - I_2) &= N_3. \end{aligned} \quad (5.39')$$

Equations (5.39) or (5.39') are Euler's equations of motion for a rigid body with one point fixed. They can also be derived from Lagrange's equations in the form of Eq. (1.53) where the generalized forces Q_j are the torques, N_j , corresponding to the Euler angles of rotation. However, only one of the Euler angles has its associated torque along one of the body axes, and the remaining two Euler's equations must be obtained by cyclic permutation (cf. Derivation 4).

Consider the case where $I_1 = I_2 \neq I_3$. A torque with components N_1 or N_2 will cause both ω_1 and ω_2 to change without affecting ω_3 . We shall return to a discussion of this in Section 5.7 when we consider the heavy symmetric top with one point fixed. Let us first consider the torque-free motion of a rigid body.

5.6 ■ TORQUE-FREE MOTION OF A RIGID BODY

One problem in rigid dynamics where Euler's equations are applicable is in the motion of a rigid body not subject to any net forces or torques. The center of mass is then either at rest or moving uniformly, and it does not decrease the generality of the solution to discuss the rotational motion in a reference frame in which the center of mass is stationary. In such a case, the angular momentum arises only from rotation about the center of mass, and Euler's equations are the equations of

*It should be obvious that Eq. (5.39), as the i th component of a vector equation, does not involve a summation over i , although summation is implied over the repeated indices j and k .

motion for the complete system. In the absence of any net torques, they reduce to

$$\begin{aligned} I_1\dot{\omega}_1 &= \omega_2\omega_3(I_2 - I_3) \\ I_2\dot{\omega}_2 &= \omega_3\omega_1(I_3 - I_1) \\ I_3\dot{\omega}_3 &= \omega_1\omega_2(I_1 - I_2). \end{aligned} \quad (5.40)$$

The same equations, of course, will also describe the motion of a rigid body when one point is fixed and there are no net applied torques. We know two immediate integrals of the motion, for both the kinetic energy and the total angular momentum vector must be constant in time. With these two integrals it is possible to integrate (5.40) completely in terms of elliptic functions, but such a treatment is not very illuminating. However, it is also possible to derive an elegant geometrical description of the motion, known as Poinsot's construction, without requiring a complete solution to the problem.

Let us consider a coordinate system oriented along the principal axes of the body but whose axes measure the components of a vector ρ along the instantaneous axis of rotation as defined by Eq. (5.33). For our purposes, it is convenient to make use of Eq. (5.17) for the kinetic energy (here constant) and write the definition of ρ in the form

$$\rho = \frac{\omega}{\omega\sqrt{I}} = \frac{\omega}{\sqrt{2T}}. \quad (5.41)$$

In this ρ space, we define a function

$$F(\rho) = \rho \cdot \mathbf{l} \cdot \rho = \rho^2 I_t, \quad (5.42)$$

where the surfaces of constant F are ellipsoids, the particular surface $F = 1$ being the inertia ellipsoid. As the direction of the axis of rotation changes in time, the parallel vector ρ moves accordingly, its tip always defining a point on the inertia ellipsoid. The gradient of F , evaluated at this point, furnishes the direction of the corresponding normal to the inertia ellipsoid. From Eq. (5.42) for $F(\rho)$, the gradient of F with respect to ρ has the form

$$\nabla_{\rho} F = 2\mathbf{l} \cdot \rho = \frac{2\mathbf{l} \cdot \omega}{\sqrt{2T}},$$

or

$$\nabla_{\rho} F = \sqrt{\frac{2}{T}} \mathbf{L}. \quad (5.43)$$

Thus, the ω vector will always move such that the corresponding normal to the inertia ellipsoid is in the direction of the angular momentum. In the particular case under discussion, the direction of \mathbf{L} is fixed in space, and it is the inertia ellipsoid (fixed with respect to the body) that must move in space in order to preserve this connection between ω and \mathbf{L} (cf. Fig. 5.4).

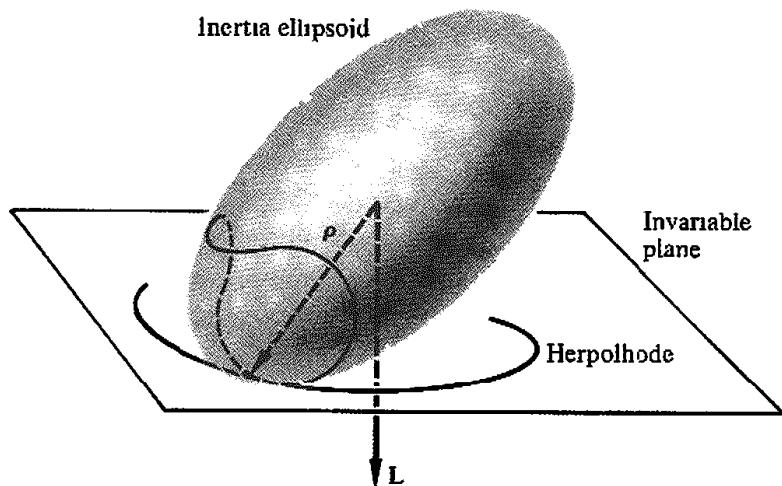


FIGURE 5.4 The motion of the inertia ellipsoid relative to the invariable plane.

It can also be shown that the distance between the origin of the ellipsoid and the plane tangent to it at the point ρ must similarly be constant in time. This distance is equal to the projection of ρ on \mathbf{L} and is given by

$$\frac{\rho \cdot \mathbf{L}}{L} = \frac{\omega \cdot \mathbf{L}}{L\sqrt{2T}}$$

or

$$\frac{\rho \cdot \mathbf{L}}{L} = \frac{\sqrt{2T}}{L}, \quad (5.44)$$

where use has been made of Eq. (5.16). Both T , the kinetic energy, and \mathbf{L} , the angular momentum, are constants of the motion, and the tangent plane is therefore always a fixed distance from the origin of the ellipsoid. Since the normal to the plane, being along \mathbf{L} , also has a fixed direction, the tangent plane is known as the *invariable plane*. We can picture the force-free motion of the rigid body as being such that the inertia ellipsoid rolls, without slipping, on the invariable plane, with the center of the ellipsoid a constant height above the plane. The rolling occurs without slipping because the point of contact is defined by the position of ρ , which, being along the instantaneous axis of rotation, is the one direction in the body momentarily at rest. The curve traced out by the point of contact on the inertia ellipsoid is known as the *polhode*, while the similar curve on the invariable plane is called the *herpolhode*.*

Poinsot's geometrical discussion is quite adequate to describe completely the force-free motion of the body. The direction of the invariable plane and the height of the inertia ellipsoid above it are determined by the values of T and \mathbf{L} , which are among the initial conditions of the problem. It is then a matter of geometry to

*Hence, the Jabberwockian-sounding statement: the polhode rolls without slipping on the herpolhode lying in the invariable plane.

trace out the polhode and the herpolhode.* The direction of the angular velocity in space is given by the direction of ρ , while the instantaneous orientation of the body is provided by the orientation of the inertia ellipsoid, which is fixed in the body. Many elaborate descriptions of force-free motion obtained in this fashion can be found in the literature.

In the special case of a symmetrical body, the inertia ellipsoid is an ellipsoid of revolution, so that the polhode on the ellipsoid is clearly a circle about the symmetry axis. The herpolhode on the invariable plane is likewise a circle. An observer fixed in the body sees the angular velocity vector ω move on the surface of a cone—called the *body cone*—whose intersection with the inertia ellipsoid is the polhode. Correspondingly, an observer fixed in the space axes sees ω move on the surface of a *space cone* whose intersection with the invariable plane is the herpolhode. Thus, the free motion of the symmetrical rigid body is sometimes described as the rolling of the body cone on the space cone. If the moment of inertia about the symmetry axis is less than that about the other two principal axes, then from Eq. (5.35) the inertia ellipsoid is prolate, i.e., football shaped—somewhat as is shown in Fig. 5.4. In that case, the body cone is outside the space cone. When the moment of inertia about the symmetry axis is the greater, the ellipsoid is oblate and the body cone rolls around the inside of the space cone. In either case, the physical description of the motion is that the direction of ω precesses in time about the axis of symmetry of the body.

The Poinsot construction shows how ω moves, but gives no information as to how the L vector appears to move in the body system of axes. Another geometrical description is available however to describe the path of the L vector as seen by an observer in the principal axes system. Equations (5.25) and (5.26) imply that in this system the kinetic energy is related to the components of the angular momentum by the equation

$$T = \frac{L_x^2}{2I_1} + \frac{L_y^2}{2I_2} + \frac{L_z^2}{2I_3}. \quad (5.45)$$

Since T is constant, this relation defines an ellipsoid, referred to as the *Binet ellipsoid*, also fixed in the body axes but *not* the same as the inertia ellipsoid.

If we adopt the convention

$$I_3 \leq I_2 \leq I_1,$$

and write the equations for the ellipsoid in the standard form

$$\frac{L_x^2}{2TI_1} + \frac{L_y^2}{2TI_2} + \frac{L_z^2}{2TI_3} = 1 \quad (5.45')$$

then we see that the ellipsoid sketched on Fig. 5.5a has semimajor axes, in order of decreasing size, of $\sqrt{2TI_1}$, $\sqrt{2TI_2}$, and $\sqrt{2TI_3}$. The conservation of the total

*The herpolhode is always concave to the origin, belying its name, which means “snakelike.”

angular momentum, \mathbf{L} , gives us

$$\frac{L_x^2 + L_y^2 + L_z^2}{L^2} = 1, \quad (5.46)$$

the equation for a sphere in $L_x L_y L_z$ space. The vector \mathbf{L} moves in such a way that it describes a path on both the ellipsoid of Eq. (5.45) and the sphere of Eq. (5.46). In other words, the path of \mathbf{L} is the intersection of the ellipsoid and the sphere. The components \mathbf{L} satisfy the equation

$$\frac{L_x^2}{2T I_1} + \frac{L_y^2}{2T I_2} + \frac{L_z^2}{2T I_3} = \frac{L_x^2 + L_y^2 + L_z^2}{L^2}.$$

It is easy to show that these two surfaces will intersect for values of L larger than the ellipsoid semiminor axis and less than the semimajor axis, that is,

$$\sqrt{2T I_3} < L < \sqrt{2T I_1}.$$

The sphere is outside the ellipsoid on the L_z axis and inside the ellipsoid along L_x . Figure 5.5 depicts curves where the sphere intersects the ellipsoid for various values of L . Fig. 5.5a shows a perspective view and Fig. 5.5b shows the view as seen from the L_y axis. The curves that appear as straight lines on Fig 5.5b correspond to the case where $L = \sqrt{2T I_2}$.

With the help of this geometrical construction, something can be said about the possible motions of a free asymmetric body. It is easy to see that a steady rotation

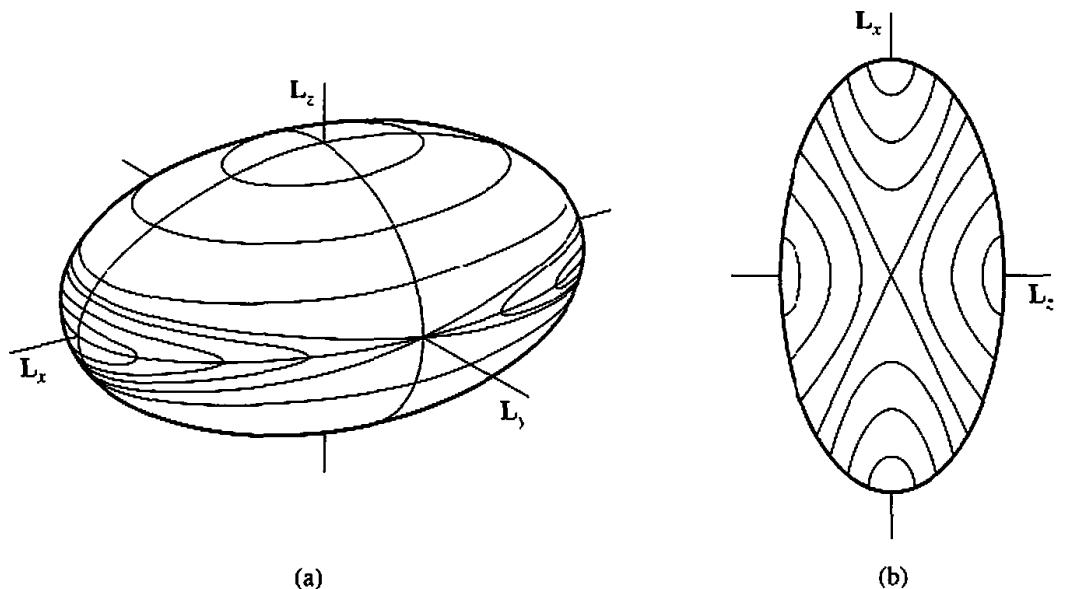


FIGURE 5.5 (a) The kinetic energy, or Binet, ellipsoid fixed in the body axes, and some possible paths of the \mathbf{L} vector in its surface. (b) Side view of Binet ellipsoid.

of such a body is possible only about one of the principal axes. From the Euler equations (5.40), all the components of ω can be constant only if

$$\omega_1\omega_2(I_1 - I_2) = \omega_2\omega_3(I_2 - I_3) = \omega_3\omega_1(I_3 - I_1) = 0,$$

which requires that at least two of the components ω_i be zero; i.e., ω is along only one of the principal axes. However, not all of these possible motions are stable—that is, not moving far from the principal axis under small perturbation. For example, steady motion about the L_z axis will occur when $L^2 = 2TI_3$. When there are slight deviations from this condition, the radius of the angular momentum sphere is just slightly smaller than this value, and the intersection with the kinetic energy ellipsoid is a small circle about the L_z axis. The motion is thus stable, the L vector never being far from the axis.

Similarly, at the other extreme, when the motion about the axis of smallest I is perturbed, the radius of the angular momentum sphere is just slightly larger than the smallest semimajor axis. The intersection is again a small closed figure around the principal axis, and the motion is stable. However, the motion about the intermediate axis is *unstable*. This is clearly shown in Fig. 5.5. For the intermediate (L_y) axis, the kinetic energy has two orbits that encircle the ellipsoid and cross each other where the $\pm L_y$ pass through the ellipsoid. Hence, there are two different orbits with values slightly less than $\sqrt{2TI_2}$ and two other distinctly different orbits with values slightly exceeding $\sqrt{2TI_2}$, all four of which have quite long paths on the surface.

This behavior can be best understood by recognizing that at the intermediate axis the radius of curvature of the ellipsoid in one direction is greater than that of the contact sphere, and less in the perpendicular direction. At the other two extremes, the radii of curvature are either greater or smaller than the sphere radius in all directions. These conclusions on the stability of free-body motion have been known for a long time, but applications, e.g., to the stability of spinning spacecraft, have brought them out of the obscurity of old monographs on rigid body dynamics.*

For a symmetrical rigid body, the analytical solution for the force-free motion is not difficult to obtain, and we can directly confirm the precessing motion predicted by the Poinsot construction. Let the symmetry axis be taken as the L_z principal axis so that $I_1 = I_2$. Euler's equations (5.40) reduce then to

*If there are dissipative mechanisms present, these stability arguments have to be modified. It is easy to see that for a body with constant L , but slowly decreasing T , the only stable rotation is about the principal axis with the largest moment of inertia. The kinetic energy of rotation about the i th principal axis for given L is $T = L^2/2I_i$, which is least for the axis with the largest I_i . If a body is set spinning about any other principal axis, the effect of a slowly decreasing kinetic energy is to cause the angular velocity vector to shift until the spinning is about the axis requiring the least value of T for the given L . Such dissipative effects are present in spacecraft because of the flexing of various members in the course of the motion, especially of the long booms carried by many of them. These facts were learned the hard way by the early designers of spacecraft!

$$\begin{aligned} I_1 \dot{\omega}_1 &= (I_1 - I_3) \omega_3 \omega_2 \\ I_1 \dot{\omega}_2 &= (I_1 - I_3) \omega_3 \omega_1 \\ I_3 \dot{\omega}_3 &= 0. \end{aligned} \quad (5.47)$$

The last of these equations states that ω_3 is a constant, and it can therefore be treated as one of the known initial conditions of the problem. The remaining two equations can now be written

$$\dot{\omega}_1 = -\Omega \omega_2, \quad \dot{\omega}_2 = \Omega \omega_1, \quad (5.48)$$

where Ω is an angular frequency

$$\Omega = \frac{I_3 - I_1}{I_1} \omega_3. \quad (5.49)$$

Elimination of ω_2 between Eqs. (5.48) leads to the standard differential equation for simple harmonic motion

$$\ddot{\omega}_1 = -\Omega^2 \omega_1,$$

with the typical solution

$$\omega_1 = A \cos \Omega t.$$

The corresponding solution for ω_2 can be found by substituting this expression for ω_1 , back in the first of Eqs. (5.48):

$$\omega_2 = A \sin \Omega t.$$

The solutions for ω_1 and ω_2 show that the vector $\omega_1 \mathbf{i} + \omega_2 \mathbf{j}$ has a constant magnitude and rotates uniformly about the z axis of the body with the angular frequency Ω (cf. Fig. 5.6). Hence, the total angular velocity $\boldsymbol{\omega}$ is also constant in magnitude and *precesses* about the z axis with the same frequency, exactly as predicted by the Poinsot construction. * Recall that the precession described here is relative to the body axes, which are themselves rotating in space with the larger frequency $\boldsymbol{\omega}$. From Eq. (5.49), it is seen that the closer I_1 is to I_3 , the slower will be the precession frequency Ω compared to the rotation frequency ω . The constants A (the amplitude of the precession) and ω_3 can be evaluated in terms of the more usual constants of the motion, namely, the kinetic energy and the magnitude of the angular momentum. Both T and L^2 can be written as functions of A and ω_3 :

*The precession can be demonstrated in another fashion by defining a vector $\boldsymbol{\Omega}$ lying along the z axis with magnitude given by (5.49). Equations (5.47) are then essentially equivalent to the vector equation

$$\dot{\boldsymbol{\omega}} = \boldsymbol{\omega} \times \boldsymbol{\Omega},$$

which immediately reveals the precession of $\boldsymbol{\omega}$ with the frequency Ω .

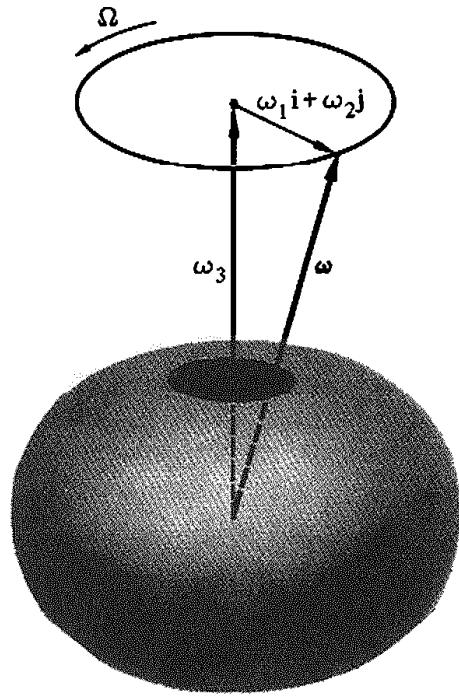


FIGURE 5.6 Precession of the angular velocity about the axis of symmetry in the force-free motion of a symmetrical rigid body.

$$T = \frac{1}{2} I_1 A^2 + \frac{1}{2} I_3 \omega_3^2,$$

$$L^2 = I_1^2 A^2 + I_3^2 \omega_3^2,$$

and these relations in turn may be solved for A and ω_3 in terms of T and L .

We would expect that Earth's axis of rotation should exhibit this precession, for the external torques acting on Earth are so weak that the *rotational* motion may be considered as that of a free body. Earth is approximately symmetrical about the polar axis and slightly flattened at the poles so that I_1 is less than I_3 . Numerically, the ratio of the moments is such that

$$\frac{I_3 - I_1}{I_1} = 0.00327,$$

and the magnitude of the precession angular frequency should therefore be

$$\Omega = \frac{\omega_3}{305.81039} \approx \frac{\omega_3}{306}.$$

Since ω_3 is practically the same as the magnitude of ω , this result predicts a period of precession of approximately 306 days or about 10 months. If some circumstance disturbed the axis of rotation from the figure axis of Earth, we would therefore expect the axis of rotation to precess around the figure axis (i.e., around the north pole) once every 10 months. Practically, such a motion should show up

as a periodic change in the apparent latitude of points on Earth's surface. Careful measurements of latitude at a network of locations around the world, carried out now for about a century, show that the rotation axis is indeed moving about the pole with an amplitude of the order of a few tenths of a second of latitude (about 10 m). But the situation is far more complicated (and interesting) than the above simple analysis would suggest.

The deviations between the figure and rotation axes are very irregular so that it's more a "wobble" than a precession. Careful frequency analysis shows the existence of an annual period in the motion, thought to arise from the annual cycle of seasons and the corresponding mean displacement of atmospheric masses about the globe. Additionally, a strong frequency component is centered about a period of 420 days, known as the *Chandler wobble*. The present belief is that this motion represents the free-body precession derived above. It is thought that the difference in period arises from the fact that Earth is not a rigid body but is to some degree elastic. In effect, some part of Earth follows along with the shift in the rotation axis, which has the effect of reducing the difference in the principal moments of inertia and therefore increasing the period. (If, for example, Earth were completely fluid, then the figure axis would instantaneously adjust to the rotation axis and there could be no precession.)

There are still other obscure features to the observed wobble. The frequency analysis indicates strong damping effects are present, believed to arise from either tidal friction or dissipative effects in the coupling between the mantle and the core. The damping period ought to be on the order of 10–20 years. But no such decay of the amplitude of the Chandler wobble has been observed; some sort of random excitation must be present to keep the wobble going. Various sources of the excitation have been suggested. Present speculation points to deep earthquakes, or the mantle phenomena underlying them, as possibly producing discontinuous changes in the inertia tensor large enough to keep exciting the free-body precession.*

5.7 ■ THE HEAVY SYMMETRICAL TOP WITH ONE POINT FIXED

As a further and more complicated example of the application of the methods of rigid dynamics, let us consider the motion of a symmetrical body in a uniform gravitational field when one point on the symmetry axis is fixed in space. A wide variety of physical systems, ranging from a child's top to complicated gyroscopic navigational instruments, are approximated by such a *heavy symmetrical top*. Both for its practical applications and as an illustration of many of the tech-

*The free precession of Earth's axis is not to be confused with its slow precession about the normal to the ecliptic. This *astronomical* precession of the equinoxes is due to the gravitational torques of the Sun and Moon, which were considered negligible in the above discussion. That the assumption is justified is shown by the long period of the precession of the equinoxes (26,000 years) compared to a period of roughly one year for the force-free precession. The astronomical precession is discussed further below.

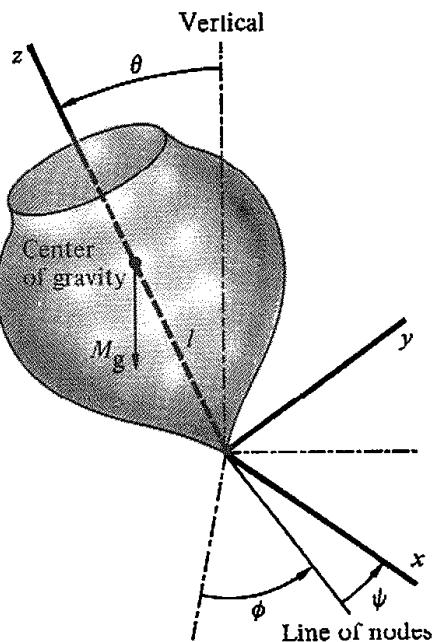


FIGURE 5.7 Euler's angles specifying the orientation of a symmetrical top.

niques previously developed, the motion of the heavy symmetrical top deserves a detailed exposition.

The symmetry axis is of course one of the principal axes and will be chosen as the z axis of the coordinate system fixed in the body.* Since one point is stationary, the configuration of the top is completely specified by the three Euler angles: θ gives the inclination of the z axis from the vertical, ϕ measures the azimuth of the top about the vertical, while ψ is the rotation angle of the top about its own z axis (cf. Fig. 5.7). The distance of the center of gravity (located on the symmetry axis) from the fixed point will be denoted by I .

The rate of change of these three angles give the characteristic motions of the top as

$\dot{\psi}$ = rotation of the top about its own figure axis, z

$\dot{\phi}$ = precession or rotation of the figure axis z about the vertical axis z'

$\dot{\theta}$ = nutation or bobbing up and down of the z figure axis relative to the vertical space axis z' .

For many cases of interest such as the top and the gyroscope, we have $\dot{\psi} \gg \dot{\theta} \gg \dot{\phi}$. Since $I_1 = I_2 \neq I_3$, Euler's equations (5.39') become

*Only the body axes need specific identification here, it will therefore be convenient to designate them in this section as the xz axes, without fear of confusing them with the space axes, which will be designated by the $x'y'z'$ axes

$$I_1\dot{\omega}_1 + \omega_2\omega_3(I_3 - I_2) = N_1,$$

$$I_2\dot{\omega}_2 + \omega_1\omega_3(I_1 - I_3) = N_2,$$

and

$$I_3\dot{\omega}_3 = N_3.$$

Let us consider the case where initially $N_3 = 0 = N_2$, $N_1 \neq 0$, and $\omega_1 = \omega_2 = 0$, $\omega_3 \neq 0$, then ω_3 will be constant. The torque N_1 will cause ω_1 to change since $\omega_1 \neq 0$. Since ω_1 is no longer zero, the second equation requires that ω_2 begin to change also. What this means in terms of an observation is not obvious. We observe the changes in the Euler angles ψ , ϕ , θ and their associated angles in the x' , y' , z' laboratory frame rather than the $\dot{\omega}_1$, $\dot{\omega}_2$, $\dot{\omega}_3$ and their associated angles in the principal axis system. This suggests that the Euler equations may not provide the most useful description of the motion.

The Lagrangian procedure, rather than Euler's equations, will be used to obtain a solution for the motion of the top. Since the body is symmetrical, the kinetic energy can be written as

$$T = \frac{1}{2}I_1(\omega_1^2 + \omega_2^2) + \frac{1}{2}I_3\omega_3^2,$$

or, in terms of Euler's angles, and using Eqs. (4.87), as

$$T = \frac{I_1}{2}(\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta) + \frac{I_3}{2}(\dot{\psi} + \dot{\phi} \cos \theta)^2, \quad (5.50)$$

where the $\dot{\phi}$, $\dot{\theta}$ cross terms in ω_1^2 and ω_2^2 cancel.

It is a well-known elementary theorem that in a constant gravitational field the potential energy is the same as if the body were concentrated at the center of mass. We will however give a brief formal proof here. The potential energy of the body is the sum over all the particles:

$$V = -m_i \mathbf{r}_i \cdot \mathbf{g},$$

where \mathbf{g} is the constant vector for the acceleration of gravity. By Eq. (1.21), defining the center of mass, this is equivalent to

$$V = -M \mathbf{R} \cdot \mathbf{g}, \quad (5.51)$$

which proves the theorem. In terms of the Euler angles,

$$V = Mgl \cos \theta, \quad (5.51')$$

so that the Lagrangian is

$$L = \frac{I_1}{2}(\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta) + \frac{I_3}{2}(\dot{\psi} + \dot{\phi} \cos \theta)^2 - Mgl \cos \theta. \quad (5.52)$$

Note that ϕ and ψ do not appear explicitly in the Lagrangian; they are therefore cyclic coordinates, indicating that the corresponding generalized momenta are constant in time. Now, we have seen that the momentum conjugate to a rotation angle is the component of the total angular momentum along the axis of rotation, which for ϕ is the vertical axis, and for ψ , the z axis in the body. We can in fact show from elementary principles that these components of the angular momentum must be constant in time. Since the torque of gravity is along the line of nodes, there is no component of the torque along either the vertical or the body z axis, for by definition both of these axes are perpendicular to the line of nodes. Hence, the components of the angular momentum along these two axes must be constant in time.

We therefore have two immediate first integrals of the motion:

$$p_\psi = \frac{\partial L}{\partial \dot{\psi}} = I_3(\dot{\psi} + \dot{\phi} \cos \theta) = I_3\omega_3 = I_1a \quad (5.53)$$

and

$$p_\phi = \frac{\partial L}{\partial \dot{\phi}} = (I_1 \sin^2 \theta + I_3 \cos^2 \theta)\dot{\phi} + I_3\dot{\psi} \cos \theta = I_1b. \quad (5.54)$$

Here the two constants of the motion are expressed in terms of new constants a and b . There is one further first integral available; since the system is conservative, the total energy E is constant in time:

$$E = T + V = \frac{I_1}{2}(\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta) + \frac{I_3}{2}\omega_3^2 + Mgl \cos \theta. \quad (5.55)$$

Only three additional quadratures are needed to solve the problem, and they are easily obtained from these three first integrals without directly using the Lagrange equations. From Eq. (5.53), ψ is given in terms of ϕ by

$$I_3\dot{\psi} = I_1a - I_3\dot{\phi} \cos \theta, \quad (5.56)$$

and this result can be substituted in (5.54) to eliminate ψ :

$$I_1\dot{\phi} \sin^2 \theta + I_1a \cos \theta = I_1b,$$

or

$$\dot{\phi} = \frac{b - a \cos \theta}{\sin^2 \theta}. \quad (5.57)$$

Thus, if θ were known as a function of time, Eq. (5.57) could be integrated to furnish the dependence of ϕ on time. Substituting Eq. (5.57) back in Eq. (5.56) results in a corresponding expression for $\dot{\psi}$:

$$\dot{\psi} = \frac{I_1 a}{I_3} - \cos \theta \frac{b - a \cos \theta}{\sin^2 \theta}. \quad (5.58)$$

which furnishes $\dot{\psi}$ if θ is known. Finally, Eqs. (5.57) and (5.58) can be used to eliminate $\dot{\phi}$ and $\dot{\psi}$ from the energy equation, resulting in a differential equation involving θ alone.

First notice that Eq. (5.53) says ω_3 is constant in time and equal to $(I_1/I_3)a$. Therefore, $E - I_3\omega_3^2/2$ is a constant of the motion, which we shall designate as E' . Making use of Eq. (5.57), the energy equation can thus be written as

$$E' = \frac{I_1 \dot{\theta}^2}{2} + \frac{I_1}{2} \frac{(b - a \cos \theta)^2}{\sin^2 \theta} + Mgl \cos \theta. \quad (5.59)$$

Equation (5.59) has the form of an equivalent one-dimensional problem in the variable θ , with the effective potential $V'(\theta)$ given by

$$V'(\theta) = Mgl \cos \theta + \frac{I_1}{2} \left(\frac{b - a \cos \theta}{\sin \theta} \right)^2. \quad (5.60)$$

Thus, we have four constants associated with the motion, the two angular momenta p_ψ and p_ϕ , the energy term $E - \frac{1}{2}I_3\omega_3^2$, and the potential energy term Mgl . It is common to define four normalized constants of the motion as

$$\begin{aligned} \alpha &= \frac{2E - I_3\omega_3^2}{I_1} \\ \beta &= \frac{2Mgl}{I_1} \\ a &= \frac{p_\psi}{I_1} \\ b &= \frac{p_\phi}{I_1}. \end{aligned} \quad (5.61)$$

In terms of these constants, the energy equation (5.55) can be written as

$$\alpha = \dot{\theta}^2 + \frac{(b - a \cos \theta)^2}{\sin^2 \theta} + \beta \cos \theta. \quad (5.62)$$

We will use this one-dimensional problem to discuss the motion in θ , very similarly to what was done in Section 3.3 in describing the radial motion for the central force problem. It is more convenient to change variables as we did for the central force problem. Using the variable $u = \cos \theta$, rewrite Eq. (5.62) as

$$\dot{u}^2 = (1 - u^2)(\alpha - \beta u) - (b - au)^2, \quad (5.62')$$

which can be reduced immediately to a quadrature:

$$t = \int_{u(0)}^{u(t)} \frac{du}{\sqrt{(1-u^2)(\alpha-\beta u) - (b-a u)^2}}. \quad (5.63)$$

With this result, and Eqs. (5.57) and (5.58), ϕ and ψ can also be reduced to quadratures. However, the polynomial in the radical is a cubic so that we have to deal with elliptic integrals. These solutions can be generated on current desk-top computers. In the case of the force-free motion, the physics tends to be obscured in the profusion of mathematics. Fortunately, the general nature of the motion can be discovered without actually performing the integrations.

Before proceeding with the study of the possible solutions of Eq. (5.63), a few comments on the constants defined in Eqs. (5.61) will be useful. Figure 5.7 shows the case where the fixed point is not at the center of mass. If the top is spinning on a horizontal surface, both α and β are greater than zero. If the top is supported by a stand that allows it to dip below horizontal, β is still larger than zero, but α could be positive or negative. Another common application is the gyroscope where the center of mass is the fixed point. In terms of Fig. 5.7, α is the energy in the system excluding the x_3 angular kinetic energy. For the gyroscope, $\beta = 0$ and $\alpha \geq 0$. We shall restrict our attention to situations in which the rotational kinetic energy about the x_3 axis is much larger than the kinetic energy about the other two axes.

It is convenient to designate the right-hand side of Eq. (5.62') as a function $f(u)$ and discuss the behavior of the cubic equation

$$f(u) = \beta u^3 - (\alpha + a^2)u^2 + (2ab - \beta)u + (\alpha - b^2).$$

For the gyroscope, $f(u)$ is only a quadratic equation since $\beta = 0$, while for the top the full cubic equation must be considered. Since many of the applications of the gyroscope use torque-free mountings, precession and nutations are suppressed so the gyroscope motions are trivial. To understand the general motions of a spinning body, we will consider only cases where $\beta > 0$.

The roots of the cubic polynomial furnish the angles at which $\dot{\theta}$ changes sign, that is, the "turning angles" in θ . Knowing these angles will give qualitative information about the motion. There are three roots to a cubic equation and three possible combinations of solutions. There can be one real root and a complex conjugate pair of roots; there can be three real roots, two of which are equal; and there can be three real and unequal roots. These possibilities depend upon the relative signs and magnitudes of the four constants in Eqs. (5.61). There is also the physical constraint that the solution u must satisfy $-1 < u \leq 1$. We will draw all figures as if $u > 0$, which would be the case if the top is supported by a horizontal surface. Recall that a point support could allow the smallest root to be less than zero.

For u large, the dominant term in $f(u)$ is βu^3 . Since β (cf. Eqs. (5.61)) is always a positive constant $f(u)$ is positive for large positive u and negative for large negative u . At points $u = \pm 1$, $f(u)$ becomes equal to $-(b \mp a)^2$ and is therefore always negative, except for the unusual case where $u = \pm 1$ is a root

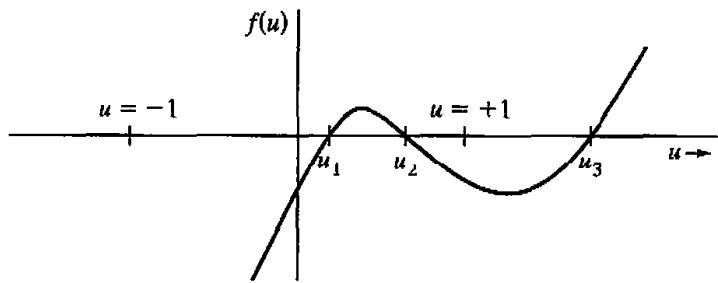


FIGURE 5.8 Illustrating the location of the turning angles of θ in the motion of a heavy symmetric top supported on a horizontal plane. A point support could allow one of the roots to be negative.

(corresponding to a vertical top). Hence, at least one root must lie in the region $u > 1$, a region that does not correspond to real angles. Indeed, physical motion of the top can occur only when u^2 is positive somewhere in the interval between $u = -1$ and $u = +1$, that is, θ between 0 and $+\pi$. We must conclude therefore that for any actual top $f(u)$ will have two roots, u_1 and u_2 , between -1 and $+1$ (cf. Fig. 5.8), and that the top moves such that $\cos \theta$ always remains between these two roots. The location of these roots, and the behavior of $\dot{\phi}$ and $\dot{\psi}$ for values of θ between them, provide much qualitative information about the motion of the top.

It is customary to depict the motion of the top by tracing the curve of the intersection of the figure axis on a sphere of unit radius about the fixed point. This curve is known as the *locus* of the figure axis. The polar coordinates of a point on the locus are identical with the Euler angles θ, ϕ for the body system. From the discussion in the preceding paragraph, we can see that the locus lies between the two bounding circles of colatitude $\theta_1 = \arccos u_1$ and $\theta_2 = \arccos u_2$, with $\dot{\theta}$ vanishing at both circles. The shape of the locus curve is in large measure determined by the value of the root of $b - au$, which we denote by u' :

$$u' = \frac{b}{a}. \quad (5.64)$$

Suppose, for example, the initial conditions are such that u' is larger than u_2 . Then, by Eq. (5.57), $\dot{\phi}$ will always have the same sign for the allowed inclination angles between θ_1 and θ_2 . Hence, the locus of the figure axis must be tangent to the bounding circles in such a manner that $\dot{\phi}$ is in the same direction at both θ_1 and θ_2 , as is shown in Fig. 5.9(a). Since $\dot{\phi}$ therefore increases secularly in one direction or the other, the axis of the top may be said to *precess* about the vertical axis. But it is not the regular precession encountered in force-free motion, for as the figure axis goes around, it nods up and down between the bounding angles θ_1 and θ_2 —the top *nutates* during the precession.

Should b/a be such that u' lies between u_1 and u_2 , the direction of the precession will be different at the two bounding circles, and the locus of the figure axis exhibits loops, as shown in Fig. 5.9(b). The average of $\dot{\phi}$ will not vanish however so that there is always a net precession in one direction or the other. It can

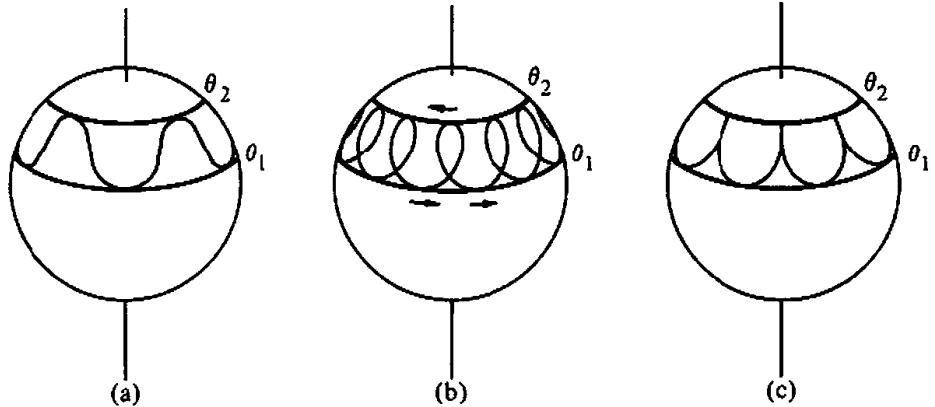


FIGURE 5.9 The possible shapes for the locus of the figure axis on the unit sphere.

also happen that u' coincides with one of the roots of $f(u)$. At the corresponding bounding circles, both $\dot{\theta}$ and $\dot{\phi}$ must then vanish, which requires that the locus have cusps touching the circle, as shown in Fig. 5.9(c).

This last case is not as exceptional as it sounds; it corresponds in fact to the initial conditions usually stipulated in elementary discussions of tops: We assume that initially the symmetrical top is spinning about its figure axis, which is fixed in some direction θ_0 . At time $t = 0$, the figure axis is released and the problem is to describe the subsequent motion. Explicitly, these initial conditions are that at $t = 0$, $\theta = \theta_0$ and $\dot{\theta} = \dot{\phi} = 0$. The quantity $u_0 = \cos \theta_0$ must therefore be one of the roots of $f(u)$; in fact, it corresponds to the upper circle:

$$u_0 = u_2 = u' = \frac{b}{a}. \quad (5.65)$$

For proof, note that with these initial conditions E' is equal to $Mgl \cos \theta_0$, and that the terms in E' derived from the top's kinetic energy can never be negative. Hence, as $\dot{\theta}$ and $\dot{\phi}$ begin to differ from their initial zero values, energy can be conserved only by a decrease in $Mgl \cos \theta$, i.e., by an increase in θ . The initial θ_0 is therefore the same as θ_2 , the minimum value θ can have. When released in this manner, *the top always starts to fall*, and continues to fall until the other bounding angle θ_1 is reached, precessing the meanwhile. The figure axis then begins to rise again to θ_2 , the complete motion being as shown in Fig. 5.9(c).

Some quantitative predictions can be made about the motion of the top under these initial conditions of vanishing $\dot{\theta}$ and $\dot{\phi}$, provided that the initial kinetic energy of rotation about the z-axis is assumed large compared to the maximum change in potential energy:

$$\frac{1}{2} I_3 \omega_3^2 \gg 2Mgl. \quad (5.66)$$

The effects of the gravitational torques, namely, the precession and accompanying nutation, will then be only small perturbations on the dominant rotation of the top about its figure axis. In this situation, we speak of the top as being a “fast top.”

With this assumption we can obtain expressions for the extent of the nutation, the nutation frequency, and the average frequency of precession.

The extent of the nutation under these given initial conditions is given by $u_1 - u_0$, where u_1 is the other physical root of $f(u)$. The initial conditions $E' = Mgl \cos \theta_0$ is equivalent to the equality

$$\alpha = \beta u_0.$$

With this relation, and the conditions of Eq. (5.65), $f(u)$ can be rewritten more simply as

$$f(u) = (u_0 - u) \left[\beta(1 - u^2) - a^2(u_0 - u) \right]. \quad (5.67)$$

The roots of $f(u)$ other than u_0 are given by the roots of the quadratic expression in the brackets, and the desired root u_1 therefore satisfies the equation

$$(1 - u_1^2) - \frac{a^2}{\beta}(u_0 - u_1) = 0. \quad (5.68)$$

Denoting $u_0 - u$ by x and $u_0 - u_1$ by x_1 , Eq. (5.68) can be rewritten as

$$x_1^2 + px_1 - q = 0, \quad (5.69)$$

where

$$p = \frac{a^2}{\beta} - 2 \cos \theta_0, \quad q = \sin^2 \theta_0.$$

The condition for a “fast” top, Eq. (5.66), implies that p is much larger than q . This can be seen by writing the ratio a^2/β as

$$\frac{a^2}{\beta} = \left(\frac{I_3}{I_1} \right) \frac{I_3 \omega_3^2}{2Mgl}.$$

Except in the case that $I_3 \ll I_1$ (which would correspond to a top in the unusual shape of a cigar), the ratio is much greater than unity, and $p \gg q$. To first order in the small quantity q/p , the only physically realizable root of Eq. (5.68) is then

$$x_1 = \frac{q}{p}.$$

Neglecting $2 \cos \theta_0$ compared to a^2/β , this result can be written

$$x_1 = \frac{\beta \sin^2 \theta_0}{a^2} = \frac{I_1}{I_3} \frac{2Mgl}{I_3 \omega_3^2} \sin^2 \theta_0. \quad (5.70)$$

Thus, the extent of the nutation, as measured by $x_1 = u_0 - u_1$, goes down as $1/\omega_3^2$. The faster the top is spun, the less is the nutation.

The *frequency* of nutation likewise can easily be found for the “fast” top. Since the amount of nutation is small, the term $(1 - u^2)$ in Eq. (5.67) can be replaced by its initial value, $\sin^2 \theta_0$. Equation (5.67) then reads, with the help of Eq. (5.70),

$$f(u) = \dot{x}^2 = a^2 x(x_1 - x).$$

If we shift the origin of x to the midpoint of its range, by changing variable to

$$y = x - \frac{x_1}{2},$$

then the differential equation becomes

$$\dot{y}^2 = a^2 \left(\frac{x_1^2}{4} - y^2 \right),$$

which on differentiation again reduces to the familiar equation for simple harmonic motion

$$\ddot{y} = -a^2 y.$$

In view of the initial condition $x = 0$ at $t = 0$, the complete solution is

$$x = \frac{x_1}{2}(1 - \cos at), \quad (5.71)$$

where x_1 is given by (5.70). The angular frequency of nutation of the figure axis between θ_0 and θ_1 is therefore

$$a = \frac{I_3}{I_1} \omega_3, \quad (5.72)$$

which *increases* the faster the top is spun initially.

Finally, the angular velocity of precession, from (5.57), is given by

$$\dot{\phi} = \frac{a(u_0 - u)}{\sin^2 \theta} \approx \frac{ax}{\sin^2 \theta_0},$$

or, substituting Eqs. (5.72) and (5.70),

$$\dot{\phi} = \frac{\beta}{2a}(1 - \cos at). \quad (5.73)$$

The rate of precession is therefore not uniform but varies harmonically with time, with the same frequency as the nutation. The *average* precession frequency however is

$$\bar{\dot{\phi}} = \frac{\beta}{2a} = \frac{Mgl}{I_3 \omega_3}, \quad (5.74)$$

which indicates that the rate of precession decreases as the initial rotational velocity of the top is increased.

We are now in a position to present a complete picture of the motion of the fast top when the figure axis initially has zero velocity. Immediately after the figure axis is released, the initial motion of the top is always to fall under the influence of gravity. But as it falls, the resultant torque around the axis of fall causes the top to pick up a precession velocity, directly proportional to the extent of its fall, which starts the figure axis moving sideways about the vertical. The initial fall results in a periodic nutation of the figure axis in addition to the precession. As the top is spun faster and faster, the extent of the nutation decreases rapidly, although the frequency of nutation increases, while at the same time the precession about the vertical becomes slower. In practice, for a sufficiently fast top the nutation is damped out by the friction at the pivot and becomes unobservable. The top then appears to precess uniformly about the vertical axis. Because the precession is regular only in appearance, Klein and Sommerfeld have dubbed it a *pseudoregular* precession. In most of the elementary discussions of precession, the phenomenon of nutation is neglected. As a consequence, such derivations seem to lead to the paradoxical conclusion that upon release the top *immediately* begins to precess uniformly, a motion that is *normal* to the forces of gravity that are the ultimate cause of the precession. Our discussion of pseudoregular precession serves to resolve the paradox; the precession builds up continuously from rest without any infinite accelerations, and the initial tendency of the top is to move in the direction of the forces of gravity.

It is of interest to determine exactly what initial conditions will result in a true regular precession. In such a case, the angle θ remains constant at its initial value θ_0 , which means that $\dot{\theta}_1 = \dot{\theta}_2 = \theta_0$. In other words, $f(u)$ must have a double root at u_0 (cf. Fig. 5.10), or

$$f(u) = \dot{u}^2 = 0, \quad \frac{df}{du} = 0; \quad u = u_0.$$

The first of these conditions, from Eq. (5.62') with $\dot{u} = 0$, implies

$$(\alpha - \beta u_0) = \frac{(b - au_0)^2}{1 - u_0^2}, \quad (5.75)$$

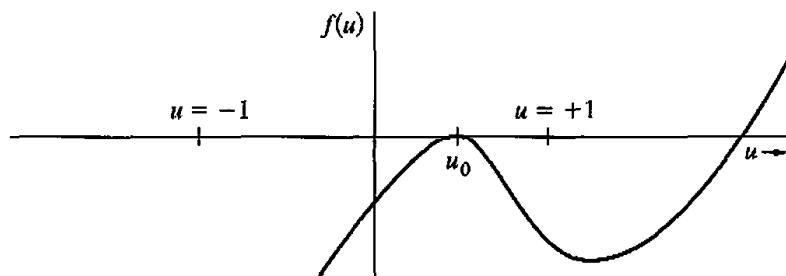


FIGURE 5.10 Appearance of $f(u)$ for a regular precession.

the second corresponds to

$$\frac{\beta}{2} = \frac{a(b - au_0)}{1 - u_0^2} - u_0 \frac{(\alpha - \beta u_0)}{1 - u_0^2}. \quad (5.76)$$

Substitution of Eq. (5.75) in Eq. (5.76) leads, in view of Eq. (5.57) for $\dot{\phi}$, to a quadratic equation for $\dot{\phi}$:

$$\frac{\beta}{2} = a\dot{\phi} - \dot{\phi}^2 \cos \theta_0. \quad (5.76')$$

With the definitions of β and a , Eq. (5.61), this can be written in two alternative forms, depending on whether a is expressed in terms of ω_3 or the (constant) $\dot{\psi}$ and $\dot{\phi}$

$$Mgl = \dot{\phi}(I_3\omega_3 - I_1\dot{\phi} \cos \theta_0), \quad (5.77)$$

or

$$Mgl = \dot{\phi}(I_3\dot{\psi} - (I_1 - I_3)\dot{\phi} \cos \theta_0). \quad (5.77')$$

The initial conditions for the problem of the heavy top require the specification of $\theta, \phi, \psi, \dot{\theta}, \dot{\phi}$, and, say, either $\dot{\psi}$ or ω_3 at the time $t = 0$. Because they are cyclic, the initial values of ϕ and ψ are largely irrelevant, and in general we can choose any desired value for each of the four others. But if in addition we require that the motion of the figure axis be one of uniform precession without nutation, then our choice of these four initial values is no longer completely unrestricted. Instead, they must satisfy either of Eqs. (5.77). For $\dot{\theta} = 0$, we may still choose initial values of θ and ω_3 , almost arbitrarily, but the value of $\dot{\phi}$ is then determined. The phrase "almost arbitrarily" is used because Eqs. (5.77) are quadratic, and for $\dot{\phi}$ to be real, the discriminant of Eq. (5.77) must be positive:

$$I_3^2\omega_3^2 > 4MglI_1 \cos \theta_0. \quad (5.78)$$

For $\theta_0 > \pi/2$ (a top mounted so its center of mass is below the fixed point), then any value of ω_3 can lead to uniform precession. But for $\theta_0 < \pi/2$, ω_3 must be chosen to be above a minimum value ω'_3 ,

$$\omega_3 > \omega'_3 = \frac{2}{I_3} \sqrt{MglI_1 \cos \theta_0} \quad (5.79)$$

to achieve the same situation. Similar conditions can be obtained from Eq. (5.77') for the allowable values of ψ . As a result of the quadratic nature of Eq. (5.77), there will in general be two solutions for $\dot{\phi}$, known as the "fast" and "slow" precession. Also note that (5.77) can never be satisfied by $\dot{\phi} = 0$ for finite $\dot{\psi}$ or ω_3 ; to obtain uniform precession, we must always give the top a shove to start it on its

way. Without this correct initial precessional velocity, we can obtain at best only a pseudoregular precession.

If the precession is slow, so that $\dot{\phi} \cos \theta_0$ may be neglected compared to a , then an approximate solution for $\dot{\phi}$ is

$$\dot{\phi} \approx \frac{\beta}{2a} = \frac{Mgl}{I_3\omega_3} \quad (\text{slow}),$$

which agrees with the average rate of pseudoregular precession for a fast top. This result is to be expected of course; if the rate of precession is slow, there is little difference between starting the gyroscope off with a little shove or with no shove at all. Note that with this value of $\dot{\phi}$, the neglect of $\dot{\phi} \cos \theta_0$ compared to a is equivalent to requiring that ω_3 be much greater than the minimum allowed value. For such large values of ω_3 , the “fast” precession is obtained when $\dot{\phi}$ is so large that Mgl is small compared to the other terms in Eq. (5.77):

$$\dot{\phi} = \frac{I_3\omega_3}{I_1 \cos \theta_0} \quad (\text{fast}).$$

The fast precession is independent of the gravitational torques and can in fact be related to the precession of a free body (see Derivation 6a in the Exercises).

One further case deserves some attention, namely, when $u = 1$ corresponds to one of the roots of $f(u)$.* Suppose, for instance, a top is set spinning with its figure axis initially vertical. Clearly then $b = a$, for I_1b and I_1a are the constant components of the angular momentum about the vertical axis and the figure axis respectively, and these axes are initially coincident. Since the initial angular velocity is only about the figure axis, the energy equation (5.59) evaluated at time $t = 0$ states that

$$E' = E - \frac{1}{2} I_3\omega_3^2 = Mgl.$$

By the definitions of α and β (Eq. (5.61)), it follows that $\alpha = \beta$.

The energy equation at any angle may therefore be written as

$$\dot{u}^2 = (1 - u^2)\beta(1 - u) - a^2(1 - u)^2$$

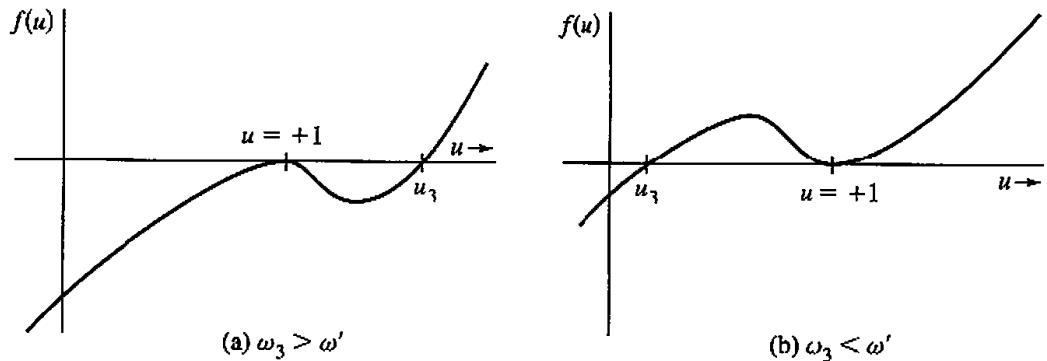
or

$$\dot{u}^2 = (1 - u)^2 \left[\beta(1 + u) - a^2 \right].$$

The form of the equation indicates that $u = 1$ is always a double root, with the third root given by

$$u_3 = \frac{a^2}{\beta} - 1.$$

*Note that this must be treated as a special case, since in the previous discussions factors of $\sin^2 \theta$ were repeatedly divided out of the expressions.

FIGURE 5.11 Plot of $f(u)$ when the figure axis is initially vertical.

If $a^2/\beta > 2$ (which corresponds to the condition for a “fast” top), u_3 is larger than 1 and the only possible motion is for $u = 1$; the top merely continues to spin about the vertical. For this state of affairs, the plot of $f(u)$ appears as shown in Fig. 5.11(a). On the other hand, if $a^2/\beta < 2$, the third root u_1 is then less than 1, $f(u)$ takes on the form shown in Fig. 5.11(b), and the top will nutate between $\theta = 0$ and $\theta = \theta_3$. There is thus a critical angular velocity, ω' , above which only vertical motion is possible, whose value is given by

$$\frac{a^2}{\beta} = \left(\frac{I_3}{I_1} \right) \frac{I_3 \omega'^2}{2Mgl} = 2$$

or

$$\omega'^2 = 4 \frac{MglI_1}{I_3^2}, \quad (5.80)$$

which is identical with Eq. (5.79) for the minimum frequency for uniform precession with $\theta_0 = 0$.

In practice, if a top is started spinning with its axis vertical and with ω_3 greater than the critical angular velocity, it will continue to spin quietly for a while about the vertical (hence the designation as a “sleeping” top). However, friction gradually reduces the frequency of rotation below the critical value, and the top then begins to wobble in ever larger amounts as it slows down.

The effects of friction (which of course cannot be directly included in the Lagrangian framework) can give rise to unexpected phenomena in the behavior of tops. A notable example is the “tippie-top,” which consists basically of somewhat more than half a sphere with a stem added on the flat surface. When set rotating with the spherical surface downwards on a hard surface, it proceeds to skid and nutate until it eventually turns upside down, pivoting on the stem, where it then behaves as a normal “sleeping” top. The complete reversal of the angular momentum vector is the result of frictional torque occurring as the top skids on its spherical surface.

A large and influential technology is based on the applications of rapidly spinning rigid bodies, particularly through the use of what are called “gyroscopes.” Basically, a three-frame gyroscope is a symmetrical top rotated very rapidly by external means about the figure axis and mounted in gimbals so that the motion of the figure axis is unrestricted about three perpendicular spatial axes while the center of gravity remains stationary. The figure axis maintains the same direction in space no matter how the mounting is reoriented, a phenomenon called gyroscopic inertia. Such an instrument can indicate the roll, pitch, and attitude directions of an airplane flying “blind” by using the xyz Euler angle convention described in Section 4.4 and Appendix A.

If external torques are suitably exerted on the gyroscope, it will undergo the precession and nutation motions described earlier for the heavy top. However, the condition for the “fast” top is abundantly satisfied, so that the extent of the nutation is always very small, and moreover is deliberately damped out by the method of mounting. The only gyroscopic phenomenon then observed is precession, and the mathematical treatment required to describe this precession can be greatly simplified. We can see how to do this by generalization from the case of the heavy symmetrical top.

If \mathbf{R} is the radius vector along the figure axis from the fixed point to the center of gravity, then the gravitational torque exerted on the top is

$$\mathbf{N} = \mathbf{R} \times M\mathbf{g}, \quad (5.81)$$

where \mathbf{g} is the *downward* vector of the acceleration of gravity. If \mathbf{L}_3 is the vector along the figure axis, describing the angular momentum of rotation about the figure axis, and $\boldsymbol{\omega}_p$, known as the precession vector, is aligned along the vertical with magnitude equal to the mean precession angular velocity $\bar{\phi}$, Eq. (5.74), then the sense and magnitude of the (pseudoregular) precession is given by

$$\boldsymbol{\omega}_p \times \mathbf{L}_3 = \mathbf{N}. \quad (5.82)$$

Since any torque about the fixed point or center of mass can be put in the form $\mathbf{R} \times \mathbf{F}$, similar to Eq. (5.81), the resulting average precession rate for a “fast” top can always be derived from Eq. (5.82), with the direction of the force \mathbf{F} defining the precession axis. Almost all engineering applications of gyroscopes involve the equilibrium behavior (i.e., neglecting transients) which can be derived from Eq. (5.82).

Free from any torques, a gyroscope spin axis will always preserve its original direction relative to an inertial system. Gyros can therefore be used to indicate or maintain specific directions, e.g., provide stabilized platforms. As indicated by Eq. (5.82), through the precession phenomena they can sense and measure angular rotation rates and applied torques. Note from Eq. (5.82) that the precession rate is proportional to the torque, whereas in a nonspinning body it is the angular acceleration that is given by the torque. Once the torque is removed, a nonspinning

body will continue to move; under similar conditions a gyro simply continues spinning without precessing.

The gyrocompass involves more complicated considerations because here we are dealing with the behavior of a gyroscope fixed in a noninertial system, while Earth rotates underneath it. In a gyrocompass, an additional precession is automatically applied by an external torque at a rate just enough to balance Earth's rotation rate. Once set in the direction of Earth's rotation, i.e., the north direction, the gyrocompass then preserves this direction, at least in slowly moving vehicles. What has been presented here is admittedly an oversimplified, highly compressed view of the fascinating technological uses of fast spinning bodies. To continue further in this direction would regrettably lead us too far afield.

There are however two examples of precession phenomena in nature for which a somewhat fuller discussion would be valuable, both for the great interest in the phenomena themselves and as examples of the techniques derived in this chapter. The first concerns the types of precession that arise from the torques induced by Earth's equatorial "bulge," and the second is the precession of moving charges in a magnetic field. The next two sections are concerned with these examples.

5.8 ■ PRECESSION OF THE EQUINOXES AND OF SATELLITE ORBITS

It has been mentioned previously that Earth is a top whose figure axis is precessing about the normal to the ecliptic, the plane of Earth's orbit, a motion known astronomically as the precession of the equinoxes. Were Earth completely spherical, none of the other members of the solar system could exert a gravitational torque on it. But, as has been pointed out, Earth deviates slightly from a sphere, being closely approximated by an oblate spheroid of revolution. It is just the net torque on the resultant equatorial "bulge" arising from gravitational attraction, chiefly of the Sun and Moon, that sets Earth's axis precessing in space.

To calculate the rate of this precession, a slight excursion into potential theory is needed to find the mutual gravitational potential of a mass point (representing the sun or the moon) and a nonspherical distribution of matter. We will find the properties of the inertia tensor as obtained above very useful in the derivation of this potential.

Consider a distribution of mass points forming one body, and a single mass point, mass M , representing the other (cf. Fig. 5.12). If r_i is the distance between the i th point in the distribution and the mass point M , then the mutual gravitational potential between the two bodies is*

$$V = -\frac{GMm_i}{r_i} = -\frac{GMm_i}{r\sqrt{1 + \left(\frac{r'_i}{r}\right)^2 - 2\frac{r'_i}{r}\cos\psi_i}}. \quad (5.83)$$

*It may be worth a reminder that summation is implied over repeated subscripts.

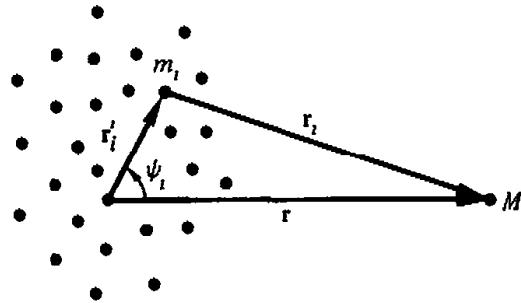


FIGURE 5.12 Geometry involved in gravitational potential between an extended body and a mass point.

In this last expression the terminology of Fig. 5.12 is used: \mathbf{r}'_i is the radius vector to the i th particle from a particular point, which will later be taken to be the center of mass of the first body, \mathbf{r} is the corresponding radius vector to the mass point M , and ψ_i is the angle between the two vectors. It is well known that a simple expansion in terms of Legendre polynomials can be given for Eq. (5.83); in fact, the reciprocal of the square root in Eq. (5.83) is known as the *generating function* for Legendre polynomials, so that

$$V = -\frac{GM}{r} \sum_{n=0}^{\infty} m_i \left(\frac{\mathbf{r}'_i}{r}\right)^n P_n(\cos \psi_i), \quad (5.84)$$

providing r , the distance from the origin to M , is much greater than any r'_i . We shall make use of only the first three Legendre polynomials that, for reference, are

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{1}{2}(3x^2 - 1). \quad (5.85)$$

For a continuous spherical body, with only a radial variation of density, all terms except the first in Eq. (5.84) can easily be shown to vanish. Thus, the n th term inside the summation, for a body with spherical symmetry and mass density $\rho(r')$, can be written

$$\iiint dV' \rho(r') \left(\frac{\mathbf{r}'_i}{r}\right)^n P_n(\cos \psi).$$

Using spherical polar coordinates, with the polar axis along \mathbf{r} , this becomes

$$\int r'^2 dr' \rho(r') \left(\frac{\mathbf{r}'_i}{r}\right)^n \int_{-1}^{+1} d(\cos \psi) P_n(\cos \psi).$$

From the orthonormal properties of P_n with respect to P_0 , the integral over $\cos \psi$ vanishes except for $n = 0$, which proves the statement.

If the body deviates only slightly from spherical symmetry, as is the case with Earth, we would expect the terms in Eq. (5.84) beyond $n = 0$ to decrease rapidly

with increasing n . It will therefore be sufficient to retain only the first nonvanishing correction term in Eq. (5.48) to the potential for a sphere. Now, the choice of the center of mass as origin causes the $n = 1$ term to vanish identically, since it can be written

$$-\frac{GM}{r^2} m_i r'_i \cos \psi_i = -\frac{GM}{r^3} \mathbf{r} \cdot \mathbf{m}_i \mathbf{r}'_i,$$

which is zero, by definition of the center of mass. The next term, for $n = 2$, can be written

$$\frac{GM}{2r^3} m_i r'^2 (1 - 3 \cos^2 \psi_i).$$

Simple tensor manipulation gives the complete second-order approximation to the nonspherical potential as

$$V = -\frac{GMm}{r} + \frac{GM}{2r^3} (3I_r - \text{Tr } I),$$

where m is the mass of the first body (Earth), I_r is the moment of inertia about the direction of \mathbf{r} , and I is the moment of inertia tensor in the principal axis system. From the diagonal representation of the inertia tensor in the principal axis system, its trace is just the sum of the principal moments of inertia, so that V can be written as

$$V = -\frac{GMm}{r} + \frac{GM}{2r^3} [3I_r - (I_1 + I_2 + I_3)]. \quad (5.86)$$

Equation (5.86) is sometimes known as *MacCullagh's formula*. So far, no assumption of rotational symmetry has been made. Let us now take the axis of symmetry to be along the third principal axis, so that $I_1 = I_2$. If α, β, γ are the direction cosines of \mathbf{r} relative to the principal axes, then the moment of inertia I_r can be expressed as

$$I_r = I_1(\alpha^2 + \beta^2) + I_3\gamma^2 = I_1 + (I_3 - I_1)\gamma^2. \quad (5.87)$$

With this form for I_r , the potential, Eq. (5.86), becomes

$$V = -\frac{GMm}{r} + \frac{GM(I_3 - I_1)}{2r^3} (3\gamma^2 - 1),$$

or

$$V = -\frac{GMm}{r} + \frac{GM(I_3 - I_1)}{r^3} P_2(\gamma). \quad (5.88)$$

The general form of Eq. (5.88) could have been foretold from the start, for the potential from a mass distribution obeys Poisson's equation. The solution appropriate to the symmetry of the body, as is well known, is an expansion of terms

of the form $P_n(\gamma)/r^{n+1}$, of which Eq. (5.88) shows the first two nonvanishing terms. However, this approach does not give the coefficients of the terms any more simply than the derivation employed here. It should also be remarked that the expansion of V is the gravitational analog of the multipole expansion of, say, the electrostatic potential of an arbitrary charged body. The $n = 1$ term is absent here because there is only one sign of gravitational "charge" and there can be no gravitational dipole moment. Further, the inertia tensor is defined analogously to the quadrupole moment tensor. Therefore, the mechanical effects we are seeking can be said to arise from the gravitational quadrupole moment of the oblate Earth.*

Of the terms in Eq. (5.88) for the potential, the only one that depends on the orientation of the body, and thus could give rise to torques, is

$$V_2 = \frac{GM(I_3 - I_1)}{r^3} P_2(\gamma). \quad (5.89)$$

For the example of Earth's precession, it should be remembered that γ is the direction cosine between the figure axis of Earth and the radius vector from Earth's center to the Sun or Moon. As these bodies go around their apparent orbits, γ will change. The relation of γ to the more customary astronomical angles can be seen from Fig. 5.13 where the orbit of the Sun or Moon is taken as being in the xy plane, and the figure axis of the body in the xz plane. The angle θ between the figure axis and the z direction is the obliquity of the figure axis. The dot product of a unit vector along the figure axis with the radius vector to the celestial body involves only the products of their x -components, so that

$$\gamma = \sin \theta \cos \eta.$$

Hence, V_2 can be written

$$V_2 = \frac{GM(I_3 - I_1)}{2r^3} (3\sin^2 \theta \cos^2 \eta - 1).$$

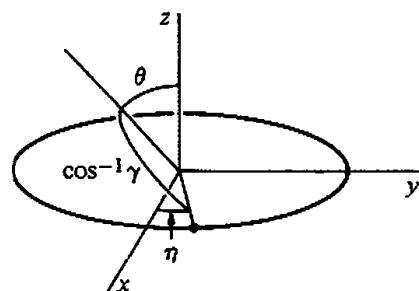


FIGURE 5.13 Figure axis of Earth relative to orbit of mass point.

*Note that so far nothing in the argument restricts the potential of Eq. (5.88) to *rigid* bodies. The constraint of rigidity enters only when we require from here on that the principal axes be fixed in the body and the associated moments of inertia be constant in time.

As we shall see, the orbital motion is very rapid compared to the precessional motion, and for the purpose of obtaining the mean precession rate, it will be adequate to average V_2 over a complete orbital period of the celestial body considered. Since the apparent orbits of the Sun and Moon have low eccentricities, r can be assumed constant and the only variation is in $\cos \eta$. The average of $\cos^2 \eta$ over a complete period is $\frac{1}{2}$, and the averaged potential is then

$$\bar{V}_2 = \frac{GM(I_3 - I_1)}{2r^3} \left(\frac{3}{2} \sin^2 \theta - 1 \right) = \frac{GM(I_3 - I_1)}{2r^3} \left(\frac{1}{2} - \frac{3}{2} \cos^2 \theta \right),$$

or, finally,

$$\bar{V}_2 = -\frac{GM(I_3 - I_1)}{2r^3} P_2(\cos \theta). \quad (5.90)$$

The torque derived from Eq. (5.90) is perpendicular to both the figure axis and the normal to the orbit (which plays the same role as the vertical axis for the heavy top). Hence, the precession is about the direction of the orbit normal vector. The magnitude of the precession rate can be obtained from Eq. (5.82), but because the potential differs in form from that for the heavy top, it may be more satisfying to obtain a more formal derivation. For any symmetric body in which the potential is a function of $\cos \theta$ only, the Lagrangian can be written, following Eq. (5.52), as

$$L = \frac{I_1}{2} (\dot{\theta}^2 + \dot{\phi}^2 \sin^2 \theta) + \frac{I_3}{2} (\dot{\psi} + \dot{\phi} \cos \theta)^2 - V(\cos \theta). \quad (5.91)$$

If we are to assume only uniform precession and are not concerned about the necessary initial conditions, we can simply take $\dot{\theta}$ and $\ddot{\theta}$ to be zero in the equations of motion. The Lagrange equation corresponding to θ is then

$$\frac{\partial L}{\partial \theta} = I_1 \dot{\phi}^2 \sin \theta \cos \theta - I_3 \dot{\phi} \sin \theta (\dot{\psi} + \dot{\phi} \cos \theta) - \frac{\partial V}{\partial \theta} = 0$$

or

$$I_3 \omega_3 \dot{\phi} - I_1 \dot{\phi}^2 \cos \theta = \frac{\partial V}{\partial (\cos \theta)}, \quad (5.92)$$

which is the analog of Eq. (5.76') for a more general potential. For slow precession, which means basically that $\dot{\phi} \ll \omega_3$, the $\dot{\phi}^2$ terms in Eq. (5.92) can be neglected, and the rate of uniform precession is given by

$$\dot{\phi} = \frac{1}{I_3 \omega_3} \frac{\partial V}{\partial (\cos \theta)}. \quad (5.93)$$

From Eq. (5.51') we see that for the heavy top Eq. (5.93) agrees with the average result of Eq. (5.74). With the potential of Eq. (5.90), the precession rate is

$$\dot{\phi} = -\frac{3GM}{2\omega_3 r^3} \frac{I_3 - I_1}{I_3} \cos \theta. \quad (5.94)$$

For the case of the precession due to the Sun, this formula can be put in a simpler form, by taking r as the semimajor axis of Earth's orbit and using Kepler's law, Eq. (3.71), in the form

$$\omega_0^2 = \left(\frac{2\pi}{\tau} \right)^2 = \frac{GM}{r^3}.$$

The precession rate, relative to the orbital angular velocity, ω_0 , is then

$$\frac{\dot{\phi}}{\omega_0} = -\frac{3}{2} \frac{\omega_0}{\omega_3} \frac{I_3 - I_1}{I_3} \cos \theta. \quad (5.95)$$

With the value of $(I_3 - I_1)/I_3$ as given in Section 5.6, and $\theta = 23^\circ 27'$, Eq. (5.95) says that the solar-induced precession would be such as to cause a complete rotation of the figure axis about the normal to the ecliptic (plane of Earth's orbit) in about 81,000 years.

The Moon is far less massive than the Sun, but it is also much closer; the net result is that the lunar-induced precession rate is over twice that caused by the Sun. Since the lunar orbit is close to the ecliptic and has the same sense as the apparent solar orbit, the two precessions nearly add together arithmetically, and the combined lunisolar precession rate is $50.25''/\text{year}$, or one complete rotation in about 26,000 years. Note that this rate of precession is so slow that the approximation of neglecting $\dot{\phi}$ compared to ω_3 is abundantly satisfied. Because the Sun, Moon, and Earth are in constant relative motion, and the Moon's orbit is inclined about 5° to the ecliptic, the precession exhibits irregularities designated as *astronomic nutation*. The extent of these periodic irregularities is not large—about $9''$ of arc in θ and about $18''$ in $\dot{\phi}$. Even so, they are far larger than the true nutation that, as Klein and Sommerfeld have shown, is manifested by the Chandler wobble whose amplitude is never more than a few tenths of an arc second.

One further application can be made of the potential, Eq. (5.88), and associated uniform precession rate, Eq. (5.93). It has been stressed that the potential represents a *mutual* gravitational interaction; if it results in torques acting on the spinning Earth, it also gives rise to (noncentral) forces acting on the mass point M . The effect of these small forces appears as a precession of the plane of the orbit of the mass point, relative to an inertial frame. It is possible to obtain an approximate formula for this precession by an argument again based on the behavior of spinning rigid bodies.

Since the precession rates are small compared to the orbital angular velocity, we can again average over the orbit. The averaging in effect replaces the particle by a rigid ring of mass M with the same radius as the (assumed circular) orbit, spinning about the figure axis of the ring with the orbital frequency. Equation (5.90) gives the potential field in which this ring is located, with θ the angle

between the figure axes of the ring and Earth. The average precession rate is still given by Eq. (5.93), but now I_3 and ω_3 refer to the spinning ring and not Earth. It would therefore be better to rewrite Eq. (5.93) for this application as

$$\dot{\phi} = \frac{\tau}{2\pi Mr^2} \frac{\partial V}{\partial(\cos \theta)}, \quad (5.93')$$

and Eq. (5.94) appears as

$$\dot{\phi} = -\frac{\tau}{2\pi} \frac{3}{2} \frac{G(I_3 - I_1)}{r^5} \cos \theta. \quad (5.94')$$

Equation (5.94') could be used, for example, to find the precession of the orbit of the Moon due to Earth's oblateness. A more current application would be to the precession of nearly circular orbits of artificial satellites revolving about Earth. The fraction of a complete precession rotation in one period of the satellite is

$$\frac{\dot{\phi}\tau}{2\pi} = -\left(\frac{\tau}{2\pi}\right)^2 \frac{3}{2} \frac{G(I_3 - I_1)}{r^5} \cos \theta.$$

An application of Kepler's law, this time for the period of the satellite, reduces this result to

$$\frac{\dot{\phi}\tau}{2\pi} = -\frac{3}{2} \frac{I_3 - I_1}{mr^2} \cos \theta, \quad (5.96)$$

where m is Earth's mass. If Earth were a uniform sphere, then the principal moments of inertia would be

$$I_3 \sim I_1 = \frac{2}{5}mR^2,$$

with R Earth's radius. Because the core is much more dense than the outer layers, the moment of inertia is smaller, such that in fact*

$$I_3 = 0.331mR^2 \approx \frac{1}{3}mR^2.$$

The approximate precession is thus given by

$$\frac{\dot{\phi}\tau}{2\pi} = -\frac{1}{2} \frac{I_3 - I_1}{I_3} \left(\frac{R}{r}\right)^2 \cos \theta. \quad (5.97)$$

For a "close" satellite where r is very close to R , and the inclination of the satellite orbit to the equator is, say, 30° , Eq. (5.97) says that the plane of the orbit precesses completely around 2π in about 700 orbits of the satellite. Since the period of a close satellite is about $1\frac{1}{2}$ hours, complete rotation of the orbital plane occurs in a little over six weeks time. Clearly the effect is quite significant. We shall rederive the precession of the satellite orbit later on, when we discuss the subject of perturbation theory (cf. Section 12.3).

*The best values of I_3 are now obtained from observation of just such effects on satellite orbits.

5.9 ■ PRECESSION OF SYSTEMS OF CHARGES IN A MAGNETIC FIELD

The motion of systems of charged particles in magnetic fields does not normally involve rigid body motion. In a number of particular instances, the motion is however most elegantly discussed using the techniques developed here for rigid body motion. For this reason, and because of their importance in atomic and nuclear physics, a few examples will be given here.

The *magnetic moment* of a system of moving charges (relative to a particular origin) is defined as

$$\mathbf{M} = \frac{1}{2} q_i (\mathbf{r}_i \times \mathbf{v}_i) \rightarrow \frac{1}{2} \int dV \rho_e(\mathbf{r}) (\mathbf{r} \times \mathbf{v}). \quad (5.98)$$

Here the first expression is a sum over discrete particles with charge q_i , while the second is the corresponding generalization to a continuous distribution of charge density $\rho_e(\mathbf{r})$. The angular momentum of the system under corresponding conventions is

$$\mathbf{L} = m_i (\mathbf{r}_i \times \mathbf{v}_i) \rightarrow \int dV \rho_m(\mathbf{r}) (\mathbf{r} \times \mathbf{v}).$$

Both the magnetic moment and the angular momentum have a similar form. We shall restrict the discussion to situations in which \mathbf{M} is directly proportional to \mathbf{L} :

$$\mathbf{M} = \gamma \mathbf{L}, \quad (5.99)$$

most naturally by having a uniform q/m ratio for all particles or at all points in the continuous system. In such cases, the *gyromagnetic ratio* γ is given by

$$\gamma = \frac{q}{2m}, \quad (5.100)$$

but, with an eye to models of particle and atomic spin, γ will often be left unspecified. The forces and torques on a magnetic dipole may be considered as derived from a potential

$$V = -(\mathbf{M} \cdot \mathbf{B}). \quad (5.101)$$

It is implied along with Eq. (5.101) that the magnetic field is substantially constant over the system. Indeed, the picture applies best to a pointlike magnetic moment whose magnitude is not affected by the motion it undergoes—a picture appropriate to permanent magnets or systems on an atomic or small scale. With uniform \mathbf{B} , the potential depends only on the orientation of \mathbf{M} relative to \mathbf{B} ; no forces are exerted on the magnetic moment, but there is a torque

$$\mathbf{N} = \mathbf{M} \times \mathbf{B}. \quad (5.102)$$

(Compare with Eq. (5.81).) The time rate of change of the total angular momentum is equal to this torque, so that in view of Eq. (5.99) we can write

$$\frac{d\mathbf{L}}{dt} = \gamma \mathbf{L} \times \mathbf{B}. \quad (5.103)$$

But this is exactly the equation of motion for a vector of constant magnitude rotating in space about the direction of \mathbf{B} with an angular velocity $\omega = -\gamma \mathbf{B}$. The effect of a uniform magnetic field on a permanent magnetic dipole is to cause the angular momentum vector (and the magnetic moment) to *precess* uniformly.

For the classical gyromagnetic ratio, Eq. (5.100), the precession angular velocity is

$$\omega_l = -\frac{q\mathbf{B}}{2m}, \quad (5.104)$$

known as the *Larmor frequency*. For electrons q is negative, and the Larmor precession is counterclockwise around the direction of \mathbf{B} .

As a second example, consider a collection of moving charged particles, without restrictions on the nature of their motion, but assumed to all have the same q/m ratio, and to be in a region of uniform constant magnetic field. It will also be assumed that any interaction potential between particles depends only on the scalar distance between the particles. The Lagrangian for the system can be written (cf. Eq. (1.63))

$$L = \frac{1}{2}m_i v_i^2 + \frac{q}{m} m_i \mathbf{v}_i \cdot \mathbf{A}_i(\mathbf{r}_i) + V(|\mathbf{r}_i - \mathbf{r}_j|), \quad (5.105)$$

where the constant magnetic field \mathbf{B} is generated by a vector potential \mathbf{A} :

$$\mathbf{A} = \frac{1}{2}\mathbf{B} \times \mathbf{r}. \quad (5.106)$$

In terms of \mathbf{B} , the Lagrangian has the form (permuting dot and cross products)

$$L = \frac{1}{2}m_i v_i^2 + \frac{q\mathbf{B}}{2m} \cdot (\mathbf{r}_i \times m_i \mathbf{v}_i) + V(|\mathbf{r}_i - \mathbf{r}_j|). \quad (5.107)$$

The interaction term with the magnetic field can be variously written (cf. Eqs. (5.101) and (5.104))

$$\frac{q\mathbf{B} \cdot \mathbf{L}}{2m} = \mathbf{M} \cdot \mathbf{B} = -\omega_l \cdot (\mathbf{r}_i \times m_i \mathbf{v}_i). \quad (5.108)$$

Suppose now we express the Lagrangian in terms of coordinates relative to “primed” axes having a common origin with the original set, but rotating uniformly about the direction of \mathbf{B} with angular velocity ω_l . Distance vectors from the origin are unchanged as of course are scalar distances such as $|\mathbf{r}_i - \mathbf{r}_j|$. However, velocities relative to the new axes differ from the original velocities by the relation

$$\mathbf{v}_i = \mathbf{v}'_i + \omega_l \times \mathbf{r}_i.$$

The two terms in the Lagrangian affected by the transformation are

$$\frac{m_i v_i^2}{2} = \frac{m_i v'_i^2}{2} + m_i \mathbf{v}'_i \cdot (\boldsymbol{\omega}_l \times \mathbf{r}_i) + \frac{m_i}{2} (\boldsymbol{\omega}_l \times \mathbf{r}_i) \cdot (\boldsymbol{\omega}_l \times \mathbf{r}_i),$$

$$-\boldsymbol{\omega}_l \cdot \mathbf{r}_i \times m_i \mathbf{v}_i = -\boldsymbol{\omega}_l \cdot (\mathbf{r}_i \times m_i \mathbf{v}'_i) - \boldsymbol{\omega}_l \cdot (\mathbf{r}_i \times m_i (\boldsymbol{\omega}_l \times \mathbf{r}_i)).$$

By permuting dot and cross product, we can see that the terms linear in $\boldsymbol{\omega}_l$ and \mathbf{v}'_i are just equal and opposite and therefore cancel in the Lagrangian. A similar permutation in the terms quadratic in $\boldsymbol{\omega}_l$ show that they are of the same form and are related to the moment of inertia of the system about the axis defined by $\boldsymbol{\omega}_l$ (cf Section 5.3). The quadratic term in the Lagrangian can in fact be written as

$$-\frac{m_i}{2} (\boldsymbol{\omega}_l \times \mathbf{r}_i) \cdot (\boldsymbol{\omega}_l \times \mathbf{r}_i) = -\frac{1}{2} \boldsymbol{\omega}_l \cdot \mathbf{I} \cdot \boldsymbol{\omega}_l = -\frac{1}{2} I_l \omega_l^2, \quad (5.109)$$

where I_l denotes the moment of inertia about the axis of $\boldsymbol{\omega}_l$. In terms of coordinates in the rotating system, the Lagrangian thus has the simple form

$$L = \frac{1}{2} m_i \mathbf{v}'_i^2 + V(|\mathbf{r}_i - \mathbf{r}_j|) - \frac{1}{2} I_l \omega_l^2, \quad (5.110)$$

from which all linear terms in the magnetic field have disappeared.

We can get an idea of the relative magnitude of the quadratic term by considering a situation in which the motion of the system consists of a rotation with some frequency ω , e.g., an electron revolving around the atomic nucleus. Then for systems not too far from spherical symmetry, the kinetic energy is approximately $\frac{1}{2} I \omega^2$ (without subscripts on the moment of inertia) and the linear term in $\boldsymbol{\omega}_l$ is on the order of $\boldsymbol{\omega}_l \cdot \mathbf{L} \approx I \omega_l \omega$. Hence, the quadratic term in Eq. (5.110) is on the order of $(\omega_l/\omega)^2$ compared to the kinetic energy, and on the order of (ω_l/ω) relative to the linear term.

In most systems on the atomic or smaller scale, the natural frequencies are much larger than the Larmor frequency. Compare, for example, the frequency of a spectral line (which is a difference of natural frequencies) to the frequency shift in the simple Zeeman effect, which is proportional to the Larmor frequency. Thus, for such systems the motion in the rotating system is the same as in the laboratory system when there is no magnetic field. What we have is *Larmor's theorem*, which states that to first order in \mathbf{B} , the effect of a constant magnetic field on a classical system is to superimpose on its normal motion a uniform precession with angular frequency $\boldsymbol{\omega}_l$.

DERIVATIONS

- If \mathbf{R}_i is an antisymmetric matrix associated with the coordinates of the i th mass point of a system, with elements $R_{mn} = \epsilon_{mnj} x_j$ show that the matrix of the inertia tensor can be written as

$$\mathbf{I} = -m_i (\mathbf{R}_i)^2.$$

2. Show directly by vector manipulation that the definition of the moment of inertia as

$$I = m_i(\mathbf{r}_i \times \mathbf{n}) \cdot (\mathbf{r}_i \times \mathbf{n})$$

reduces to Eq. (5.18)

3. Prove that for a general rigid body motion about a fixed point, the time variation of the kinetic energy T is given by

$$\frac{dT}{dt} = \boldsymbol{\omega} \cdot \mathbf{N}.$$

4. Derive Euler's equations of motion, Eq. (5.39'), from the Lagrange equation of motion, in the form of Eq. (1.53), for the generalized coordinate ψ .

5. Equation (5.38) holds for the motions of systems that are not rigid, relative to a chosen rotating set of coordinates. For general nonrigid motion, if the rotating axes are chosen to coincide with the (instantaneous) principal axes of the continuous system, show that Eqs. (5.39) are to be replaced by

$$\frac{d(I_i\omega_i)}{dt} + \epsilon_{ijk}\omega_j\omega_k I_k - \omega_i \frac{dI_i}{dt} = N_i, \quad i = 1, 2, 3,$$

where

$$I_i = \int dV \rho(\mathbf{r})\epsilon_{ijk}x_i v'_k$$

with $\rho(\mathbf{r})$ the mass density at point \mathbf{r} , and \mathbf{v}' the velocity of the system point at \mathbf{r} relative to the rotating axes. These equations are sometimes known as the *Liouville equations* and have applications for discussing almost-rigid motion, such as that of Earth including the atmosphere and oceans.

6. (a) Show that the angular momentum of the torque-free symmetrical top rotates in the body coordinates about the symmetry axis with an angular frequency Ω . Show also that the symmetry axis rotates in space about the fixed direction of the angular momentum with the angular frequency

$$\dot{\phi} = \frac{I_3\omega_3}{I_1 \cos \theta},$$

where ϕ is the Euler angle of the line of nodes with respect to the angular momentum as the space z axis.

- (b) Using the results of Exercise 15, Chapter 4, show that $\boldsymbol{\omega}$ rotates in space about the angular momentum with the same frequency $\dot{\phi}$, but that the angle θ' between $\boldsymbol{\omega}$ and \mathbf{L} is given by

$$\sin \theta' = \frac{\Omega}{\dot{\phi}} \sin \theta'',$$

where θ'' is the inclination of $\boldsymbol{\omega}$ to the symmetry axis. Using the data given in Section 5.6, show therefore that Earth's rotation axis and the axis of angular momentum are never more than 1.5 cm apart on Earth's surface.

- (c) Show from parts (a) and (b) that the motion of the force-free symmetrical top can be described in terms of the rotation of a cone fixed in the body whose axis is the symmetry axis, rolling on a fixed cone in space whose axis is along the angular momentum. The angular velocity vector is along the line of contact of the two cones. Show that the same description follows immediately from the Poinsot construction in terms of the inertia ellipsoid.
7. For the general asymmetrical rigid body, verify analytically the stability theorem shown geometrically above on p. 204 by examining the solution of Euler's equations for small deviations from rotation about each of the principal axes. The direction of ω is assumed to differ so slightly from a principal axis that the component of ω along the axis can be taken as constant, while the product of components perpendicular to the axis can be neglected. Discuss the boundedness of the resultant motion for each of the three principal axes.
8. When the rigid body is not symmetrical, an analytic solution to Euler's equation for the torque-free motion cannot be given in terms of elementary functions. Show, however, that the conservation of energy and angular momentum can be used to obtain expressions for the body components of ω in terms of elliptic integrals.
9. Apply Euler's equations to the problem of the heavy symmetrical top, expressing ω , in terms of the Euler angles. Show that the two integrals of motion, Eqs. (5.53) and (5.54), can be obtained directly from Euler's equations in this form.
10. Obtain from Euler's equations of motion the condition (5.77) for the uniform precession of a symmetrical top in a gravitational field, by imposing the requirement that the motion be a uniform precession without nutation.
11. Show that the magnitude of the angular momentum for a heavy symmetrical top can be expressed as a function of θ and the constants of the motion only. Prove that as a result the angular momentum vector precesses uniformly only when there is uniform precession of the symmetry axis.
12. (a) Consider a primed set of axes coincident in origin with an inertial set of axes but rotating with respect to the inertial frame with fixed angular velocity ω_0 . If a system of mass points is subject to forces derived from a conservative potential V depending only on the distance to the origin, show that the Lagrangian for the system in terms of coordinates relative to the primed set can be written as

$$L = T' + \omega_0 \cdot \mathbf{L}' + \frac{1}{2} \omega_0 \cdot \mathbf{I}' \cdot \omega_0 - V,$$

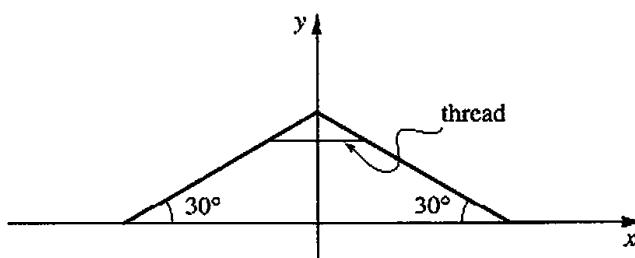
where primes indicate the quantities evaluated relative to the primed set of axes. What is the physical significance of each of the two additional terms?

- (b) suppose that ω_0 is in the $x'_2 x'_3$ plane, and that a symmetric top is constrained to move with its figure axis in the $x'_2 x'_1$ plane, so that only two Euler angles are needed to describe its orientation. If the body is mounted so that the center of mass is fixed at the origin and $V = 0$, show that the figure axis of the body oscillates about the x'_3 axis according to the plane-pendulum equation of motion and find the frequency of small oscillations. This illustrates the principle of the gyro compass.

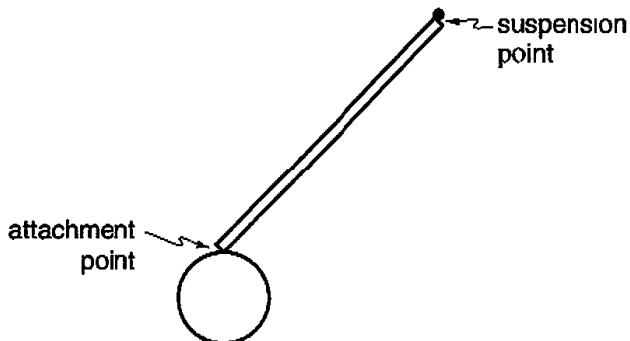
EXERCISES

13. Two thin rods each of mass m and length l are connected to an ideal (no friction) hinge and a horizontal thread. The system rests on a smooth surface as shown in the figure. At time $t = 0$, the thread is cut. Neglecting the mass of the hinge and the thread, and considering only motion in the xy plane

- (a) Find the speed at which the hinge hits the floor.
- (b) Find the time it takes for the hinge to hit the floor.



14. What is the height-to-diameter ratio of a right cylinder such that the inertia ellipsoid at the center of the cylinder is a sphere?
15. Find the principal moments of inertia about the center of mass of a flat rigid body in the shape of a 45° right triangle with uniform mass density. What are the principal axes?
16. Three equal mass points are located at $(a, 0, 0)$, $(0, a, 2a)$, $(0, 2a, a)$. Find the principal moments of inertia about the origin and a set of principal axes.
17. A uniform right circular cone of height h , half-angle α , and density ρ rolls on its side without slipping on a uniform horizontal plane in such a manner that it returns to its original position in a time τ . Find expressions for the kinetic energy and the components of the angular momentum of the cone.
18. (a) A bar of negligible weight and length l has equal mass points m at the two ends. The bar is made to rotate uniformly about an axis passing through the center of the bar and making an angle θ with the bar. From Euler's equations find the components along the principal axes of the bar of the torque driving the bar.
 (b) From the fundamental torque equation (1.26) find the components of the torque along axes fixed in space. Show that these components are consistent with those found in part (a).
19. A uniform bar of mass M and length $2l$ is suspended from one end by a spring of force constant k . The bar can swing freely only in one vertical plane, and the spring is constrained to move only in the vertical direction. Set up the equations of motion in the Lagrangian formulation.



20. A plane pendulum consists of a uniform rod of length l and negligible thickness with mass m , suspended in a vertical plane by one end. At the other end a uniform disk of radius a and mass M is attached so it can rotate freely in its own plane, which is the vertical plane. Set up the equations of motion in the Lagrangian formulation.
21. A compound pendulum consists of a rigid body in the shape of a lamina suspended in the vertical plane at a point other than the center of gravity. Compute the period for small oscillations in terms of the radius of gyration about the center of gravity and the separation of the point of suspension from the center of gravity. Show that if the pendulum has the same period for two points of suspension at unequal distances from the center of gravity, then the sum of these distances is equal to the length of the equivalent simple pendulum.
22. A uniform rod slides with its ends inside a smooth vertical circle. If the rod subtends an angle of 120° at the center of the circle, show that the equivalent simple pendulum has a length equal to the radius of the circle.
23. An automobile is started from rest with one of its doors initially at right angles. If the hinges of the door are toward the front of the car, the door will slam shut as the automobile picks up speed. Obtain a formula for the time needed for the door to close if the acceleration f is constant, the radius of gyration of the door about the axis of rotation is r_0 , and the center of mass is at a distance a from the hinges. Show that if f is 0.3 m/s^2 and the door is a uniform rectangle 1.2 m wide, the time will be approximately 3.04 s .
24. A wheel rolls down a flat inclined surface that makes an angle α with the horizontal. The wheel is constrained so that its plane is always perpendicular to the inclined plane, but it may rotate about the axis normal to the surface. Obtain the solution for the two-dimensional motion of the wheel, using Lagrange's equations and the method of undetermined multipliers.
25. (a) Express in terms of Euler's angles the constraint conditions for a uniform sphere rolling without slipping on a flat horizontal surface. Show that they are nonholonomic.
(b) Set up the Lagrangian equations for this problem by the method of Lagrange multipliers. Show that the translational and rotational parts of the kinetic energy are separately conserved. Are there any other constants of motion?
26. For the axially symmetric body precessing uniformly in the absence of torques, find analytical solutions for the Euler angles as a function of time.

27. In Section 5.6, the precession of Earth's axis of rotation about the pole was calculated on the basis that there were no torques acting on Earth. Section 5.8, on the other hand, showed that Earth is undergoing a forced precession due to the torques of the Sun and Moon. Actually both results are valid: The motion of the axis of rotation about the symmetry axis appears as the nutation of the Earth in the course of its forced precession. To prove this statement, calculate θ and $\dot{\phi}$ as a function of time for a heavy symmetrical top that is given an initial velocity $\dot{\phi}_0$, which is large compared with the net precession velocity $\beta/2a$, but which is small compared with ω_3 . Under these conditions, the bounding circles for the figure axis still lie close together, but the orbit of the figure axis appears as in Fig. 5.9(b), that is, shows large loops that move only slowly around the vertical. Show for this case that (5.71) remains valid but now

$$x_1 = \left(\frac{\beta}{a^2} - \frac{2\dot{\phi}_0}{a} \right) \sin^2 \theta_0.$$

From these values of θ and $\dot{\phi}$, obtain ω_1 and ω_2 , and show that for $\beta/2a$ small compared with $\dot{\phi}_0$, the vector ω precesses around the figure axis with an angular velocity

$$\Omega = \frac{I_3 - I_1}{I_1} \omega_3$$

in agreement with Eq. (5.49). Verify from the numbers given in Section 5.6 that $\dot{\phi}_0$ corresponds to a period of about 1600 years, so that $\dot{\phi}_0$ is certainly small compared with the daily rotation and is sufficiently large compared with $\beta/2a$, which corresponds to the precession period of 26,000 years.

28. Suppose that in a symmetrical top each element of mass has a proportionate charge associated with it, so that the e/m ratio is constant—the so-called charged symmetric top. If such a body rotates in a uniform magnetic field the Lagrangian, from (5.108), is

$$L = T - \omega_l \cdot \mathbf{L}.$$

Show that T is a constant (which is a manifestation of the property of the Lorentz force that a magnetic field does no work on a moving charge) and find the other constants of motion. Under the assumption that ω_l is much smaller than the initial rotational velocity about the figure axis, obtain expressions for the frequencies and amplitudes of nutation and precession. From where do the kinetic energies of nutation and precession come?

29. A homogeneous cube of sides l is initially at rest in unstable equilibrium with one edge in contact with a horizontal plane. The cube is given a small angular displacement and allowed to fall. What is the angular velocity of the cube when one face contacts the plane if:
- the edge in contact with the plane cannot slide?
 - the plane is frictionless so the edge can slide?
30. A door is constructed of a thin homogeneous material. It has a height of 2 m and a width of 0.9 m. If the door is opened by 90° and released from rest, it is observed that the door closes itself in 3 s. Assuming that the hinges are frictionless, what angle do these hinges make with the vertical?

CHAPTER

6

Oscillations

A class of mechanical motions that can best be treated in the Lagrangian formulation is that of the oscillations of a system about positions of equilibrium. The theory of small oscillations finds widespread physical applications in acoustics, molecular spectra, vibrations of mechanisms, and coupled electrical circuits. If the deviations of the system from stable equilibrium conditions are small enough, the motion can generally be described as that of a system of coupled linear harmonic oscillators. It will be assumed the reader is familiar with the properties of a simple harmonic oscillator of one degree of freedom, both in free and forced oscillation, with and without damping. Here the emphasis will be on methods appropriate to discrete systems with more than one degree of freedom. As will be seen, the mathematical techniques required turn out to be very similar to those employed in studying rigid body motion, although the mechanical systems considered need not involve rigid bodies at all. Analogous treatments of oscillations about stable motions can also be developed, but these are most easily done in the Hamiltonian formulation presented in Chapter 8.

6.1 ■ FORMULATION OF THE PROBLEM

We consider conservative systems in which the potential energy is a function of position only. It will be assumed that the transformation equations defining the generalized coordinates of the system, q_1, \dots, q_n , do not involve the time explicitly. Thus, time-dependent constraints are to be excluded. The system is said to be in *equilibrium* when the generalized forces acting on the system vanish:

$$Q_i = - \left(\frac{\partial V}{\partial q_i} \right)_0 = 0. \quad (6.1)$$

The potential energy therefore has an extremum at the equilibrium configuration of the system, $q_{01}, q_{02}, \dots, q_{0n}$. If the configuration is initially at the equilibrium position, with zero initial velocities \dot{q}_n , then the system will continue in equilibrium indefinitely. Examples of the equilibrium of mechanical systems are legion—a pendulum at rest, a suspension galvanometer at its zero position, an egg standing on end.

An equilibrium position is classified as *stable* if a small disturbance of the system from equilibrium results only in small bounded motion about the rest po-

sition. The equilibrium is *unstable* if an infinitesimal disturbance eventually produces unbounded motion. A pendulum at rest is in stable equilibrium, but the egg standing on end is an obvious illustration of unstable equilibrium. It can be readily seen that when the extremum of V is a minimum the equilibrium must be stable. Suppose the system is disturbed from the equilibrium by an increase in energy dE above the equilibrium energy. If V is a minimum at equilibrium, any deviation from this position will produce an increase in V . By the conservation of energy, the velocities must then decrease and eventually come to zero, indicating bound motion. On the other hand, if V decreases as the result of some departure from equilibrium, the kinetic energy and the velocities increase indefinitely, corresponding to unstable motion. The same conclusion may be arrived at graphically by examining the shape of the potential energy curve, as shown symbolically in Fig. 6.1. A more rigorous mathematical proof that stable equilibrium requires a minimum in V will be given in the course of the discussion.

We shall be interested in the motion of the system within the immediate neighborhood of a configuration of stable equilibrium. Since the departures from equilibrium are too small, all functions may be expanded in a Taylor series about the equilibrium, retaining only the lowest-order terms. The deviations of the generalized coordinates from equilibrium will be denoted by η_i :

$$q_i = q_{0i} + \eta_i, \quad (6.2)$$

and these may be taken as the new generalized coordinates of the motion. Expanding the potential energy about q_{0i} , we obtain

$$V(q_1, \dots, q_n) = V(q_{01}, \dots, q_{0n}) + \left(\frac{\partial V}{\partial q_i} \right)_0 \eta_i + \frac{1}{2} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 \eta_i \eta_j + \dots \quad (6.3)$$

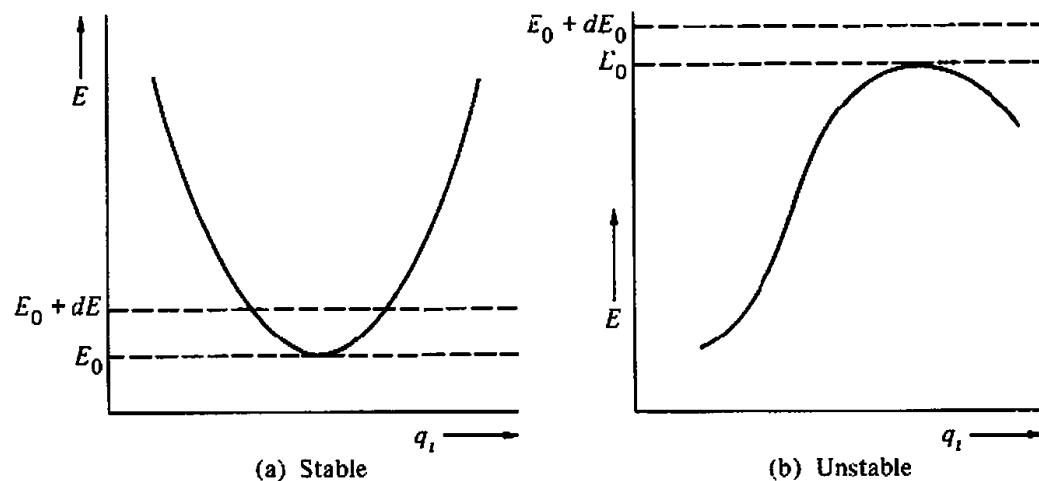


FIGURE 6.1 Shape of the potential energy curve at equilibrium.

where the summation convention has been invoked, as usual. The terms linear in η_i vanish automatically in consequence of the equilibrium conditions (6.1). The first term in the series is the potential energy of the equilibrium position, and by shifting the arbitrary zero of potential to coincide with the equilibrium potential, this term may also be made to vanish. We are therefore left with the quadratic terms as the first approximation to V :

$$V = \frac{1}{2} \left(\frac{\partial^2 V}{\partial q_i \partial q_j} \right)_0 \eta_i \eta_j = \frac{1}{2} V_{ij} \eta_i \eta_j, \quad (6.4)$$

where the second derivatives of V have been designated by the constants V_{ij} depending only upon the equilibrium values of the q_i 's. It is obvious from their definition that the V_{ij} 's are symmetrical, that is, that $V_{ij} = V_{ji}$. The V_{ij} coefficients can vanish under a variety of circumstances. Thus, the potential can simply be independent of a particular coordinate, so that equilibrium occurs at any arbitrary value of that coordinate. We speak of such cases as *neutral* or *indifferent* equilibrium. It may also happen, for example, that the potential behaves like a quadratic at that point, again causing one or more of the V_{ij} 's to vanish. Either situation calls for special treatment in the mathematical discussion that follows.

A similar series expansion can be obtained for the kinetic energy. Since the generalized coordinates do not involve the time explicitly, the kinetic energy is a homogeneous quadratic function of the velocities (cf. Eq. (1.71)):

$$T = \frac{1}{2} m_{ij} \dot{q}_i \dot{q}_j = \frac{1}{2} m_{ij} \dot{\eta}_i \dot{\eta}_j. \quad (6.5)$$

The coefficients m_{ij} are in general functions of the coordinates q_k , but they may be expanded in a Taylor series about the equilibrium configuration:

$$m_{ij}(q_1, \dots, q_n) = m_{ij}(q_{01}, \dots, q_{0n}) + \left(\frac{\partial m_{ij}}{\partial q_k} \right)_0 \eta_k + \dots$$

As Eq. (6.5) is already quadratic in the $\dot{\eta}_i$'s, the lowest nonvanishing approximation to T is obtained by dropping all but the first term in the expansions of m_{ij} . Denoting the constant values of the m_{ij} functions at equilibrium by T_{ij} , we can therefore write the kinetic energy as

$$T = \frac{1}{2} T_{ij} \dot{\eta}_i \dot{\eta}_j. \quad (6.6)$$

It is again obvious that the constants T_{ij} must be symmetric, since the individual terms in Eq. (6.6) are unaffected by an interchange of indices. From Eqs. (6.4) and (6.6), the Lagrangian is given by

$$L = \frac{1}{2} (T_{ij} \dot{\eta}_i \dot{\eta}_j - V_{ij} \eta_i \eta_j). \quad (6.7)$$

Taking the η 's as the general coordinates, the Lagrangian of Eq. (6.7) leads to the following n equations of motion:

$$T_{ij} \ddot{\eta}_{ij} + V_{ij} \eta_j = 0. \quad (6.8)$$

where explicit use has been made of the symmetry property of the V_{ij} and T_{ij} coefficients. Each of Eqs. (6.8) will involve, in general, all of the coordinates η_i , and it is this set of simultaneous differential equations that must be solved to obtain the motion near the equilibrium.

In almost all cases of interest, the kinetic energy term can be easily written so as to have no cross terms.* This corresponds to the Lagrangian

$$L = \frac{1}{2}(T_i \dot{\eta}_i^2 - V_{ij} \eta_i \eta_j), \quad (6.9)$$

which generates the following equations of motion

$$T_i \ddot{\eta}_i + V_{ij} \eta_j = 0. \quad (\text{no sum over } i) \quad (6.10)$$

6.2 ■ THE EIGENVALUE EQUATION AND THE PRINCIPAL AXIS TRANSFORMATION

The equations of motion (6.8) are linear differential equations with constant coefficients, of a form familiar from electrical circuit theory. We are therefore led to try an oscillatory solution of the form

$$\eta_i = Ca_i e^{-i\omega t}. \quad (6.11)$$

Here Ca_i gives the complex amplitude of the oscillation for each coordinate η_i , the factor C being introduced for convenience as a scale factor, the same for all coordinates. It is understood of course that it is the real part of Eq. (6.9) that is to correspond to the actual motion. Substitution of the trial solution (6.9) into the equations of motion leads to the following equations for the amplitude factors:

$$(V_{ij}a_j - \omega^2 T_{ij}a_j) = 0. \quad (6.12)$$

Equations (6.12) constitute n linear homogeneous equations for the a_i 's, and consequently can have a nontrivial solution only if the determinant of the coefficients vanishes:

*Mathematically, we could go even further when the coordinates are Cartesian and making the $T_{ij} = \delta_{ij}$ by rescaling the coordinates. Such coordinates are called *mass-weighted coordinates* since they are generated by dividing the coordinates by the square root of the mass. This transforms the kinetic energy to the form

$$T = \frac{\dot{\eta}_i \dot{\eta}_i}{2}.$$

This reduces the problem to the eigenvalue problem of Chapters 4 and 5, only in n dimensions instead of three, however, the mathematical simplification can obscure the physics, since each coordinate can have a different characteristic scale.

$$\begin{vmatrix} V_{11} - \omega^2 T_{11} & V_{12} - \omega^2 T_{12} & \dots \\ V_{21} - \omega^2 T_{21} & V_{22} - \omega^2 T_{22} & \\ V_{31} - \omega^2 T_{31} & & \\ \vdots & & \end{vmatrix} = 0. \quad (6.13)$$

This determinantal condition is in effect an algebraic equation of the n th degree for ω^2 , and the roots of the determinant provide the frequencies for which Eq. (6.11) represents a correct solution to the equations of motion. For each of these values of ω^2 , Eqs. (6.12) may be solved for the amplitudes of a_i , or more precisely, for $n - 1$ of the amplitudes in terms of the remaining a_i .

Equations (6.12) represent a type of eigenvalue equation, for writing T_{ij} as an element of the matrix \mathbf{T} , the equations may be written

$$\mathbf{V}\mathbf{a} = \lambda \mathbf{T}\mathbf{a}. \quad (6.14)$$

Here the effect of \mathbf{V} on the eigenvector \mathbf{a} is not merely to reproduce the vector times the factor λ , as in the ordinary eigenvalue problem. Instead, the eigenvector is such that \mathbf{V} acting on \mathbf{a} produces a multiple of the result of \mathbf{T} acting on \mathbf{a} . We shall show that the eigenvalues λ for which Eq. (6.14) can be satisfied are all real in consequence of the symmetric and reality properties of \mathbf{T} and \mathbf{V} , and, in fact, must be positive. It will also be shown that the eigenvectors \mathbf{a} are orthogonal—in a sense. In addition, the matrix of the eigenvectors, \mathbf{A} , diagonalizes both \mathbf{T} and \mathbf{V} , the former to the unit matrix $\mathbf{1}$ and the latter to a matrix whose diagonal elements are the eigenvalues λ . Most importantly it is necessary to show that \mathbf{a} and λ are real.

Proceeding as in Section 5.4, let \mathbf{a}_k be a column matrix representing the k th eigenvector, satisfying the eigenvalue equation*

$$\mathbf{V}\mathbf{a}_k = \lambda_k \mathbf{T}\mathbf{a}_k. \quad (6.15)$$

Assume now that the only solution to Eq. (6.15) involves complex λ and \mathbf{a}_k . The adjoint equation, i.e., the transposed complex conjugate equation, for λ_l has the form

$$\mathbf{a}_l^\dagger \mathbf{V} = \lambda_l^* \mathbf{a}_l^\dagger \mathbf{T}. \quad (6.16)$$

Here \mathbf{a}_l^\dagger stands for the adjoint vector—the complex conjugate row matrix—and explicit use has been made of the fact that the \mathbf{V} and \mathbf{T} matrices are real and symmetric. Multiply Eq. (6.16) from the right by \mathbf{a}_k and subtract the result of the similar product of Eq. (6.15) from the left with \mathbf{a}_l^\dagger . The left-hand side of the difference equation vanishes, leaving only

$$0 = (\lambda_k - \lambda_l^*) \mathbf{a}_l^\dagger \mathbf{T} \mathbf{a}_k. \quad (6.17)$$

*It hardly need be added that there is no summation over k in Eq. (6.15). Indeed, in this chapter the summation convention will apply only to the components of matrices or tensors (of any rank) and not to the matrices and tensors themselves.

When $l = k$, Eq. (6.17) becomes

$$(\lambda_k - \lambda_k^*) \mathbf{a}_k^\dagger \mathbf{T} \mathbf{a}_k = 0. \quad (6.18)$$

That the matrix product in Eq. (6.18) is real can be shown immediately by taking its complex conjugate and using the symmetry property of \mathbf{T} . However, we want to prove that the matrix product is not only real but is positive definite. For this purpose, separate \mathbf{a}_k into its real and imaginary components,

$$\mathbf{a}_k = \boldsymbol{\alpha}_k + i \boldsymbol{\beta}_k.$$

The matrix product can then be written as

$$\mathbf{a}_k^\dagger \mathbf{T} \mathbf{a}_k = \tilde{\boldsymbol{\alpha}}_k^\dagger \mathbf{T} \boldsymbol{\alpha}_k + \tilde{\boldsymbol{\beta}}_k^\dagger \mathbf{T} \boldsymbol{\beta}_k + i(\tilde{\boldsymbol{\alpha}}_k^\dagger \mathbf{T} \boldsymbol{\beta}_k - \tilde{\boldsymbol{\beta}}_k^\dagger \mathbf{T} \boldsymbol{\alpha}_k). \quad (6.19)$$

The imaginary term vanishes by virtue of the symmetry of \mathbf{T} and therefore, as noted earlier, the matrix product is real. Further, the kinetic energy in Eq. (6.6) can be rewritten in terms of a column matrix $\dot{\boldsymbol{\eta}}$ as

$$T = \frac{1}{2} \tilde{\boldsymbol{\eta}}^\dagger \mathbf{T} \dot{\boldsymbol{\eta}}. \quad (6.20)$$

Hence, the first two terms in Eq. (6.18) are twice the kinetic energies when the velocity matrix $\dot{\boldsymbol{\eta}}_k$ has the values $\boldsymbol{\alpha}_k$ and $\boldsymbol{\beta}_k$, respectively. Now, a kinetic energy by its physical nature must be positive definite for real velocities, and therefore the matrix product in Eq. (6.18) cannot be zero. It follows that the eigenvalues λ_k must be real.

Since the eigenvalues are real, the ratios of the eigenvector components $a_{j,k}$ determined by Eqs. (6.15) must all be real. There is still some indeterminateness of course since the value of a particular one of the $a_{j,k}$'s can still be chosen at will without violating Eqs. (6.15). We can require however that this component shall be real, and the reality of λ_k then ensures the reality of all the other components. (Any complex phase factor in the amplitude of the oscillation will be thrown into the factor C , Eq. (6.11).) Multiply now Eq. (6.15) by $\tilde{\mathbf{a}}_k$ from the left and solve for λ_k :

$$\lambda_k = \frac{\tilde{\mathbf{a}}_k^\dagger \mathbf{V} \mathbf{a}_k}{\tilde{\mathbf{a}}_k^\dagger \mathbf{T} \mathbf{a}_k}. \quad (6.21)$$

The denominator of this expression is equal to twice the kinetic energy for velocities $a_{i,k}$ and since the eigenvectors are all real, the sum must be positive definite. Similarly, the numerator is the potential energy for coordinates $a_{i,k}$, and the condition that V be a minimum at equilibrium requires that the sum must be positive or zero. Neither numerator nor denominator can be negative, and the denominator cannot be zero, hence λ is always finite and positive. (It may however be zero.) Recall that λ stands for ω^2 , so that positive λ corresponds to real frequencies of oscillation. Were the potential not a local minimum, the numerator in Eq. (6.21)

might be negative, giving rise to imaginary frequencies that would produce an unbounded exponential increase of the η_i with time. Such motion would obviously be unstable, and we have here the promised mathematical proof that a minimum of the potential is required for stable motion.

Let us return for the moment to Eq. (6.17) which, in view of the reality of the eigenvalues and eigenvectors, can be written

$$(\lambda_k - \lambda_l) \tilde{\mathbf{a}}_l \mathbf{T} \mathbf{a}_k = 0. \quad (6.17')$$

If all the roots of the secular equation are distinct, then Eq. (6.17') can hold only if the matrix product vanishes for l not equal to k :

$$\tilde{\mathbf{a}}_l \mathbf{T} \mathbf{a}_k = 0, \quad l \neq k. \quad (6.22a)$$

It has been remarked several times that the values of the a_{jk} 's are not completely fixed by the eigenvalue equations (6.12). We can remove this indeterminacy by requiring further that

$$\tilde{\mathbf{a}}_k \mathbf{T} \mathbf{a}_k = 1. \quad (6.22b)$$

There are n such equations (6.22), and they uniquely fix the one arbitrary component of each of the n eigenvectors \mathbf{a}_k .* If we form all the eigenvectors \mathbf{a}_k into a square matrix \mathbf{A} with components a_{jk} (cf. Section 4.6), then the two equations (6.22a and b) can be combined into one matrix equation:

$$\tilde{\mathbf{A}} \mathbf{T} \mathbf{A} = \mathbf{1}. \quad (6.23)$$

When two or more of the roots are repeated, the argument leading to Eq. (6.22a) falls through for $\lambda_l = \lambda_k$. We shall reserve a discussion of this exceptional case of *degeneracy* for a later time. For the present, suffice it to state that a set of a_{jk} coefficients can always be found that satisfies both the eigenvalue conditions Eqs. (6.10), and Eq. (6.22a), so that Eq. (6.23) always holds.

In Chapter 4, the *similarity* transformation of a matrix \mathbf{C} by a matrix \mathbf{B} was defined by the equation (cf. Eq. (4.41)):

$$\mathbf{C}' = \mathbf{B} \mathbf{C} \mathbf{B}^{-1}.$$

*Equation (6.22b) may be put in a form that explicitly shows that it suffices to remove the indeterminacy in the a_{jk} 's. Suppose it is the magnitude of a_{1k} that is to be evaluated; the ratio of all the other a_{jk} 's to a_{1k} is obtained from Eqs. (6.12). Then Eq. (6.22b) can be written as

$$\sum_{i,j} T_{ij} \frac{a_{ik}}{a_{1k}} \frac{a_{jk}}{a_{1k}} = \frac{1}{a_{1k}^2}.$$

The left-hand side is completely determined from the eigenvalue equations and may be evaluated directly to provide a_{1k} .

We now introduce the related concept of the *congruence* transformation of \mathbf{C} by \mathbf{A} according to the relation

$$\mathbf{C}' = \tilde{\mathbf{A}}\mathbf{C}\mathbf{A}. \quad (6.24)$$

If \mathbf{A} is orthogonal, so that $\tilde{\mathbf{A}} = \mathbf{A}^{-1}$, there is no essential difference between the two types of transformation (as may be seen by denoting \mathbf{A}^{-1} by the matrix \mathbf{B}). Equation (6.23) can therefore be read as the statement that \mathbf{A} transforms \mathbf{T} by a congruence transformation into a diagonal matrix, in particular into the unit matrix.

If a diagonal matrix $\boldsymbol{\lambda}$ with elements $\lambda_{ik} = \lambda_k \delta_{ik}$ is introduced, the eigenvalue equations (6.15) may be written

$$V_{ij}a_{jk} = T_{ij}a_{jl}\lambda_{lk},$$

which becomes in matrix notation

$$\mathbf{V}\mathbf{A} = \mathbf{T}\mathbf{A}\boldsymbol{\lambda}. \quad (6.25)$$

Multiplying by $\tilde{\mathbf{A}}$ from the left, Eq. (6.25) takes the form

$$\tilde{\mathbf{A}}\mathbf{V}\mathbf{A} = \tilde{\mathbf{A}}\mathbf{T}\mathbf{A}\boldsymbol{\lambda},$$

which by Eq. (6.23) reduces to

$$\tilde{\mathbf{A}}\mathbf{V}\mathbf{A} = \boldsymbol{\lambda}. \quad (6.26)$$

Our final equation (6.26) states that a congruence transformation of \mathbf{V} by \mathbf{A} changes it into a diagonal matrix whose elements are the eigenvalues λ_k . Eq. (6.26) has solutions

$$|\mathbf{V} - \boldsymbol{\lambda}\mathbf{1}| = 0. \quad (6.26')$$

In summary we can use normalized Cartesian coordinates so that $T_{ij} = \delta_{ij}$ which reduces the physics to solving

$$\tilde{\mathbf{A}}\mathbf{A} = \mathbf{1} \quad (4.36) \quad \text{and} \quad \tilde{\mathbf{A}}\mathbf{V}\mathbf{A} = \mathbf{V}_{\text{diagonal}} \quad (6.26),$$

or we may choose more general coordinates where $T_{ij} \neq \delta_{ij}$, even allowing $T_{ij} = T_{ji} \neq 0$ for $i \neq j$, and use

$$\tilde{\mathbf{A}}\mathbf{T}\mathbf{A} = \mathbf{1} \quad (6.23) \quad \text{and} \quad \tilde{\mathbf{A}}\mathbf{V}\mathbf{A} = \mathbf{V}_{\text{diagonal}} \quad (6.26),$$

to solve the general problem.

As an example, we consider a particle of mass m with two degrees of freedom (x_1, x_2) that obeys the Lagrangian (cf. Eq. (6.9))

$$L = \frac{1}{2}m(\dot{x}_1^2 + \dot{x}_2^2) - \frac{1}{2}V_{ij}x_i x_j$$

where the V_{ij} are constants. The congruence transformation (6.26) has solutions only when Eq. (6.26') is satisfied, so

$$\begin{vmatrix} V_{11} - \lambda & V_{12} \\ V_{21} & V_{22} - \lambda \end{vmatrix} = 0$$

This equation has two solutions:

$$\begin{aligned} \lambda_1 &= \frac{1}{2} \left(V_{11} + V_{22} + \sqrt{(V_{11} - V_{22})^2 + 4V_{12}V_{21}} \right) \\ \lambda_2 &= \frac{1}{2} \left(V_{11} + V_{22} - \sqrt{(V_{11} - V_{22})^2 + 4V_{12}V_{21}} \right). \end{aligned}$$

Associated with the eigenvalues λ_i are the eigenvectors a_{ij} that satisfy

$$a_{ij}(V_{ij} - \lambda_i \delta_{ij}) = 0 \quad \text{and} \quad a_{i1}^2 + a_{i2}^2 = 1 \quad (\text{no sum on } i)$$

We consider two limiting cases. The first case assumes $V_{11} > V_{22} > 0$ and $0 \neq V_{21} = V_{12} \ll (V_{11} - V_{22})$. We write the small quantity $\delta = [V_{12}/(V_{11} - V_{22})]$ then, to first order in δ , the eigenvalues are

$$\begin{aligned} \lambda_1 &= V_{11} + V_{12}\delta \\ \lambda_2 &= V_{22} - V_{12}\delta \end{aligned} \tag{6.27}$$

whose eigenvectors are, to lowest order in δ ,

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \begin{bmatrix} 1 - \frac{\delta^2}{2} & -\delta + \frac{\delta^3}{2} \\ \delta - \frac{\delta^3}{2} & 1 - \frac{\delta^2}{2} \end{bmatrix}. \tag{6.28}$$

These correspond to the relations

$$a_{11} = a_{22} \quad \text{and} \quad a_{12} = -a_{21}.$$

The other limiting case assumes $V_{12} > V_{22} > 0$ and $(V_{11} - V_{22}) \ll V_{12} = V_{21}$. We now write $\varepsilon = (V_{11} - V_{22})/8V_{12}$, which is a small quantity. To first order in ε the eigenvalues are

$$\begin{aligned} \lambda_1 &= \frac{1}{2}(V_{11} + V_{22}) + V_{12} + (V_{11} - V_{22})\varepsilon \\ \lambda_2 &= \frac{1}{2}(V_{11} + V_{22}) - V_{12} - (V_{11} - V_{22})\varepsilon \end{aligned} \tag{6.29}$$

whose eigenvectors are, to lowest order in ε ,

$$\mathbf{a} = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}}(1 + 2\varepsilon) & -\frac{1}{\sqrt{2}}(1 - 2\varepsilon) \\ \frac{1}{\sqrt{2}}(1 - 2\varepsilon) & \frac{1}{\sqrt{2}}(1 + 2\varepsilon) \end{bmatrix}. \tag{6.30}$$

The relations among the components of the eigenvectors are different than in the previous example. Here $a_{12} = -a_{21}$ is slightly less than $1/\sqrt{2}$ while $a_{11} = a_{22}$ is slightly greater than $1/\sqrt{2}$.

The preceding approximations looked at the behavior of the eigenvalues and eigenvectors in limiting cases. The qualitative changes in these quantities as a function of $V_{12}/(V_{11} - V_{22})$ from zero to three are shown in Fig. 6.2. We shall return to this example after considering the general problem of multiple roots of the eigenvalue equation (6.26').

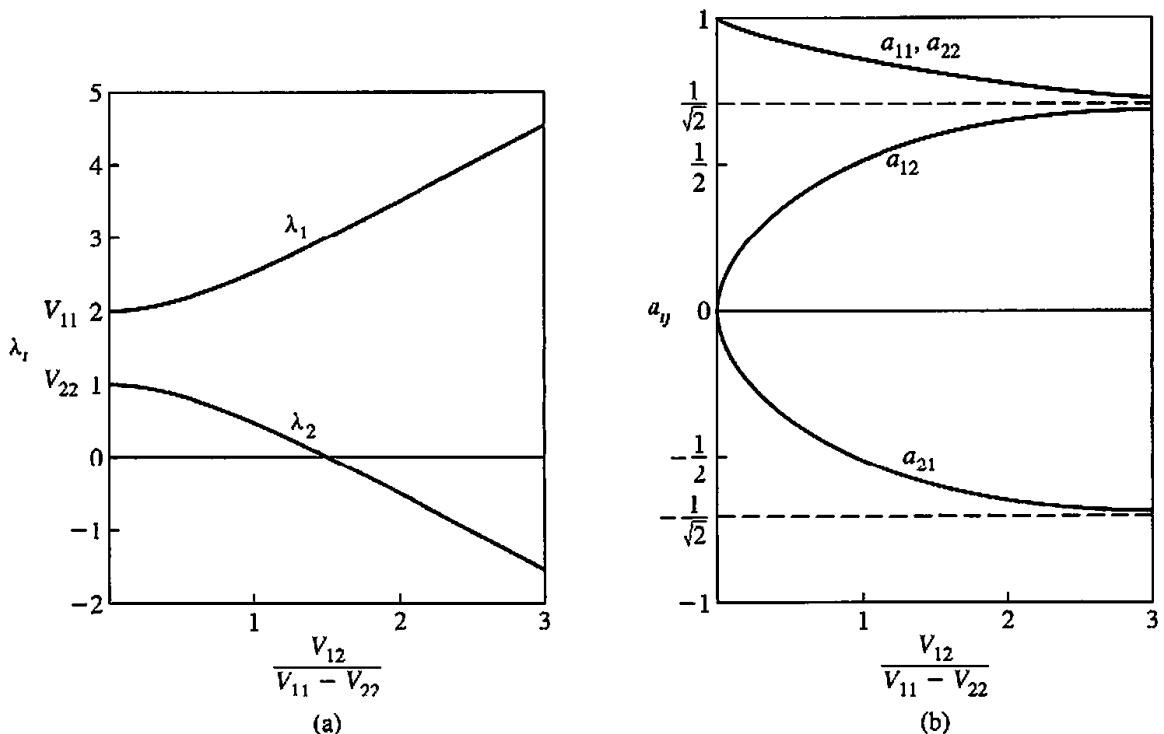


FIGURE 6.2 Behavior of the (a) eigenvalues and (b) eigenvector components as the energy ratio $\frac{V_{12}}{V_{11} - V_{22}}$ changes from 0 to 3.

It remains only to consider the case of multiple roots to the secular equation, a situation that is more annoying in the mathematical theory than it is in practice. If one or more of the roots is repeated, it is found that the number of independent equations among the eigenvalues is insufficient to determine even the ratio of the eigenvector components. Thus, if the eigenvalue λ is a double root, any two of the components a_i may be chosen arbitrarily, the rest being fixed by the eigenvalue equations.

In general, any pair of eigenvectors randomly chosen out of the infinite set of allowed vectors will not be orthogonal. Nevertheless, it is always possible to construct a pair of allowed vectors that are orthogonal, and these can be used to form the orthogonal matrix \mathbf{A} . Consider for simplicity the procedure to be followed for a double root. Let \mathbf{a}'_k and \mathbf{a}'_l be any two allowable eigenvectors for a given

double root λ , which have been normalized so as to satisfy Eq. (6.22b). Any linear combination of \mathbf{a}'_k and \mathbf{a}'_l will also be an eigenvector for the root λ . We therefore seek to construct a vector \mathbf{a}_l ,

$$\mathbf{a}_l = c_1 \mathbf{a}'_k + c_2 \mathbf{a}'_l, \quad (6.31)$$

where c_1 and c_2 are constants such that \mathbf{a}_l is orthogonal to \mathbf{a}'_k . The orthogonality condition, Eq. (6.22a), then requires that

$$\tilde{\mathbf{a}}_l^T \mathbf{T} \mathbf{a}'_k = c_1 + c_2 \tilde{\mathbf{a}}_l^T \mathbf{T} \mathbf{a}'_l = 0,$$

where use has been made of the normalization of \mathbf{a}'_k . It therefore follows that the ratio of c_1 to c_2 must be given by

$$\frac{c_1}{c_2} = -\tilde{\mathbf{a}}_l^T \mathbf{T} \mathbf{a}'_k \equiv -\tau_l. \quad (6.32)$$

We can illustrate these ideas by again considering our two-dimensional example given by Eqs (6.27) through (6.30). The two limiting cases of the off-diagonal potential term V_{12} , being much less than and much greater than the difference factor ($V_{11} - V_{22}$), provide an excellent example of the problems introduced by degeneracy. When

$$V_{11} = V_{22} = V_0, \quad V_{12} = 0,$$

the two eigenvalues become the same, $\lambda_1 = \lambda_2 = V_0$.

If the limit is taken by letting $V_{12} \rightarrow 0$ first and then taking the limit ($V_{11} \rightarrow V_{22}$), the eigenvectors in Eqs. (6.28) become

$$\mathbf{a}_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}. \quad (6.33)$$

If the limit is taken in the reverse order, Eqs. (6.30) give

$$\mathbf{b}_1 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix} \quad \text{and} \quad \mathbf{b}_2 = \begin{pmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{pmatrix}, \quad (6.34)$$

where \mathbf{b} is used for the eigenvectors in Eqs. (6.34) to avoid confusion with the eigenvectors in Eqs. (6.33). Each of the eigenvectors in (6.33) and (6.34) are linear combinations of the other set of eigenvectors. For example,

$$\mathbf{b}_1 = \frac{1}{\sqrt{2}}(\mathbf{a}_1 + \mathbf{a}_2), \quad \text{and} \quad \mathbf{b}_2 = \frac{1}{\sqrt{2}}(\mathbf{a}_2 - \mathbf{a}_1),$$

so either set of eigenvectors is a linear combination of the other, as was discussed in this section. These results obviously generalize to the infinite set

$$\mathbf{a}_1 = \begin{pmatrix} a \\ b \end{pmatrix} \quad \text{and} \quad \mathbf{a}_2 = \begin{pmatrix} -b \\ a \end{pmatrix},$$

where a and b are any pairs of numbers that satisfy

$$a^2 + b^2 = 1.$$

This shows that there is an infinite set of possible eigenvectors in the case of degeneracy.

There is another way to consider the significance of these results. The approximate eigenvectors in Eqs. (6.28) are for the case where the main potential energy terms are V_{11} and V_{22} , which are at diagonal positions, and the V_{12} are in the off-diagonal positions. If we take the eigenvectors of Eq. (6.30) in the limit $\varepsilon \rightarrow 0$ and let the eigenvectors of Eqs. (6.30) transform V as $V' = \tilde{\mathbf{A}}\mathbf{V}\mathbf{A}$, we obtain the transformed potential energy tensor

$$\mathbf{V}' = \begin{pmatrix} \frac{1}{2}(V_{11} + V_{22}) + V_{12} & \frac{1}{2}(V_{11} - V_{22}) \\ \frac{1}{2}(V_{11} - V_{22}) & \frac{1}{2}(V_{11} + V_{22}) - V_{12} \end{pmatrix}$$

in which the difference term ($V_{11} - V_{22}$) is off-diagonal. Thus, the set of eigenvectors given by Eqs. (6.30) are for the physical situation in which the small energy term ($V_{11} - V_{22}$) is off-diagonal.

Returning to the main discussion, the requirement that \mathbf{a}_l of Eq. (6.32) be normalized provides another condition on the two coefficients, which in terms of τ_l defined by Eq. (6.32) takes the form

$$\tilde{\mathbf{a}}_l^T \mathbf{T} \mathbf{a}_l = 1 = c_1^2 + c_2^2 + 2c_1c_2\tau_l.$$

Together the two equations fix the coefficients c_1 and c_2 , and therefore the vector \mathbf{a}_l . Both \mathbf{a}_l and $\mathbf{a}_k \equiv \mathbf{a}'_k$ are automatically orthogonal to the eigenvectors of the other distinct eigenvalues, for then the argument based on Eq. (6.17') remains valid. Hence, we have a set of n eigenvectors \mathbf{a}_j whose components form the matrix \mathbf{A} satisfying Eq. (6.23).

A similar procedure is followed for a root of higher multiplicity. If λ is an m -fold root, then orthogonal normalized eigenvectors are formed out of linear combinations of any of the m corresponding eigenvectors $\mathbf{a}'_1, \dots, \mathbf{a}'_m$. The first of the "orthonormal" eigenvectors \mathbf{a}_1 is then chosen as a multiple of \mathbf{a}'_1 ; \mathbf{a}_2 is taken as a linear combination of \mathbf{a}'_1 and \mathbf{a}'_2 ; and so on. In this manner, the number of constants to be determined is equal to the sum of the first m integers, or $\frac{1}{2}m(m-1)$. The normalization requirements provide m conditions, while there are $\frac{1}{2}m(m-1)$ orthogonality conditions, and together these are just enough to fix the constants uniquely.

This process of constructing orthogonalized eigenvectors in the case of multiple roots is completely analogous to the Gram-Schmidt method of constructing a sequence of orthogonal functions out of any arbitrary set of functions. Phrased in geometrical language, it is also seen to be identical with the procedure followed in Chapter 5 for multiple eigenvalues of the inertia tensor. For example, the added indeterminacy in the eigenvector components for a double root means that all of the vectors in a *plane* are eigenvectors. We merely choose any two perpendicular

directions in the plane as being the new principal axes, with the eigenvectors in \mathbf{A} as unit vectors along these axes.

6.3 ■ FREQUENCIES OF FREE VIBRATION, AND NORMAL COORDINATES

The somewhat lengthy arguments of the preceding section demonstrate that the equations of motion will be satisfied by an oscillatory solution of the form (6.11), not merely for one frequency but in general for a set of n frequencies ω_k . A complete solution of the equations of motion therefore involves a superposition of oscillations with all the allowed frequencies. Thus, if the system is displaced slightly from equilibrium and then released, the system performs small oscillations about the equilibrium with the frequencies $\omega_1, \dots, \omega_n$. The solutions of the secular equation are therefore often designated as the frequencies of *free vibration* or as the *resonant frequencies* of the system.

The general solution of the equations of motion may now be written as a summation over an index k :

$$\eta_i = C_k a_{ik} e^{-i\omega_k t}, \quad (6.35)$$

there being a complex scale factor C_k for each resonant frequency. It might be objected that for each solution λ_k of the secular equation there are two resonant frequencies $+\omega_k$ and $-\omega_k$. The eigenvector \mathbf{a}_k would be the same for the two frequencies, but the scale factors C_k^+ and C_k^- could conceivably be different. On this basis, the general solution should appear as

$$\eta_i = a_{ik} (C_k^+ e^{+i\omega_k t} + C_k^- e^{-i\omega_k t}). \quad (6.35')$$

Recall however that the actual motion is the real part of the complex solution, and the real part of either (6.35) or (6.35') can be written in the form

$$\eta_i = f_k a_{ik} \cos(\omega_k t + \delta_k), \quad (6.36)$$

where the amplitude f_k and the phase δ_k are determined from the initial conditions. Either of the solutions ((6.35) and (6.36)) will therefore represent the actual motion, and the former of course is the more convenient.

The orthogonality properties of \mathbf{A} greatly facilitate the determination of the scale factors C_k in terms of the initial conditions. At $t = 0$, the real part of Eq. (6.35) reduces to

$$\eta_i(0) = \operatorname{Re} C_k a_{ik}, \quad (6.37)$$

where Re stands for “real part of.” Similarly, the initial value of the velocities is obtained as

$$\dot{\eta}_i(0) = \operatorname{Im} C_k a_{ik} \omega_k, \quad (6.38)$$

where $\text{Im } C_k$ denotes the imaginary part of C_k . From these $2n$ equations, the real and imaginary parts of the n constants C_k may be evaluated. To solve Eq. (6.37), for example, let us first write it in terms of column matrices $\boldsymbol{\eta}(0)$ and \mathbf{C} :

$$\boldsymbol{\eta}(0) = \mathbf{A} \text{Re } \mathbf{C}. \quad (6.37')$$

If we multiply by $\tilde{\mathbf{A}}\mathbf{T}$ from the left and use Eq. (6.23), we immediately obtain a solution for $\text{Re } \mathbf{C}$:

$$\text{Re } \mathbf{C} = \tilde{\mathbf{A}}\mathbf{T}\boldsymbol{\eta}(0).$$

or, taking the l th component,

$$\text{Re } C_l = a_{jl} T_{jk} \dot{\eta}_k(0). \quad (6.39)$$

A similar procedure leads to the imaginary part of the scale factors as*

$$\text{Im } C_l = \frac{1}{\omega_l} \sum_{j,k} a_{jl} T_{jk} \dot{\eta}_k(0). \quad (6.40)$$

Equations (6.39) and (6.40) thus permit the direct computation of the complex factors C_l (and therefore the amplitudes and phases) in terms of the initial conditions and the matrices \mathbf{T} and \mathbf{A} .

The solution for each coordinate, Eq. (6.35), is in general a sum of simple harmonic oscillations in all of the frequencies ω_k satisfying the secular equation. Unless it happens that all of the frequencies are commensurable, that is, rational fractions of each other, η_i never repeats its initial value and is therefore not itself a periodic function of time. However, it is possible to transform from the η_i to a new set of generalized coordinates that are all simple periodic functions of time—a set of variables known as the *normal coordinates*.

We define a new set of coordinates ζ_j

$$\eta_i = a_{ii} \zeta_j, \quad (6.41)$$

or, in terms of single column matrices $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$,

$$\boldsymbol{\eta} = \mathbf{A} \boldsymbol{\zeta}. \quad (6.41')$$

The potential energy, Eq. (6.4), is written in matrix notation as

$$V = \frac{1}{2} \tilde{\boldsymbol{\eta}} \mathbf{V} \boldsymbol{\eta}. \quad (6.42)$$

Now, the single-row transpose matrix $\tilde{\boldsymbol{\eta}}$ is related to $\boldsymbol{\zeta}$ by the equation

$$\tilde{\boldsymbol{\eta}} = \widetilde{\mathbf{A}} \boldsymbol{\zeta} = \boldsymbol{\zeta} \widetilde{\mathbf{A}},$$

*The summation over j and k is shown explicitly because there is no summation over the repeated subscript l .

so that the potential energy can be written also as

$$V = \frac{1}{2} \tilde{\zeta} \tilde{\mathbf{A}} \mathbf{V} \mathbf{A} \zeta.$$

But \mathbf{A} diagonalizes \mathbf{V} by a congruence transformation (cf. Eq. (6.26)), and the potential energy therefore reduces simply to

$$V = \frac{1}{2} \tilde{\zeta} \tilde{\mathbf{A}} \mathbf{V} \mathbf{A} \zeta = \frac{1}{2} \omega_k^2 \zeta_k^2. \quad (6.43)$$

The kinetic energy has an even simpler form in the new coordinates. Since the velocities transform as the coordinates, T as given in Eq. (6.20) transforms to

$$T = \frac{1}{2} \tilde{\zeta} \tilde{\mathbf{A}} \mathbf{T} \mathbf{A} \dot{\zeta}$$

which by virtue of Eq. (6.23) reduces to

$$T = \frac{1}{2} \tilde{\zeta} \dot{\zeta} = \frac{1}{2} \dot{\zeta}_k \dot{\zeta}_k. \quad (6.44)$$

Equations (6.43) and (6.44) state that in the new coordinates both the potential and kinetic energies are sums of squares only, without any cross terms. Of course, this result is simply another way of saying that \mathbf{A} produces a principal axis transformation. Recall that the principal axis transformation of the inertia tensor was specifically designed to reduce the moment of inertia to a sum of squares; the new axes being the principal axes of the inertia ellipsoid. Here the kinetic and potential energies are also quadratic forms (as was the moment of inertia) and both are diagonalized by \mathbf{A} . For this reason, the principal axis transformation employed here is a particular example of the well-known algebraic process of the *simultaneous diagonalization of two quadratic forms*.

The equations of motion share in the simplification resulting from their use. The new Lagrangian is

$$L = \frac{1}{2} (\dot{\zeta}_k \dot{\zeta}_k - \omega_k^2 \zeta_k^2) \quad (6.45)$$

so that the Lagrange equations for ζ_k are

$$\ddot{\zeta}_k + \omega_k^2 \zeta_k = 0. \quad (6.46)$$

Equations (6.47) have the immediate solutions

$$\zeta_k = C_k e^{-i\omega_k t}, \quad (6.47)$$

which could have been seen of course directly from Eqs. (6.35) and (6.41). Each of the new coordinates is thus a simply periodic function involving only *one* of the resonant frequencies. As mentioned earlier, it is therefore customary to call the ζ 's the *normal coordinates* of the system.

Each normal coordinate corresponds to a vibration of the system with only one frequency, and these component oscillations are spoken of as the *normal modes of vibration*. All of the particles in each mode vibrate with the same frequency and with the same phase,* the relative amplitudes being determined by the matrix

*Particles may be exactly out of phase if the a 's have opposite sign

elements a_{jk} . The complete motion is then built up out of the sum of the normal modes weighted with appropriate amplitude and phase factors contained in the C_k 's.

Harmonics of the fundamental frequencies are absent in the complete motion essentially because of the stipulation that the amplitude of oscillation be small. We are then allowed to represent the potential as a quadratic form, which is characteristic of simple harmonic motion. The normal coordinate transformation emphasizes this point, for the Lagrangian in the normal coordinates (6.45) is seen to be the sum of the Lagrangians for harmonic oscillators of frequencies ω_k . We can thus consider the complete motion for small oscillations as being obtained by exciting the various harmonic oscillators with different intensities and phases.*

6.4 ■ FREE VIBRATIONS OF A LINEAR TRIATOMIC MOLECULE

To illustrate the technique for obtaining the resonant frequencies and normal modes, we shall consider in detail a model based on a linear symmetrical triatomic molecule. In the equilibrium configuration of the molecule, two atoms of mass m are symmetrically located on each side of an atom of mass M (cf. Fig. 6.3). All three atoms are on one straight line, the equilibrium distances apart being denoted by b . For simplicity, we shall first consider only vibrations along the line of the molecule, and the actual complicated interatomic potential will be approximated by two springs of force constant k joining the three atoms. There are three obvious coordinates marking the position of the three atoms on the line. In these coordinates, the potential energy is

$$V = \frac{k}{2}(x_2 - x_1 - b)^2 + \frac{k}{2}(x_3 - x_2 - b)^2. \quad (6.48)$$

We now introduce coordinates relative to the equilibrium positions:

$$\eta_i = x_i - x_{0i},$$

where

$$x_{02} - x_{01} = b = x_{03} - x_{02}.$$

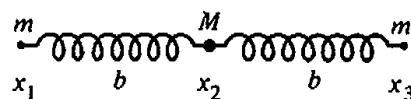


FIGURE 6.3 Model of a linear symmetrical triatomic molecule.

*Note for future reference that the same sort of picture appears in the quantization of the electromagnetic field. The frequencies of the harmonic oscillators are identified with the photon frequencies, and the amplitudes of excitation become the discrete quantized "occupation numbers"—the number of photons of each frequency.

The potential energy then reduces to

$$V = \frac{k}{2}(\eta_2 - \eta_1)^2 + \frac{k}{2}(\eta_3 - \eta_2)^2,$$

or

$$V = \frac{k}{2}(\eta_1^2 + 2\eta_2^2 + \eta_3^2 - 2\eta_1\eta_2 - 2\eta_2\eta_3). \quad (6.49)$$

Hence, the **V** tensor has the form

$$\mathbf{V} = \begin{pmatrix} k & -k & 0 \\ -k & 2k & -k \\ 0 & -k & k \end{pmatrix}. \quad (6.50)$$

The kinetic energy has an even simpler form:

$$T = \frac{m}{2}(\dot{\eta}_1^2 + \dot{\eta}_3^2) + \frac{M}{2}\dot{\eta}_2^2, \quad (6.51)$$

so that the **T** tensor is diagonal:

$$\mathbf{T} = \begin{pmatrix} m & 0 & 0 \\ 0 & M & 0 \\ 0 & 0 & m \end{pmatrix}. \quad (6.52)$$

Combining these two tensors, the secular equation appears as

$$|\mathbf{V} - \omega^2 \mathbf{T}| = \begin{vmatrix} k - \omega^2 m & -k & 0 \\ -k & 2k - \omega^2 M & -k \\ 0 & -k & k - \omega^2 m \end{vmatrix} = 0. \quad (6.53)$$

Direct evaluation of the determinant leads to the cubic equation in ω^2 :

$$\omega^2(k - \omega^2 m)(k(M + 2m) - \omega^2 Mm) = 0, \quad (6.54)$$

with the obvious solutions

$$\omega_1 = 0, \quad \omega_2 = \sqrt{\frac{k}{m}}, \quad \omega^3 = \sqrt{\frac{k}{m} \left(1 + \frac{2m}{M}\right)}. \quad (6.55)$$

The first eigenvalue, $\omega_1 = 0$, may appear somewhat surprising and even alarming at first sight. Such a solution does not correspond to an oscillatory motion at all, for the equation of motion for the corresponding normal coordinate is

$$\ddot{\xi}_1 = 0,$$

which produces a uniform translational motion. But this is precisely the key to the difficulty. The vanishing frequency arises from the fact that the molecule

may be translated rigidly along its axis without any change in the potential energy, an example of neutral equilibrium mentioned previously. Since the restoring force against such motion is zero, the effective "frequency" must also vanish. We have made the assumption that the molecule has three degrees of freedom for vibrational motion, whereas in reality one of them is a rigid body degree of freedom.

A number of interesting points can be discussed in connection with a vanishing resonant frequency. It is seen from Eq. (6.21) that a zero value of ω can occur only when the potential energy is positive but is not positive definite; that is, it can vanish even when not all the η_i 's are zero. An examination of V , Eq. (6.49), shows that it is not positive definite and that V does in fact vanish when all the η 's are equal (uniform translation).

Since the zero frequency found here is of no consequence for the vibration frequencies of interest, it is often desirable to phrase the problem so that the root is eliminated from the outset. We can do this here most simply by imposing the condition or constraint that the center of mass remain stationary at the origin:

$$m(x_1 + x_3) + Mx_2 = 0. \quad (6.56)$$

Equation (6.56) can then be used to eliminate one of the coordinates from V and T , reducing the problem to one of two degrees of freedom (cf. Derivation 1, this chapter).

The restriction of the motion to be along the molecular axis allows only one possible type of uniform rigid body motion. However, if the more general problem of vibrations in all three directions is considered, the number of rigid body degrees of freedom will be increased to six. The molecule may then translate uniformly along the three axes or perform uniform rotations about the axes. Hence, in any general system of n degrees of freedom, there will be six vanishing frequencies and only $n - 6$ true vibration frequencies. Again, the reduction in the number of degrees of freedom can be performed beforehand by imposing the conservation of linear and angular momentum upon the coordinates.

In addition to rigid body motion, it has been pointed out that zero resonant frequencies may also arise when the potential is such that both the first *and* second derivatives of V vanish at equilibrium. Small oscillations may still be possible in this case if the fourth derivatives do not also vanish (the third derivatives must vanish for a stable equilibrium), but the vibrations will not be simple harmonic. Such a situation therefore constitutes a breakdown of the customary method of small oscillations, but fortunately it is not of frequent occurrence.

Returning now to the examination of the resonant frequencies, ω_2 will be recognized as the well-known frequency of oscillation for a mass m suspended by a spring of force constant k . We are therefore led to expect that only the end atoms partake in this vibration; the center molecule remains stationary. It is only in the third mode of vibration, ω_3 , that the mass M can participate in the oscillatory motion. These predictions are verified by examining the eigenvectors for the three normal modes.

The components a_{ij} are determined for each frequency by the equations

$$\begin{aligned} (k - \omega_j^2 m) a_{1j} & - k a_{2j} = 0 \\ - k a_{1j} + (2k - \omega_j^2 M) a_{2j} & - k a_{3j} = 0 \\ - k a_{2j} + (k - \omega_j^2 m) a_{3j} & = 0, \end{aligned} \quad (6.57a)$$

along with the normalization condition:

$$m(a_{1j}^2 + a_{3j}^2) + Ma_{2j}^2 = 1. \quad (6.57b)$$

For $\omega_1 = 0$, it follows immediately from the first and third of Eqs. (6.57a) that all three coefficients are equal: $a_{11} = a_{21} = a_{31}$. This of course is exactly what was expected from the translational nature of the motion (cf. Fig. 6.4a). The normalization condition then fixes the value of a_{1j} so that

$$a_{11} = \frac{1}{\sqrt{2m + M}}, \quad a_{12} = \frac{1}{\sqrt{2m + M}}, \quad a_{13} = \frac{1}{\sqrt{2m + M}}. \quad (6.58a)$$

The factors $(k - \omega_2^2 m)$ vanish for the second mode, and Eqs. (6.57a) show immediately that $a_{22} = 0$ (as predicted) and $a_{12} = -a_{32}$. The numerical value of these quantities is then determined by Eq. (6.57b):

$$a_{12} = \frac{1}{\sqrt{2m}}, \quad a_{22} = 0, \quad a_{32} = -\frac{1}{\sqrt{2m}}. \quad (6.58b)$$

In this mode the center atom is at rest, while the two outer ones vibrate exactly out of phase (as they must in order to conserve linear momentum) (cf. Fig. 6.4b). Finally, when $\omega = \omega_3$, it can be seen from the first and third of Eqs. (6.57a) that a_{13} and a_{33} must be equal. The rest of the calculation for this mode is not quite as simple as for the others, and it will be sufficient to state the final result:

$$a_{13} = \frac{1}{\sqrt{2m \left(1 + \frac{2m}{M}\right)}}, \quad a_{23} = \frac{-2}{\sqrt{2M \left(2 + \frac{M}{m}\right)}}, \quad a_{33} = \frac{1}{\sqrt{2m \left(1 + \frac{2m}{M}\right)}}. \quad (6.58c)$$

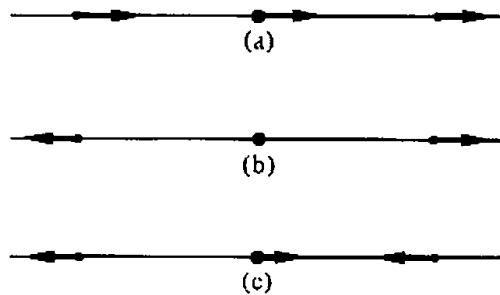


FIGURE 6.4 Longitudinal normal modes of the linear symmetric triatomic molecule

Here the two outer atoms vibrate with the same amplitude, while the inner one oscillates out of phase with them and has a different amplitude, (cf. Fig. 6.4c.)

The normal coordinates may be found by inverting Eq. (6.41) as

$$\begin{aligned}\xi_1 &= \frac{1}{\sqrt{2m+M}}(\sqrt{m}\eta_1 + \sqrt{M}\eta_2 + \sqrt{m}\eta_3) \\ \xi_2 &= \sqrt{\frac{1}{2}}(\eta_1 - \eta_3) \\ \xi_3 &= \frac{1}{\sqrt{2m+M}}\left[\sqrt{\frac{M}{2}}(\eta_1 + \eta_3) - \sqrt{2m}\eta_2\right].\end{aligned}\quad (6.59)$$

These normal modes describe each of the behaviors shown on Fig. 6.4. Any general longitudinal vibration of the molecule that does not involve a rigid translation will be some linear combination of the normal modes ω_2 and ω_3 . The amplitudes of the normal modes, and their phases relative to each other, will of course be determined by the initial conditions (cf. Exercise 5).

We have spoken so far only of vibrations along the axis; in the actual molecule there will also be normal modes of vibration perpendicular to the axis. The complete set of normal modes is naturally more difficult to determine than merely the longitudinal modes, for the general motion in all directions corresponds to nine degrees of freedom. While the procedure is straightforward, the algebra rapidly becomes quite complicated, and it is not feasible to present the detailed calculation here. However, it is possible to give a qualitative discussion on the basis of general principles, and most of the conclusions of the complete solution can be predicted beforehand.

The general problem will have a number of zero resonant frequencies corresponding to the possibility of rigid body motion. For a molecule with n atoms there are $3n$ degrees of freedom. Subtracting the three translational and three rigid rotational degrees of freedom, there will be in general $3n - 6$ vibrational modes. For the linear molecule, there will be three degrees of freedom for rigid translation, but rigid rotation can account for only *two* degrees of freedom. Rotation about the axis of the molecule is obviously meaningless and will not appear as a mode of rigid body motion. We are therefore left with four true modes of vibration. Two of these are the longitudinal modes, which have already been examined so that there can only be two modes of vibration perpendicular to the axis. However, the symmetry of the molecule about its axis shows that these two modes of perpendicular vibration must be degenerate. There is nothing to distinguish a vibration in the y direction from a vibration in the z direction, and the two frequencies must be equal.

The additional indeterminacy of the eigenvectors of a degenerate mode appears here, in that all directions perpendicular to the molecular axis are alike. Any two orthogonal axes in the plane normal to the molecule may be chosen as the directions of the degenerate modes of vibration. The complete motion of the atoms

normal to the molecular axis will depend upon the amplitudes and relative phases of the two degenerate modes. If both are excited, and they are exactly in phase, then the atoms will move on a straight line passing through the equilibrium configuration. But if they are out of phase, the composite motion is an elliptical Lissajous figure, exactly as in a two-dimensional isotropic oscillator. The two modes then represent a rotation, rather than a vibration.

It is obvious from the symmetry of the molecules that the amplitudes of the end atoms must be identical in magnitude. The complete calculation shows that the end atoms also travel in the same direction along the Lissajous figure. Hence, the center atom must revolve in the opposite direction, in order to conserve angular momentum. Figure 6.5 illustrates the motion for the two degenerate modes when they are 90° out of phase.

As the complexity of the molecule increases, the size of the secular determinant becomes very large, and finding the normal frequencies and amplitudes becomes a problem of considerable magnitude. We have seen however that even in a situation as simple as the linear triatomic molecule, a study of the symmetries to be expected in the vibrations greatly simplifies the calculations. Considerable mathematical ingenuity has been devoted to exploiting the symmetries inherent in complex molecules to reduce the labor involved in finding their vibration frequencies. The theory of symmetry groups has been applied with great success in factoring the large secular determinant into smaller blocks that may be diagonalized separately. It has been pointed out however that such elaborate mathematical manipulation was more appropriate in a time when numerical computations were difficult and tedious. Considering the speed and memory capacity of present-day computers, a straightforward approach may be easier and more accurate in the long run. Fast and accurate routines for solving the eigenvalue problems of large matrices are the stock-in-trade today of scientific computers of even moderate size. There has therefore been a trend toward a more brute-force approach in which mass-weighted Cartesian coordinates (see p. 241) are used to formulate the problem. The kinetic energy ellipsoid for the molecular vibrations is then already a sphere, and finding the normal modes reduces to diagonalizing the potential energy. These approaches are extensively applied in infrared and Raman spectroscopy.

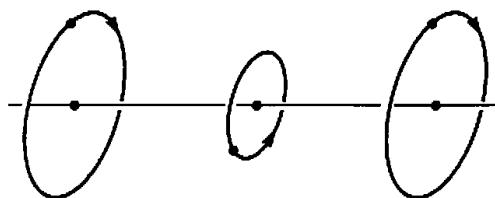


FIGURE 6.5 Degenerate modes of the symmetrical triatomic molecule.

6.5 ■ FORCED VIBRATIONS AND THE EFFECT OF DISSIPATIVE FORCES

Free vibrations occur when the system is displaced initially from its equilibrium configuration and is then allowed to oscillate by itself. Very often, however, the system is set into oscillation by an external driving force that continues to act on the system after $t = 0$. The frequency of such a *forced oscillation* is then determined by the frequency of the driving force and not by the resonant frequencies. Nevertheless, the normal modes are of great importance in obtaining the *amplitudes* of the forced vibration, and the problem is greatly simplified by use of the normal coordinates obtained from the free modes.

If F_j is the generalized force corresponding to the coordinate η_j , then by Eq. (1.49) the generalized force Q_i for the normal coordinate ξ_i is

$$Q_i = a_{ji} F_j. \quad (6.60)$$

The equations of motion when expressed in normal coordinates now become

$$\ddot{\xi}_i + \omega_i^2 \xi_i = Q_i. \quad (6.61)$$

Equations (6.61) are a set of n inhomogeneous differential equations that can be solved only when we know the dependence of Q_i on time. While the solution will not be as simple as in the free case, note that the normal coordinates preserve their advantage of separating the variables, and each equation involves only a single coordinate.

Frequently, the driving force varies sinusoidally with time. In an acoustic problem, for example, the driving force might arise from the pressure of a sound wave impinging on the system, and Q_i then has the same frequency as the sound wave. Or, if the system is a polyatomic molecule, a sinusoidal driving force is present if the molecule is illuminated by a monochromatic light beam. Each atom in the molecule is then subject to an electromagnetic force whose frequency is that of the incident light. Even where the driving force is not sinusoidal with a single frequency, it can often be considered as built up as a superposition of such sinusoidal terms. Thus, if the driving force is periodic, it can be represented by a Fourier series; other times, a Fourier integral representation is suitable. Since Eqs. (6.61) are linear equations, its solutions for particular frequencies can be superposed to find the complete solution for given Q_i .

It is therefore of general interest to study the nature of the oscillations when the force Q_i can be written as

$$Q_i = Q_{0i} \cos(\omega t + \delta_i), \quad (6.62)$$

where ω is the angular frequency of an external force. The equations of motion now appear as

$$\ddot{\xi}_i + \omega_i^2 \xi_i = Q_{0i} \cos(\omega t + \delta_i). \quad (6.63)$$

A complete solution of Eq. (6.63) consists of the general solution to the homogeneous equation (that is, the free modes of vibration) plus a particular solution to the inhomogeneous equation. By a proper choice of initial conditions, the superimposed free vibrations can be made to vanish,* centering our interest on the particular solution of Eqs. (6.63) that will obviously have the form

$$\zeta_i = B_i \cos(\omega t + \delta_i). \quad (6.64)$$

Here the amplitudes B_i are determined by substituting the solution in Eqs. (6.63):

$$B_i = \frac{Q_{0i}}{\omega_i^2 - \omega^2}. \quad (6.65)$$

The complete motion is then

$$\eta_j = a_{ji} \zeta_i = \frac{a_{ji} Q_{0i} \cos(\omega t + \delta_i)}{\omega_i^2 - \omega^2}. \quad (6.66)$$

Thus, the vibration of each particle is again composed of linear combinations of the normal modes, but now each normal oscillation occurs at the frequency of the driving force.

Two factors determine the extent to which each normal mode is excited. One is the amplitude of the generalized driving force, Q_{0i} . If the force on each particle has no component in the direction of vibration of some particular normal mode, then obviously the generalized force corresponding to the mode will vanish and Q_{0i} will be zero. *An external force can excite a normal mode only if it tends to move the particles in the same direction as in the given mode.* The second factor is the closeness of the driving frequency to the free frequency of the mode. As a consequence of the denominators in Eq. (6.66), the closer ω approaches to any ω_i , the stronger will that mode be excited relative to the other modes. Indeed, Eq. (6.66) apparently predicts infinite amplitude when the driving frequency agrees exactly with one of the ω_i 's—the familiar phenomenon of resonance. Actually, of course, the theory behind Eq. (6.66) presumes only small oscillations about equilibrium positions; when the amplitude predicted by the formula becomes large, this assumption breaks down and Eq. (6.66) is then no longer valid. Note that the oscillations are in phase with the driving force when the frequency is less than the resonant frequency, but that there is a phase change of π in going through the resonance.

Our discussion has been unrealistic in that the absence of dissipative or frictional forces has been assumed. In many physical systems, these forces, when present, are proportional to the particle velocities and can therefore be derived

*The free vibrations are essentially the transients generated by the application of the driving forces. If we consider the system to be initially in an equilibrium configuration, and then slowly build up the driving forces from zero, these transients will not appear. Alternatively, dissipative forces can be assumed present (see pages following) that will damp out the free vibrations.

from a dissipation function \mathcal{F} (cf. Section 1.5). Let us first consider the effects of frictional forces on the free modes of vibration.

From its definition, \mathcal{F} must be a homogeneous quadratic function of the velocities:

$$\mathcal{F} = \frac{1}{2} \mathcal{F}_{ij} \dot{\eta}_i \dot{\eta}_j. \quad (6.67)$$

The coefficients \mathcal{F}_{ij} are clearly symmetric, $\mathcal{F}_{ii} = \mathcal{F}_{jj}$, and in general will be functions of the coordinates. Since we are concerned with only small vibrations about equilibrium, it is sufficient to expand the coefficients about equilibrium and retain only the first, constant term, exactly as was done for the kinetic energy. In future applications of Eq. (6.67), we shall take \mathcal{F}_{ij} as denoting these constant factors. Recall that $2\mathcal{F}$ is the rate of energy dissipation due to the frictional forces (cf. Eq. (2.60)). The dissipation function \mathcal{F} therefore can never be negative. The complete set of Lagrange equations of motion now become (cf. Section 1.5)

$$T_{ij} \ddot{\eta}_j + \mathcal{F}_{ij} \dot{\eta}_j + V_{ij} \eta_j = 0. \quad (6.68)$$

Clearly in order to find normal coordinates for which the equations of motion would be decoupled, it is necessary to find a principal axis transformation that simultaneously diagonalizes the three quadratic forms T , V , and \mathcal{F} . As was shown above, this is not in general possible; normal modes cannot usually be found for any arbitrary dissipation function.

There are however some exceptional cases when simultaneous diagonalization is possible. For example, if the frictional force is proportional both to the particle's velocity *and* its mass, then \mathcal{F} will be diagonal whenever T is. When such simultaneous diagonalization is feasible, then the equations of motion are decoupled in the normal coordinates with the form

$$\ddot{\xi}_i + \mathcal{F}_i \dot{\xi}_i + \omega_i^2 \xi_i = 0. \quad (\text{no summation}) \quad (6.69)$$

Here the \mathcal{F}_i 's are the nonnegative coefficients in the diagonalized form of \mathcal{F} when expressed in terms of ξ_i . Being a set of linear differential equations with constant coefficients, Eqs. (6.69) may be solved by functions of the form

$$\xi_i = C_i e^{-i\omega'_i t},$$

where ω'_i satisfies the quadratic equation

$$\omega'^2 + i\omega'_i \mathcal{F}_i - \omega_i^2 = 0. \quad (\text{no summation}) \quad (6.70)$$

Equation (6.70) has the two solutions

$$\omega'_i = \pm \sqrt{\omega_i^2 - \frac{\mathcal{F}_i^2}{4}} - i \frac{\mathcal{F}_i}{2}. \quad (6.71)$$

The motion is therefore not a pure oscillation, for ω' is complex. It is seen from Eq. (6.71) that the imaginary part of ω' results in a factor $\exp(-\mathcal{F}_i t/2)$, and by reason of the nonnegative nature of the \mathcal{F}_i 's, this is always an exponentially decreasing function of time.* The presence of a damping factor due to the friction is hardly unexpected. As the particles vibrate, they do work against the frictional forces, and the energy of the system (and hence the vibration amplitudes) must decrease with time. The real part of Eq. (6.71) corresponds to the oscillatory factor in the motion; note that the presence of friction also affects the frequency of the vibration. However, if the dissipation is small, the squared term in \mathcal{F}_i may be neglected, and the frequency of oscillation reduces to the friction-free value. The complete motion is then simply an exponential damping of the free modes of vibration:

$$\zeta_i = C_i e^{-\mathcal{F}_i t/2} e^{-i\omega_i t}. \quad (6.72)$$

If the dissipation function cannot be diagonalized along with T and V , the solution is much more difficult to obtain. The general nature of the solution remains pretty much the same, however: an exponential damping factor times an oscillatory exponential function. Suppose we seek a solution to Eqs. (6.68) of the form

$$\eta_j = C_a e^{-i\omega t} = C_a e^{-\kappa t} e^{-2\pi i\nu t}. \quad (6.73)$$

With this solution, Eqs. (6.68) become a set of simultaneous linear equations

$$V_{ij}a_j - i\omega\mathcal{F}_{ij}a_j - \omega^2 T_{ij}a_j = 0. \quad (6.74)$$

It is convenient to write ω as $i\gamma$, so that

$$\gamma = -i\omega = -\kappa - 2\pi i\nu, \quad (6.75)$$

and thus $-\kappa$ is the real part of γ . In terms of the square tensors of V , T , and \mathcal{F} , the set of equations (6.74) become a column matrix equation involving γ :

$$\mathbf{Va} + \gamma \mathbf{Fa} + \gamma^2 \mathbf{Ta} = 0. \quad (6.76)$$

The set of homogeneous equations (6.74) or (6.76) can be solved for the a_j only for certain values of ω or γ .

Without actually evaluating the corresponding secular equation, we can show that κ must always be nonnegative. Convert the matrix equation (6.76) into a scalar equation for γ by multiplying from the left with \mathbf{a}^\dagger :

$$\mathbf{a}^\dagger \mathbf{Va} + \gamma \mathbf{a}^\dagger \mathbf{Fa} + \gamma^2 \mathbf{a}^\dagger \mathbf{Ta} = 0. \quad (6.77)$$

*Some (but not all) \mathcal{F}_i 's may be zero, which simply means there are no frictional effects in the corresponding normal modes. The important point is that the \mathcal{F}_i 's cannot be negative.

Equation (6.77) is a quadratic equation for γ with coefficients that are matrix products of the same general type as those encountered in Eq. (6.19). By virtue of the symmetry of \mathbf{V} , \mathbf{F} , and \mathbf{T} , the matrix products are all real, as can be seen by expanding \mathbf{a} as $\alpha + i\beta$ (cf. Eq. (6.19)). Hence, if γ is a solution of the quadratic equation, its complex conjugate γ^* must also be a solution. Now, the sum of the two roots of a quadratic equation is the negative of the coefficient of the linear term divided by the coefficient of the square term

$$\gamma + \gamma^* = -2\kappa = -\frac{\mathbf{a}^\dagger \mathbf{F} \mathbf{a}}{\mathbf{a}^\dagger \mathbf{T} \mathbf{a}}. \quad (6.78)$$

Hence, κ can be expressed in terms of the real and imaginary parts of a_j as

$$\kappa = \frac{1}{2} \frac{\mathcal{F}_{ij}(\alpha_i \alpha_j + \beta_i \beta_j)}{T_{kl}(\alpha_k \alpha_l + \beta_k \beta_l)}. \quad (6.79)$$

The dissipation function \mathcal{F} must always be positive, and T is positive definite; hence κ cannot be negative. The oscillations of the system may decrease exponentially with time, but they can never increase with time. Note that if \mathcal{F} is positive definite, κ must be different from zero (and positive), and all modes will have an exponential damping factor. The frequencies of oscillation, given by the real part of ω , will of course be affected by the dissipative forces, but the change will be small if the damping is not very large during a period of oscillation.

Finally, we may consider forced sinusoidal oscillations in the presence of dissipative forces. Representing the variation of the driving force with time by

$$F_j = F_{0j} e^{-i\omega t},$$

where F_{0j} may be complex, the equations of motion are

$$V_{ij}\eta_j + \mathcal{F}_{ij}\dot{\eta}_j + T_{ij}\ddot{\eta}_j = F_{0i} e^{-i\omega t}. \quad (6.80)$$

If we seek a particular solution to these equations of the form

$$\eta_j = A_j e^{-i\omega t},$$

we obtain the following set of inhomogeneous linear equations for the amplitudes A_j :

$$(V_{ij} - i\omega \mathcal{F}_{ij} - \omega^2 T_{ij})A_j - F_{0i} = 0. \quad (6.81)$$

The solution to these equations* may easily be obtained from Cramer's rule:

$$A_j = \frac{D_j(\omega)}{D(\omega)}. \quad (6.82)$$

*They are of course merely the inhomogeneous version of Eqs. (6.74)

where $D(\omega)$ is the determinant of the coefficients of A_j in Eq. (6.81) and $D_j(\omega)$ is the modification in $D(\omega)$ resulting when the j th column is replaced by $F_{01} \dots F_{0n}$. It is the denominator $D(\omega)$ that is of principal interest to us here, for the resonances arise essentially out of the algebraic form of the denominator. Now, D is the determinant appearing in the secular equation corresponding to the homogeneous equations (6.74); its roots are the complex frequencies of the free modes of vibration. The requirement that both γ and γ^* are roots of Eq. (6.77) means, on the basis of Eq. (6.75), that if ω_t is a root of $D(\omega)$, then $-\omega_t^*$ is a root. For a system of n degrees of freedom, it is therefore possible to represent $D(\omega)$ as

$$D(\omega) = G(\omega - \omega_1)(\omega - \omega_2) \dots (\omega - \omega_n)(\omega + \omega_1^*)(\omega + \omega_2^*) \dots (\omega + \omega_n^*),$$

where G is some constant. Using product notation, and denoting ω by $2\pi v$, this representation can be written as

$$D(\omega) = G \prod_{t=1}^n (2\pi(v - v_t) + i\kappa_t)(2\pi(v + v_t) + i\kappa_t). \quad (6.83)$$

When we rationalize Eq. (6.83) to separate A_t into its real and imaginary parts, the denominator will be

$$D^*(\omega)D(\omega) = GG^* \prod_{t=1}^n (4\pi^2(v - v_t)^2 + \kappa_t^2)(4\pi^2(v + v_t)^2 + \kappa_t^2). \quad (6.84)$$

The amplitudes of the forced oscillation thus exhibit typical resonance behavior in the neighborhood of the frequencies of free oscillations $\pm v_t$. As a result of the presence of the damping constants κ_t , the resonance denominators no longer vanish at the free mode frequencies, and the amplitudes remain finite. The driving frequency at which the amplitude peaks is no longer exactly at the free frequencies because of frequency dependence of terms in A_j other than the particular resonance denominator. However, so long as the damping is small enough to preserve a recognizable resonant peak, the shift in the resonance frequencies is usually small.

We have discussed the properties of small oscillations solely in terms of mechanical systems. The reader however has undoubtedly noticed the similarity with the theory of the oscillations of electrical networks. The equations of motion (6.68) become the circuit equations for n coupled circuits if we read the V_{ij} coefficients as reciprocal capacitances, the \mathcal{F}_{ij} 's as resistances, and the T_{ij} 's as inductances. Driving forces are replaced by generators of frequency ω applied to one or more of the circuits, and the equations of forced vibration (6.80) reduce to the electrical circuit equations (2.42) mentioned in Chapter 2.

We have presented here only a fraction of the techniques that have been devised for handling small oscillations, and of the general theorems about the motion. For example, space does not permit a discussion of the powerful Laplace transform techniques to study the response of a linearly oscillating system to driving forces

with arbitrary time dependencies. Nor is it appropriate here to fully consider the extensive subject of nonlinear oscillations, where the potential energy contains terms beyond the quadratic, and the motion is no longer simple harmonic. (Some relevant portions of this field will be introduced later when we treat chaos and perturbation theory). As mentioned earlier, a formal development of the theory of small oscillations about steady motion will be given later in connection with the Hamiltonian version of mechanics. Another generalization that will deserve our attention relates to the oscillation of systems with continuously infinite numbers of degrees of freedom. The question is how we can construct a way of handling continuous systems that is analogous to the classical mechanics of discrete systems. We shall postpone such considerations of continuous systems to Chapter 13—after we have developed the canonical formulation of discrete mechanics, and after we have seen how the structure of Newtonian mechanics must be modified in the special theory of relativity.

6.6 ■ BEYOND SMALL OSCILLATIONS: THE DAMPED DRIVEN PENDULUM AND THE JOSEPHSON JUNCTION

As an example of forced vibrations with dissipative forces, we consider the motion of the pendulum sketched in Fig. 6.6, which is subjected to an applied torque N , and is permitted to rotate through its full range of motion $-\pi \leq \phi \leq \pi$. In addition, the pendulum is subject to damping by the viscosity η of the medium in which it rotates. For simplicity, we will assume that the rod is massless, and that all of the pendulum mass is concentrated at the end of the rod.

Let us begin by recalling the dynamics of a simple pendulum of length R and mass m . The angular acceleration of the pendulum is produced by the restoring

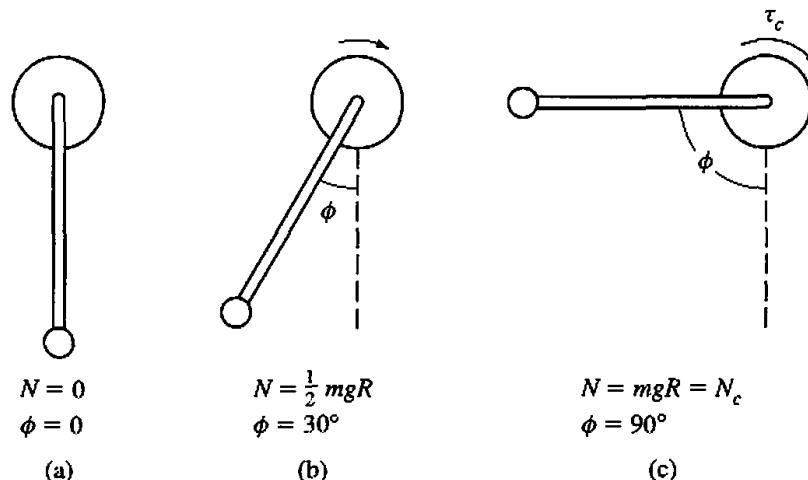


FIGURE 6.6 Pendulum (a) with no applied torque, $N = 0$, (b) with the torque $N = \frac{1}{2} mgR$, and (c) with the critical torque applied, $N_c = mgR$. Figures 6.6, 6.8, 6.10, and 6.11 are adapted from C. P. Poole, Jr., H. A. Farach and R. J. Creswick, “Superconductivity,” Wiley, NY, 1995.

gravitational torque $mgR \sin \phi$ corresponding to the equation of motion

$$mR^2 \frac{d^2\phi}{dt^2} + mgR \sin \phi = 0, \quad (6.85)$$

where $I = mR^2$ is the moment of inertia. For small angular displacements, the approximation $\sin \phi \approx \phi$ linearizes the problem by making the torque proportional to the displacement, and the motion is simple harmonic, $\phi = \phi_0 \sin \omega t$ with the *characteristic frequency* ω_0

$$\omega_0 = \left(\frac{g}{R} \right)^{1/2} \quad (6.86)$$

If a torque N is applied to a stationary pendulum, it will swing out through an angle ϕ . The force of gravity acting on the mass m provides the restoring torque $mgR \sin \phi$, as we noted above, and the pendulum assumes an equilibrium position at the angle ϕ given by

$$N = mgR \sin \phi \quad \left(\frac{d\phi}{dt} = 0 \right), \quad (6.87)$$

as indicated in Fig. 6.6b. The greater the torque, the larger the angle ϕ . There is a critical torque N_c indicated on Fig. 6.6(c) for which the angle ϕ assumes the values $\pi/2$:

$$N_c = mgR. \quad (6.88)$$

If N exceeds this critical value, then the applied torque becomes larger than the restoring torque, $N > mgR \sin \phi$, for all angles ϕ . As a result, the pendulum will begin to rotate beyond $\phi = \pi/2$, and it will continue to rotate as long as the torque $N > N_c$ is applied. The motion will take place at a variable angular speed ω

$$\omega = \frac{d\phi}{dt}, \quad (6.89)$$

and it can persist if the torque is later removed.

With these facts in mind, let us proceed to examine the case of the damped pendulum assuming that the damping force $F_{\text{damp}} = \eta\omega$ is proportional to the angular velocity ω . To write the differential equation of its motion, we add the restoring and damping torques $mgR \sin \phi$ and $\eta d\phi/dt$, respectively, to Eq. (6.85):

$$N = mR^2 \frac{d^2\phi}{dt^2} + \eta \frac{d\phi}{dt} + mgR \sin \phi. \quad (6.90)$$

If we define a *critical frequency* ω_c corresponding to the angular speed at which the damping torque $\eta\omega$ equals the *critical torque* mgR ,

$$\omega_c = \frac{mgR}{\eta} = \frac{N_c}{\eta}, \quad (6.91)$$

then we can write the pendulum equation (6.90) in the normalized form

$$\frac{N}{N_c} = \frac{1}{\omega_0^2} \frac{d^2\phi}{dt^2} + \frac{1}{\omega_c} \frac{d\phi}{dt} + \sin \phi. \quad (6.92)$$

The solutions of this equation exhibit complex time variations of the angular position $\phi(t)$.

When a constant torque is applied to the pendulum at rest, there will be a initial transient behavior that eventually settles down to a *dynamic steady state* after the transients die out. We shall examine several cases of this dynamic steady state.

1. For low applied torques, $N \leq N_c$, there is a *static steady state*

$$N = N_c \sin \phi, \quad (6.93)$$

in which all time derivatives vanish after the initial oscillations have died out. This is illustrated in Fig. 6.6b with the pendulum stationary at the angle ϕ .

2. For undamped motion ($\eta = 0$) with a constant applied torque, N , Eq. (6.90) assumes the form

$$\text{torque} = N - mgR \sin \phi = mR^2 \frac{d^2\phi}{dt^2}, \quad (6.94)$$

so we see that the acting torque is angularly dependent. This torque has special values at four particular angles:

$$\text{torque} = N \quad \phi = 0 \quad (6.95a)$$

$$\text{torque} = N - N_c \quad \phi = \pi/2 \quad (6.95b)$$

$$\text{torque} = N \quad \phi = \pi \quad (6.95c)$$

$$\text{torque} = N + N_c \quad \phi = 3\pi/2 \quad (6.95d)$$

If the applied torque N exceeds the critical torque N_c , the motion will be continuously accelerated rotation, and the pendulum increases its energy as time goes on. The angular speed also increases with time, but with fluctuations that repeat every cycle, as indicated in Fig. 6.7. Note that Fig. 6.7 is drawn for the case where damping is present. The average over these oscillations provides the average angular speed

$$\langle \omega \rangle = \left\langle \frac{d\phi}{dt} \right\rangle, \quad (6.96)$$

which continually increases linearly with the time.

3. When damping is present with $\omega_c \ll \omega_0$ and $N > N_c$, the angular speed ω continues to increase until the damping term $\eta d\phi/dt$ approaches the

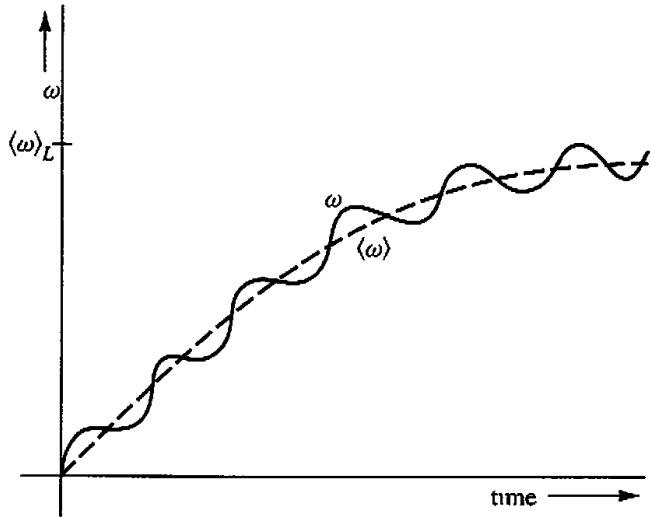


FIGURE 6.7 Dependence of the angular velocity $\omega = d\phi/dt$ on the time for an applied torque $N > N_c$. The average value $\langle \omega \rangle$ increases linearly with time in the absence of damping (linear region), and the overall curve applies to the case $\omega_e \ll \omega_0$ with damping.

value of the applied torque. When this occurs, the average angular speed $\langle \omega \rangle$ approaches a limiting value $\langle \omega \rangle_L$, as shown in Fig. 6.7, and the acceleration fluctuates around an average that is zero: $\langle d\phi^2/dt^2 \rangle = 0$. The pendulum undergoes what is called *quasi-static motion*, rotating with an angular speed ω that undergoes periodic variations but always remains close to the average $\langle \omega \rangle_L$.

To obtain more insight into this quasi-static behavior, we neglect the acceleration term in the equation of motion (6.92), and write

$$\frac{N}{N_c} = \frac{1}{\omega_c} \frac{d\phi}{dt} + \sin \phi, \quad (6.97)$$

which is an equation that can be solved analytically with the solutions

$$\langle \omega \rangle = 0 \quad \text{for } N < N_c \quad (6.98a)$$

$$\langle \omega \rangle = \omega_c [(N/N_c)^2 - 1]^{1/2} \quad \text{for } N > N_c \quad (6.98b)$$

$$\langle \omega \rangle = (N/N_c)\omega_c \quad \text{for } N \gg N_c, \quad (6.98c)$$

which are plotted in Fig. 6.8. The actual cyclic variations in ω for points *A* and *B* on this plot are presented in Fig. 6.9. At point *A*, the applied torque has the value $N = 1.2N_c$, so from Eqs. (6.95) the net torque varies between $0.2N_c$ and $2.2N_c$ around the cycle, and the angular speed is fast at the bottom and slow at the top, with the variations shown at the lower part of Fig. 6.9. For point *B*, we have $N = 2N_c$ so the net torque varies between N_c and $3N_c$, producing the more regular variations in angular speed presented

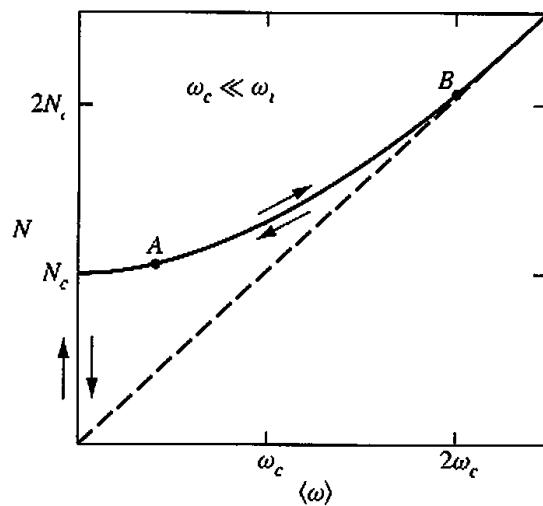


FIGURE 6.8 Relationship between the applied torque N and the average angular velocity $\langle\omega\rangle$ for $\omega_c \ll \omega_0$. We see that $\langle\omega\rangle = 0$ for $N < N_c$ and $\langle\omega\rangle$ increases with increasing $N > N_c$.

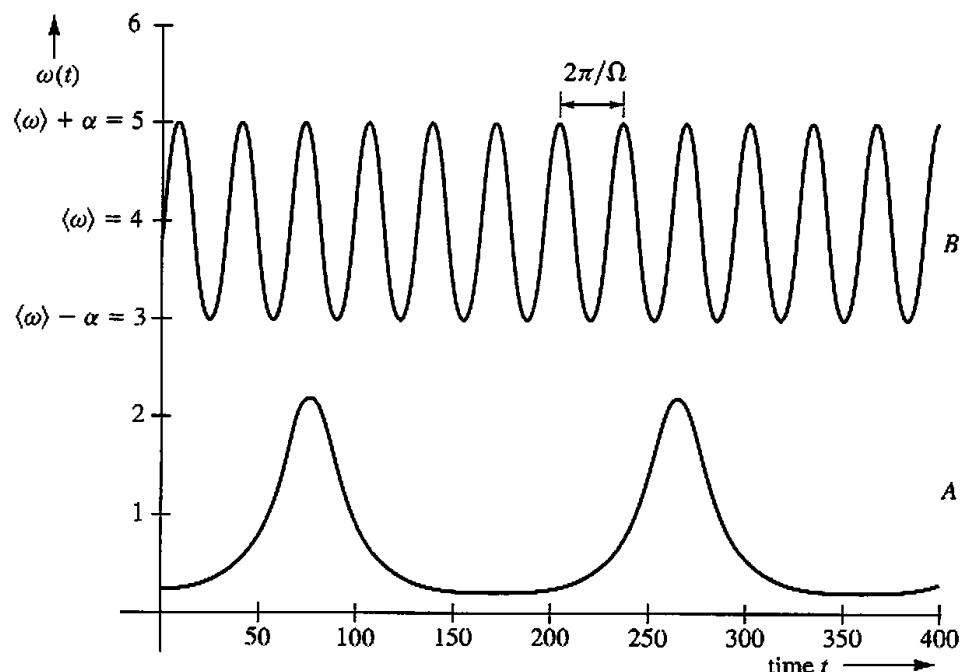


FIGURE 6.9 Oscillations at points A ($N = 1.2N_c$) and B ($2N_c$) for $\omega_c \ll \omega_0$ indicated on Fig. 6.8 for the damped harmonic oscillator. Adapted from A. Barone and G. Paterno, "Physics and Applications of the Josephson Effect," Wiley, NY, 1982.

at the top of Fig. 6.9. In the limit $N \gg N_c$, meaning $\langle \omega \rangle \gg \omega_c$, the angular speed begins to approximate a sinusoidal variation with time

$$\omega(t) \approx \langle \omega \rangle + \alpha \sin \Omega t, \quad (6.99)$$

which approximates point *B* in Fig. 6.8.

4. For the negligible damping case ($\eta \rightarrow 0$ and $\omega_c \gg \omega_0$), the steady-state solution (6.98a) can still occur for $N < N_c$ with the pendulum held fixed at the angle ϕ defined by Eq. (6.93), which means that $\omega = \langle \omega \rangle = 0$. In addition, the solution, (6.98c), in which the torque balances the time averaged damping force, now applies for all values of N , both less than and greater than N_c , and so we have

$$\omega = 0 \quad \text{for } N \leq N_c \quad (6.100a)$$

$$\langle \omega \rangle = (N/N_c)\omega_c \quad \text{for } 0 \leq N \quad (6.100b)$$

These solutions are plotted in Fig. 6.10. Note from the figure that the system exhibits hysteresis, meaning that the behavior differs for increasing and decreasing torques. When the torque is increased for $N < N_c$, the pendulum is stabilized at the angle ϕ satisfying the relation $N = N_c \sin \phi$ of Eq. (6.87), so $\omega = 0$ via Eq. (6.100a). When N reaches the critical torque N_c , the angular speed jumps to the value ω_c , and then rises linearly with further increases in N , as shown in the figure. For decreasing torques, Eq. (6.100b) applies, and $\langle \omega \rangle$ remains proportional to N all the way to the origin, as shown.

5. Figure 6.8 shows the response for $\omega_c \ll \omega_0$, Fig. 6.10 presents it for $\omega_c \gg \omega_0$, and the question arises as to what is the behavior for an intermediate condition such as $\omega_c \approx \omega_0$? This requires solving the general

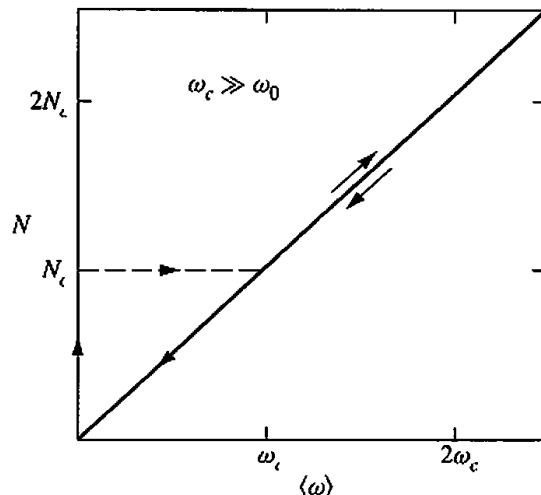


FIGURE 6.10 Relationship between the applied torque N and the average angular velocity $\langle \omega \rangle$ for $\omega_c \gg \omega_0$. There is hysteresis for the behavior when $\langle \omega \rangle < \omega_c$.

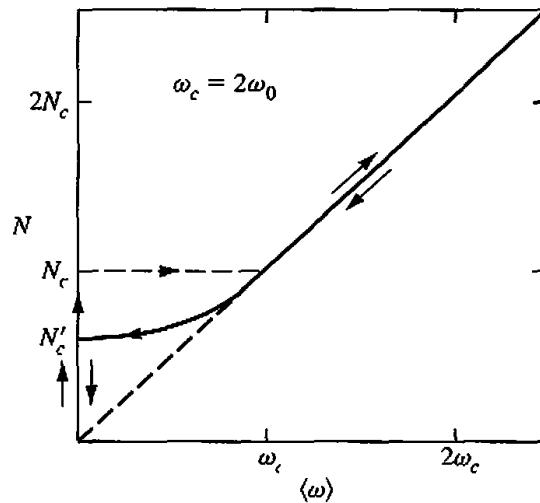


FIGURE 6.11 Relationship between the average angular velocity of the pendulum $\langle \omega \rangle$ and the applied torque N . For low applied torques, the pendulum oscillates and the average velocity is zero, whereas at high torques, $N > N_c$, motion is continuous with $\langle \omega \rangle$ proportional to N . Note the hysteresis for increasing and decreasing torques.

equation (6.92) since no approximations can be made. The N versus $\langle \omega \rangle$ characteristic for the particular case $\omega_c = 2\omega_0$ is plotted in Fig. 6.11. We see from the figure that for increasing torques there is the usual initial rise in N at zero frequency until the critical value N_c is reached, at which point the average angular speed jumps to ω_c , as in the $\omega_c \gg \omega_0$ case of Fig. 6.10. For decreasing torques, there is hysteresis with zero average frequency reached at a torque N'_c , which is less than N_c .

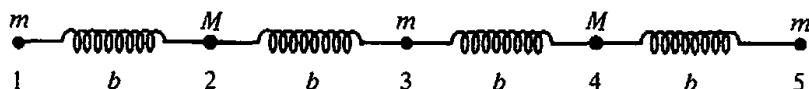
The damped-driven pendulum equation (6.92) has a particularly important application in solid-state physics. When two superconductors are in close proximity with a thin layer of insulating material between them, the arrangement constitutes a *Josephson junction*, which has the property that electric current I can flow across the junction with zero applied voltage, up to a certain critical value I_c . Current exceeding this value is accompanied by the presence of a voltage, and plots of current I versus voltage V for the junction exhibit *hysteresis*. The Josephson junction satisfies the same differential equation (6.93) as the damped oscillator with the current playing the role of the torque, the voltage playing the role of the average angular speed, the capacitance acting like a moment of inertia, and the electrical conductance serving as the viscosity. The variable, which is the angle ϕ for the oscillator, becomes the phase difference ψ across the Josephson junction. Many physicists find it helpful to obtain an intuitive understanding of the operation of the Josephson junction by studying properties of the damped driven pendulum that mimics its behavior.

DERIVATIONS

1. The problem of the linear triatomic molecule can be reduced to one of two degrees of freedom by introducing coordinates $y_1 = x_2 - x_1$, $y_2 = x_3 - x_2$, and eliminating x_2 by requiring that the center of mass remain at rest. Obtain the frequencies of the normal modes in these coordinates and show that they agree with the results of Section 6.4. The distances between the atoms, y_1 and y_2 , are known as *internal coordinates*.
2. Obtain the frequencies of longitudinal vibration of the molecule discussed in Section 6.4, except that now the center atom is to be considered bound to the origin by a spring of force constant k . Show that the translational mode disappears

EXERCISES

3. A bead of mass m is constrained to move on a hoop of radius R . The hoop rotates with constant angular velocity ω around a diameter of the hoop, which is a vertical axis (line along which gravity acts).
 - (a) set up the Lagrangian and obtain the equations of motion of the bead.
 - (b) Find the critical angular velocity Ω below which the bottom of the hoop provides a stable equilibrium for the bead.
 - (c) Find the stable equilibrium position for $\omega > \Omega$.
4. Obtain the normal modes of vibration for the double pendulum shown in Fig. 1.4, assuming equal lengths, but not equal masses. Show that when the lower mass is small compared to the upper one, the two resonant frequencies are almost equal. If the pendula are set in motion by pulling the upper mass slightly away from the vertical and then releasing it, show that subsequent motion is such that at regular intervals one pendulum is at rest while the other has its maximum amplitude. This is the familiar phenomenon of “beats.”
5. (a) In the linear triatomic molecule, suppose the initial condition is that the center atom is at rest but displaced by an amount a_0 from equilibrium, the other two being at their equilibrium points. Find the amplitudes of the longitudinal small oscillations about the center of mass. Give the amplitudes of the normal modes
 - (b) Repeat part (a) but with the center atom initially at its equilibrium position but with an initial speed v_0 .
6. (a) A five-atom linear molecule is simulated by a configuration of masses and ideal springs that looks like the following diagram.



All force constants are equal. Find the eigenfrequencies and normal modes for longitudinal vibrations. [Hint: Transform the coordinates η_i to ζ_i defined by

$$\eta_3 = \zeta_3, \quad \eta_1 = \frac{\zeta_1 + \zeta_5}{\sqrt{2}}, \quad \eta_5 = \frac{\zeta_1 - \zeta_5}{\sqrt{2}}$$

with symmetrical expressions for η_2 and η_4 . The secular determinant will then factor into determinants of lower rank.]

- (b) Solve this problem using computer techniques.

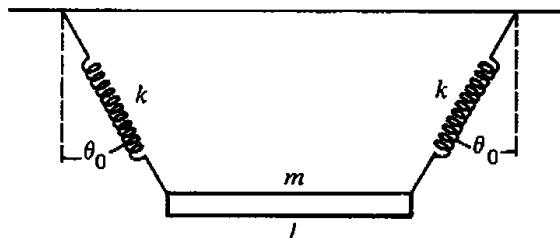
7. In the linear triatomic molecule, suppose that motion in the y and z directions is governed by the potentials

$$V_y = \frac{k}{2}(y_2 - y_1)^2 + \frac{k}{2}(y_3 - y_2)^2,$$

$$V_z = \frac{k}{2}(z_2 - z_1)^2 + \frac{k}{2}(z_3 - z_2)^2.$$

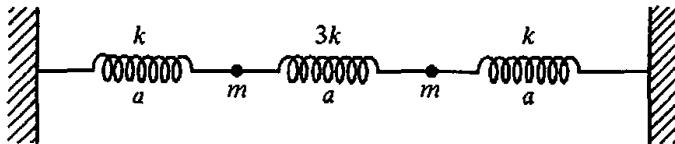
Find the eigenfrequencies for small vibrations in three dimensions and describe the normal modes. What symmetries do the zero frequencies represent? You may want to use the kind of intermediate coordinates suggested in Exercise 6.

8. The equilibrium configuration of a molecule is represented by three atoms of equal mass at the vertices of a 45° right triangle connected by springs of equal force constant. Obtain the secular determinant for the modes of vibration in the plane and show by rearrangement of the columns that the secular equation has a triple root $\omega = 0$. Reduce the determinant to one of third rank and obtain the nonvanishing frequencies of free vibration.
9. Show directly that the equations of motion of the preceding problem are satisfied by
 (a) a uniform translation of all atoms along the x axis, (b) a uniform translation along the y axis, and (c) a uniform rotation about the z axis
10. (a) Three equal mass points have equilibrium positions at the vertices of an equilateral triangle. They are connected by equal springs that lie along the arcs of the circle circumscribing the triangle. Mass points and springs are constrained to move only on the circle, so that, for example, the potential energy of a spring is determined by the arc length covered. Determine the eigenfrequencies and normal modes of small oscillations in the plane. Identify physically any zero frequencies.
 (b) Suppose one of the springs has a change in force constant δk , the others remaining unchanged. To first order in δk , what are the changes in the eigenfrequencies and normal modes?
 (c) Suppose what is changed is the mass of one of the particles by an amount δm . Now how do the normal eigenfrequencies and normal modes change?
11. A uniform bar of length l and mass m is suspended by two equal springs of equilibrium length b and force constant k , as shown in the diagram.



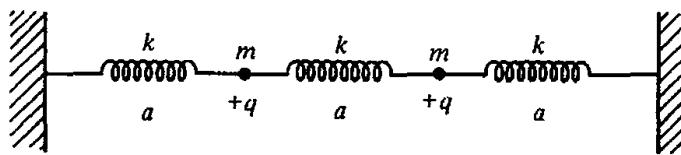
Find the normal modes of small oscillation in the plane.

12. Two particles move in one dimension at the junction of three springs, as shown in the figure. The springs all have unstretched lengths equal to a , and the force constants and masses are shown



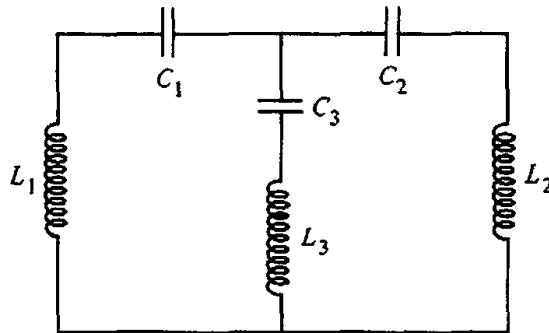
Find the eigenfrequencies and normal modes of the system.

13. Two mass points of equal mass m are connected to each other and to fixed points by three equal springs of force constant k , as shown in the diagram.



The equilibrium length of each spring is a . Each mass point has a positive charge $+q$, and they repel each other according to the Coulomb law. Set up the secular equation for the eigenfrequencies.

14. Find expressions for the eigenfrequencies of the following electrical coupled circuit.



15. If the generalized driving forces Q_i are not sinusoidal, show that the forced vibrations of the normal coordinates in the absence of damping are given by

$$\zeta_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{G_i(\omega)}{\omega_i^2 - \omega^2} e^{-i\omega t} d\omega,$$

where $G_i(\omega)$ is the Fourier transform of Q_i , defined by

$$Q_i(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} G_i(\omega) e^{-i\omega t} d\omega.$$

If the dissipation function is simultaneously diagonalized along with T and V , show that the forced vibrations are given by

$$\zeta_i = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} \frac{G_i(\omega)(\omega_i^2 - \omega^2 + i\omega\mathcal{F}_i)}{(\omega_i^2 - \omega^2)^2 + \omega^2\mathcal{F}_i^2} e^{-i\omega t} dt,$$

which has the typical resonance denominator form. These results are simple illustrations of the powerful technique of the *operational calculus* for handling transient vibrations.

16. A mass particle moves in a constant vertical gravitational field along the curve defined by $y = ax^4$, where y is the vertical direction. Find the equation of motion for small oscillations about the position of equilibrium.
17. A plane triatomic molecule consists of equal masses m at vertices of an equilateral triangle of sides a . Assume the molecule is held together by forces that are harmonic for small oscillations and that the force constants are identical and equal to k . Allow motion only in the plane of the molecule.
 - (a) Without writing the equations of motion, justify your reasoning on the number of normal modes of the system and how many of these modes have zero frequency.
 - (b) One of the normal modes corresponds to a symmetrical stretching of all three vertices of the molecule. Find the frequency of this mode.
18. A particle in an isotropic three-dimensional harmonic oscillator potential has a natural frequency of ω_0 . Assume the particle is charged and that crossed static electric and magnetic fields are applied. Find the vibration frequencies with these electromagnetic fields present. Discuss the results for the limits of strong and weak fields.
19. Show for the case $V_{11} > V_{22} > 0$ and $V_{12} = V_{21} = 0$ in Eq. (6.27) that there are two normal modes with frequencies $\omega_1 = (V_{11})^{1/2}$ and $\omega_2 = (V_{22})^{1/2}$. Reintroduce the mass factor m and describe a physical system that would show this behavior for small oscillations.
20. Write the Lagrangian for the case $V_{11} = V_{22} = 0$ and $V_{12} = V_{21} > 0$ for the example discussed in Eqs. (6.27) to (6.30). Show there is one normal mode of simple harmonic motion with the frequency $\omega_1 = (V_{12})^{1/2}$, and a second mode in which the particle is unbound, receding exponentially to infinity for long time $t > \tau$ in accordance with the expression $e^{-t/\tau}$, where the parameter τ is given by $\tau = (V_{12})^{-1/2}$. For this unbounded mode, how does the distance depend upon time when $t < \tau$? What is the nature of the point $x_1 = x_2 = 0$? Restate your results with the mass parameter m included explicitly.
21. Write the Lagrangian discussed in Eqs. (6.27) to (6.30) in polar coordinates for the case $V_{11} = V_{22} > 0$ and $V_{12} = V_{21} = 0$. Show that there is a radial normal mode $r = r_0 \cos(\omega t)$ with frequency $\omega = (V_{11})^{1/2}$ when the angular momentum is zero. Show that in the case of nonzero angular momentum, the angular momentum is conserved and the particle can no longer reach $r = 0$. Write the fictitious potential energy $V'(r)$ (Chapter 3) for nonzero angular momentum. When finished, reintroduce the mass parameter, m , into all equations.
22. Repeat Exercise 21 with the conditions $V_{11} > V_{22} < 0$ and $V_{12} = V_{21} = 0$ and discuss your results in terms of the effective potential energy of Chapter 3.
23. Make a full analysis of the example discussed in Eqs. (6.27) to (6.30).

CHAPTER

7

The Classical Mechanics of the Special Theory of Relativity

At the end of the nineteenth century, the physics community had two incompatible descriptions of phenomena, Newtonian mechanics and Maxwellian electromagnetic theory. Newtonian mechanics assumed that all inertial frames were equivalent, while Maxwell's wave equations gave a universal speed of light that was the same in all inertial frames. Albert Einstein developed the special theory of relativity to replace Newtonian mechanics with a theory that was consistent with electromagnetic theory. After a brief historical survey, we shall review the assumptions of the special theory and the consequences of these assumptions. We shall then examine the formalism of the geometric picture of spacetime that results. Lastly, we develop a Lagrangian formalism and study attempts to express the results in a proper relativistic form.

In Newtonian mechanics, a set of well-verified laws applies in an inertial frame of reference defined by the first law. Any frame moving at constant velocity with respect to an inertial frame is also an inertial frame. Consider two frames denoted by S and S' with (t, x, y, z) and (t', x', y', z') the coordinates in S and S' , respectively. Without loss of generality, we assume the coordinate axes are aligned, x along x' , and so on. Let S' be moving relative to S in the $+x$ -direction at a speed v , as shown in Figure 7.1.

Newtonian mechanics assumes the spacetime coordinates in S are related to those in S' by the simple expressions

$$\begin{aligned} t' &= t \\ x' &= x - vt \\ y' &= y \\ z' &= z. \end{aligned} \tag{7.1}$$

Transformations of this type are called *Galilean transformations*. Under this assumption, it follows that Newton's second law,

$$\mathbf{F} = \frac{d}{dt} \mathbf{p},$$

relating the applied force, \mathbf{F} , and the momentum, \mathbf{p} , remains invariant, and

$$\mathbf{F} = \mathbf{F}', \quad t = t', \quad \text{and} \quad \mathbf{p} = \mathbf{p}'. \tag{7.2}$$

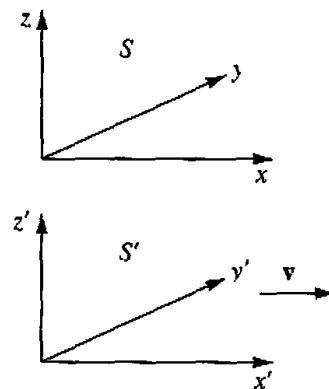


FIGURE 7.1 Galilean transformation from S to S' by a velocity v in the $+x$ -direction.

The time in both the S and S' frames is assumed to be ($t = t'$). The Newtonian world view is that the universe consists of three spatial directions and one time direction. All observers agree on the time direction up to a possible choice of units. Under these assumptions, there are no universal velocities. If \mathbf{u} and \mathbf{u}' are the velocities of a particle as measured in two frames moving with relative velocity \mathbf{v} as defined by Figure 7.1, then

$$\mathbf{u}' = \mathbf{u} - \mathbf{v}. \quad (7.3)$$

Maxwell's electromagnetic equations, on the other hand, have a universal constant (denoted by c), which is interpreted as the speed of light. Since this is inconsistent with Newtonian mechanics, either Newtonian or Maxwellian mechanics would have to be modified. After carefully thinking about how the universe would appear to an observer traveling at the speed of light, Albert Einstein decided that Maxwell's equations are correct to all inertial observers and the assumed transformations for Newtonian mechanics are incorrect. The correct transformations make the speed of light the same to all inertial observers.

7.1 ■ BASIC POSTULATES OF THE SPECIAL THEORY

Einstein used two postulates to develop what became known as the special theory:

1. The laws of physics are the same to all inertial observers.
2. The speed of light is the same to all inertial observers.

A formulation of physics that explicitly incorporates these two postulates is said to be *covariant*. Since the speed of light, c , is the same in all coordinate systems, it is reasonable to consider the numerical value of c as a conversion factor between the units used in measuring space and the units used in measuring time. So, $c dt$ is the time interval measured in the same units used to measure space units. In the SI system of units, $c dt$ has dimensions of meters. Many books

and articles on relativity set $c = 1$ and measure time and space in meters. In the material that follows, we shall show the explicit dependence upon c .

To satisfy the two postulates, the space and time of the special theory consist of a single entity that we refer to as *spacetime*. This spacetime is the geometric framework within which we perform physics. We cannot assume that all observers make the same division into time and space in the same way. The separation is unique to each inertial frame. The square of the distance in that spacetime, Δs^2 , between two points A and B is given by

$$(\Delta s)^2 = c^2(\text{time interval})^2 - (\text{space interval})^2, \quad (7.4)$$

where the interval is between the two points A and B . If the separation of the interval is assumed to be infinitesimal, the Δ is replaced by the differential symbol ds . Since a point in spacetime consists of a specification of three spatial coordinate values and one time value, the usual convention is to refer to a point in spacetime as an **event**. The term *event* is used because such a point has a definite location and a definite time in any frame.

The choice of opposite signs for the time and space intervals is intrinsic to the theory; however, the choice of a positive sign for $(c dt)^2$ is arbitrary. Some authors define a $(ds)^2$, which is the negative of the choice given in Eq. (7.4). All sign choices makes $(ds)^2 = 0$ according to the definition in Eq. (7.4) for light, since the space interval is $\pm(c \times \text{time interval})$. The choice made here for the relative signs used for space and time is such that real bodies moving at a velocity less than light have $(ds)^2 > 0$. This makes ds real for bodies moving slower than light speed. If $(ds)^2 > 0$, the interval is called *timelike*. If $(ds)^2 < 0$, the interval is called *spacelike*. Intervals for which $(ds)^2 = 0$ are called *lightlike* or *null*.

Since, to all inertial observers, objects that travel on timelike paths move less than the speed of light, they are called *tardyons*. Hypothetical bodies that always move faster than light are called *tachyons*, but such bodies will not concern us here. Objects moving at the speed of light are called *null* or *lightlike*.

In the limit of small displacements (differential displacements), Eq. (7.4) becomes, in a Cartesian coordinate system,

$$(ds)^2 = (c dt)^2 - (dx^2 + dy^2 + dz^2). \quad (7.4')$$

The four-dimensional space with an interval defined by Eqs. (7.4) or (7.4'), is often called *Minkowski* space to distinguish it from a four-dimensional Euclidean space for which there would be no minus sign in Eqs. (7.4) or (7.4'). The idea of using *ict* for the time coordinate to make the space Euclidean is no longer useful since it obscures the non-Euclidean nature of spacetime and makes the generalization to noninertial frames more difficult.

Since the interval between two events of spacetime is a geometric quantity, all inertial observers measure coordinates that preserve the value of the interval squared, $(ds)^2$. If S and S' are two different inertial frames, then

$$ds'^2 = ds^2. \quad (7.5)$$

Thus, $(ds)^2$ is called the square of the **invariant spacetime interval**. For this to be possible, the transformations between the coordinates in S' and those in S , must involve the relative velocity between the frames in both the space and the time parts; that is, the time coordinate can no longer stand independent of the transformation. This means the relative splitting of spacetime into space and time will be different for different inertial observers. Since the time measured in a laboratory frame is different from that measured by an observer at rest with respect to the body under study, we must distinguish these times. We distinguish them by calling the time measured by clocks at rest with respect to a body the *proper time*, while the other inertial observer uses a time that is often called *laboratory time*.

As a special case of Eq. (7.4), consider the relation between the proper time, τ , measured by an observer at rest with respect to an object in frame S' with coordinates (τ, x', y', z') , which is moving at a velocity, \mathbf{v} , with respect to a laboratory frame S with coordinates (t, x, y, z) . In the rest frame of the object, there is no motion, so Eqs. (7.4') and (7.5) give

$$c^2(d\tau)^2 = c^2(dt)^2 - v^2(dt)^2 = c^2(dt)^2 \left(\frac{1 - v^2}{c^2} \right)$$

or

$$dt = \frac{d\tau}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (7.6)$$

Since Eq. (7.6) makes $d\tau < dt$, this effect on dt is called “time dilation”: moving clocks appear to run slower.

The invariance of the interval expressed in Eq. (7.5), naturally divides spacetime into four regions, sketched in Fig. 7.2 relative to any event A at time t_A (A is located at $x = y = t = 0$ in Figure 7.2). If an event B at time t_B is such that $(ds_{AB})^2 > 0$, then all inertial observers will agree on the time order of the events A and B . It is even possible to choose an inertial frame where B has the same space coordinates as A . If t_B is less than t_A in one inertial frame, then t_B is less than t_A in all inertial frames. We call this region the *past*. Likewise, there is a region called the *future* where for event C (shown in Figure 7.2), t_C is greater than t_A for all inertial observers. Both the past and the future could be causally related to the event A . For any event inside the light cone, there exists a frame in which that event and the origin have the same x, y, z coordinates.

If $(ds_{AD})^2 < 0$, then there exist a set of inertial frames in which the relative order of t_A and t_D can be reversed or even made equal. This region has sometimes been referred to as the *elsewhere*, or as the *elsewhen*. In the region in which event D is located, there exists an inertial frame S' with its origin at event the A in which D is at the same time as A (but somewhere else). There also exist frames in which the time of D occurs before A and frames in which the time of D is after event A . Separating the past-future and the elsewhere is the null or **light cone**, where $ds^2 = 0$. The null cone is the set of spacetime points from which emitted

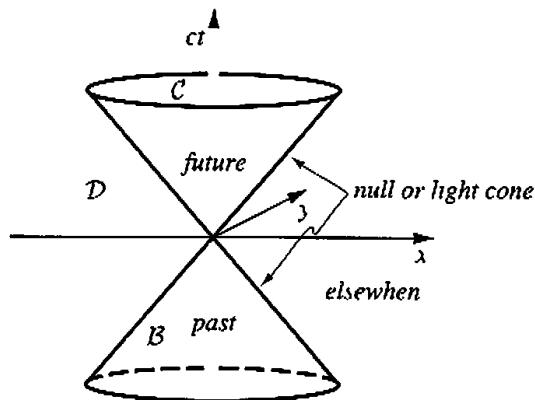


FIGURE 7.2 The three dimensions (ct , x , and y) of the light cone. The third spatial dimension has been suppressed. The event A referenced in the text is located at $x = y = ct = 0$. The light cone is the set of (ct, x, y) traced out by light emitted from $ct = x = y = 0$ or by light that reaches $x = y = 0$ at time $ct = 0$. The past and future lie inside the light cone. This figure is of necessity misleading because all points on the light cone have zero separation in spacetime.

light could reach event A , and those points from which light emitted from event A could reach. Any interval between the origin and a point inside the light cone is timelike, and any interval between the origin to a point outside the light cone is spacelike. Understanding the implication of the division of spacetime by the light cone is usually all that is needed to resolve the apparent paradoxes of the special theory.

7.2 ■ LORENTZ TRANSFORMATIONS

The simplest set of transformations that preserve the invariance of the interval, ds^2 , are called the *Lorentz transformations*. These transformations are simplest in the sense that they are linear in the coordinates and as the relative velocity goes to zero, the transformations become identity transformations. If we consider parallel Cartesian coordinate systems, S and S' , whose origins coincide at $t = t' = 0$, and whose relative velocity is v along the x axis as measured by S , and define

$$\beta = \frac{v}{c}, \quad \text{and} \quad \gamma = \frac{1}{\sqrt{1 - \beta^2}}, \quad (7.7)$$

then the following four equations relate the two sets of coordinates

$$ct' = \frac{ct - \beta x}{\sqrt{1 - \beta^2}} = \gamma(ct - \beta x) \quad (7.8a)$$

$$x' = \frac{x - \beta ct}{\sqrt{1 - \beta^2}} = \gamma(x - \beta ct) \quad (7.8b)$$

$$y' = y \quad (7.8c)$$

$$z' = z. \quad (7.8d)$$

Here we are only interested in transformations for which $t' \rightarrow t$ and $x' \rightarrow x$ as $\beta \rightarrow 0$. As matrices, these transformations appear as

$$\begin{bmatrix} ct' \\ x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} ct \\ x \\ y \\ z \end{bmatrix}. \quad (7.8')$$

In the limit of $\beta \ll 1$, Eqs. (7.8) reduce to the Galilean transformations as expected.

The generalization to arbitrary orientation of the velocity relative to the axes is straightforward. Since we are considering spacetime a four-dimensional entity, we would expect to deal with four-dimensional vectors. Using the notation $(ct, x, y, z) = (ct, \mathbf{r})$ allows the writing of the generalization of Eqs. (7.8') to the case where \mathbf{v} is not parallel to an axis, as

$$\begin{aligned} ct' &= \gamma(ct - \boldsymbol{\beta} \cdot \mathbf{r}) \\ \mathbf{r}' &= \mathbf{r} + \frac{(\boldsymbol{\beta} \cdot \mathbf{r})\boldsymbol{\beta}(\gamma - 1)}{\beta^2} - \boldsymbol{\beta}\gamma ct, \end{aligned} \quad (7.9)$$

provided the two sets of axes are aligned. Another way to express this arbitrary velocity is to consider the Lorentz transformation between two inertial coordinate systems with aligned axes, as a matrix transformation relating the two 4-quantities, $\mathbf{x} = (ct, \mathbf{r})$ and $\mathbf{x}' = (ct', \mathbf{r}')$, where

$$\mathbf{x}' = \mathbf{L}\mathbf{x} \quad (7.10)$$

We treat \mathbf{x}' and \mathbf{x} as column matrices and \mathbf{L} as the symmetric matrix

$$\mathbf{L} = \begin{bmatrix} \gamma & -\gamma\beta_x & -\gamma\beta_y & -\gamma\beta_z \\ -\gamma\beta_x & 1 + (\gamma - 1)\frac{\beta_x^2}{\beta^2} & (\gamma - 1)\frac{\beta_x\beta_y}{\beta^2} & (\gamma - 1)\frac{\beta_x\beta_z}{\beta^2} \\ -\gamma\beta_y & (\gamma - 1)\frac{\beta_x\beta_y}{\beta^2} & 1 + (\gamma - 1)\frac{\beta_y^2}{\beta^2} & (\gamma - 1)\frac{\beta_y\beta_z}{\beta^2} \\ -\gamma\beta_z & (\gamma - 1)\frac{\beta_x\beta_z}{\beta^2} & (\gamma - 1)\frac{\beta_z\beta_y}{\beta^2} & 1 + (\gamma - 1)\frac{\beta_z^2}{\beta^2} \end{bmatrix}. \quad (7.11)$$

This reduces to the results given in Eqs. (7.8') when $\beta_x = \beta$, $\beta_y = \beta_z = 0$.

These transformations map the origin of S and the origin of S' to $(0, 0, 0, 0)$. Hence the coordinates of both origins correspond to the same location in spacetime. If this is not desired, there is a more general transformation of the form

$$\mathbf{x}' = \mathbf{L}\mathbf{x} + \mathbf{a} \quad (7.12)$$

where \mathbf{L} is a spacetime rotation (boost) and \mathbf{a} is a spacetime translation. This is the *Poincaré transformation* or the *inhomogeneous Lorentz transformation*. We shall consider only homogeneous transformations for which \mathbf{a} of Eq. (7.12) is zero.

7.3 ■ VELOCITY ADDITION AND THOMAS PRECESSION

The most general homogeneous Lorentz transformation will involve both a velocity change and a rotation of the coordinates. The velocity transformation is termed a boost and has the form of Eq. (7.11). Any homogeneous Lorentz transformation, \mathbf{L} , can be written as

$$\mathbf{L} = \mathbf{R}\mathbf{L}_0 = \mathbf{L}'_0\mathbf{R}' \quad (7.13)$$

where \mathbf{R} is a rotation matrix as discussed in Chapter 4, and \mathbf{L}_0 , which is called a restricted or proper Lorentz transformation, corresponds to a pure boost. The restricted Lorentz transformations form a representation of the Lorentz group.* Since \mathbf{R} is not symmetric and \mathbf{L}_0 is symmetric, \mathbf{L} will, in general, have no symmetry. Also, since \mathbf{L}_0 and \mathbf{R} are matrices, $\mathbf{R}\mathbf{L}_0 \neq \mathbf{L}_0\mathbf{R}$. There will exist two other transformations \mathbf{L}'_0 and \mathbf{R}' such that $\mathbf{R}\mathbf{L}_0 = \mathbf{L}'_0\mathbf{R}'$.

For any Lorentz transformation, \mathbf{L} , there is an inverse transformation, \mathbf{L}^{-1} , such that

$$\mathbf{L}\mathbf{L}^{-1} = \mathbf{L}^{-1}\mathbf{L} = \mathbf{1}, \quad (7.14)$$

where $\mathbf{1}$ is the diagonal unit 4×4 matrix with elements $\delta_{\alpha\beta}$. The existence of an inverse places four constraints on the diagonal element and six on the off-diagonal elements for a total of ten constraints on the Lorentz transformation. There are then only six independent components. Three of these correspond to the components of the relative velocity vector and three correspond to the Euler angles of the rotation (see Section 4.4).

Consider three inertial systems, S_1 , S_2 , and S_3 , with x axes aligned. Let S_2 be moving at a velocity v along the common x -direction with respect to S_1 and let S_3 be moving at velocity v' along the common x -direction with respect to S_2 . The Lorentz transformation from S_1 to S_3 is given by

$$\begin{aligned} \mathbf{L}_{1-3} &= \begin{bmatrix} \gamma' & -\gamma'\beta' & 0 & 0 \\ -\gamma'\beta' & \gamma' & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \\ &= \begin{bmatrix} \gamma\gamma'(1+\beta\beta') & -\gamma\gamma'(\beta+\beta') & 0 & 0 \\ -\gamma\gamma'(\beta+\beta') & \gamma\gamma'(1+\beta\beta') & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \end{aligned}$$

*Group concepts are discussed in Appendix B.

where Eq. (7.7) defines β and γ for v and β' and γ' for v' . Let β'' be the speed of S_3 relative to S_1 and γ'' the associated factor, then since L_{1-3} can be written as a single Lorentz transformation with a velocity β'' with its associated γ'' as

$$L_{1-3} = \begin{bmatrix} \gamma'' & -\gamma''\beta'' & 0 & 0 \\ -\gamma''\beta'' & \gamma'' & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},$$

and, since these two forms of L_{1-3} must be the same, we have

$$\beta'' = \frac{\beta + \beta'}{1 + \beta\beta'} \quad (7.15)$$

This is the relativistic addition of velocity formula for parallel velocities.

The product of any two transformations, L_1 and L_2 is itself a Lorentz transformation, L_3 . Such a Lorentz transformation will, in general, involve not only a boost, but may also include a rotation of coordinate axes. If both L_1 and L_2 are pure boosts but their two velocities are not parallel, L_3 will involve a rotation in addition to a boost. This rotation is called the *Thomas precession* rotation. The usual form for the Thomas precession assumes the second boost, L_2 has a velocity small compared to the first boost, L_1 and also that it is small compared to the speed of light. For example, the Thomas precession can be observed for a gyroscope orbiting the Earth or for electrons in atoms.

Consider three inertial frames S_1 , S_2 , and S_3 , with S_2 moving at a velocity β with respect to S_1 and S_3 moving at a velocity of β' with respect to S_2 . Without loss of generality, we can arrange the axes of S_1 so that β is along the x axis of S_1 and β' lies in the $x'y'$ plane of S_2 ; that is, β , β' define the $x'y'$ plane of S_2 . Let L represent the transformation from S_1 to S_2 and L' the transformation from S_2 to S_3 with γ and γ' associated with β and β' . Then from Eq. (7.11),

$$L = \begin{pmatrix} \gamma & -\gamma\beta & 0 & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7.16)$$

and

$$L' = \begin{bmatrix} \gamma' & -\gamma'\beta'_x & -\gamma'\beta'_y & 0 \\ -\gamma'\beta'_x & 1 + (\gamma' - 1)\frac{\beta'^2}{\beta'^2} & (\gamma' - 1)\frac{\beta'_x\beta'_y}{\beta'^2} & 0 \\ -\gamma'\beta'_y & (\gamma' - 1)\frac{\beta'_x\beta'_y}{\beta'^2} & 1 + (\gamma' - 1)\frac{\beta'^2}{\beta'^2} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7.17)$$

We assume that the components of β' are small and only need be retained to first order giving via matrix multiplications of Eq. (7.16) and Eq. (7.17)

$$\mathbf{L}'' = \mathbf{L}'\mathbf{L} = \begin{bmatrix} \gamma\gamma' & -\gamma\gamma'\beta & -\gamma'\beta'_y & 0 \\ -\gamma\beta & \gamma & 0 & 0 \\ -\gamma\gamma'\beta'_y & \gamma\beta\gamma'\beta'_y & \gamma' & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (7.18)$$

Since \mathbf{L}'' is not symmetric, it must correspond to a rotation and a boost. We shall write the velocity of S_3 as observed by S_1 as β'' .

Since the off-diagonal elements corresponding to the z axis are zero, this rotation is about an axis perpendicular to the xy plane. The boost from S_1 to S_3 is denoted by β'' , and we assume that β' is small compared to β and also small compared to the speed of light ($\gamma' \approx 1$). Then, to first order, the nonvanishing components of β'' are (Since the velocity perpendicular to x is small we can ignore to first order the distinction among y , y' , and y'')

$$\beta''_x = \beta, \quad \beta''_y = \frac{\beta'_y}{\gamma}, \quad \beta''^2 = \beta^2, \quad \text{and} \quad \gamma'' = \gamma, \quad (7.19)$$

and Eq. (7.18) becomes

$$\mathbf{L}'' \approx \begin{bmatrix} \gamma'' & -\gamma''\beta''_x & -\gamma''\beta''_y & 0 \\ -\gamma''\beta''_x & \gamma'' & 0 & 0 \\ -\gamma''\beta''_y & \gamma''\beta''_x\beta''_y & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7.18')$$

In this approximation, a pure Lorentz transformation from S_3 to S_1 (the inverse transformation) would correspond to a large boost in the x'' axis of $-\beta''_x$ and a small boost in the y'' axis of $-\beta''_y$. The Lorentz boost for that transformation

$$\mathbf{L}_{3-1} = \begin{bmatrix} \gamma'' & \gamma''\beta''_x & \gamma''\beta''_y & 0 \\ \gamma''\beta''_x & \gamma'' & (\gamma'' - 1)\frac{\beta''_y}{\beta''} & 0 \\ \gamma''\beta''_y & (\gamma'' - 1)\frac{\beta''_y}{\beta''} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7.20)$$

Finally, the rotation matrix induced by the rotation from S_1 to S_3 , after some algebraic simplification and the dropping of higher-order terms in β'' , is found to be

$$\mathbf{R} = \mathbf{L}''\mathbf{L}_{3-1} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & (\gamma - 1)\frac{\beta''_y}{\beta} & 0 \\ 0 & -(\gamma - 1)\frac{\beta''_y}{\beta} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \quad (7.21)$$

Comparison with Eq. (4.44) shows that \mathbf{R} implies S_3 is rotated with respect to S_1 about the z axis through an infinitesimal angle:

$$\Delta\Omega = (\gamma - 1) \frac{\beta_y''}{\beta} = \beta_y'' \beta \left(\frac{\gamma - 1}{\beta^2} \right). \quad (7.22)$$

The spatial rotation resulting from the successive application of two nonparallel Lorentz transformations has been declared every bit as paradoxical as the more frequently discussed apparent violations of common sense, such as the so-called "twin paradox." But the present apparent paradox has important applications, especially in atomic physics, and therefore has been abundantly verified experimentally.

Consider a particle moving in the laboratory system with a velocity \mathbf{v} that is not constant. Since the system in which the particle is at rest is accelerated with respect to the laboratory, the two systems should not be connected by a Lorentz transformation. We can circumvent this difficulty by a frequently used stratagem (elevated by some to the status of an additional postulate of relativity). We imagine an infinite number of inertial systems moving uniformly relative to the laboratory system, one of which instantaneously matches the velocity of the particle. The particle is thus instantaneously at rest in an inertial system that can be connected to the laboratory system by a Lorentz transformation. It is assumed that this Lorentz transformation will also describe the properties of the particle and its true rest system as seen from the laboratory system.

Suppose now that S_1 is the laboratory system, while S_2 and S_3 are two of the instantaneous rest systems a time Δt apart in the particle's motion. By Eq. (7.22), the laboratory observer will see a change in the particle's velocity in this time, $\Delta\mathbf{v}$, which has only a y -component $\beta_y''c = \Delta v$. Since the initial x axis has been chosen along the direction of $\mathbf{v} = \beta c$, the vector of the infinitesimal rotation in this time can be written as

$$\Delta\boldsymbol{\Omega} = -(\gamma - 1) \frac{\mathbf{v} \times \Delta\mathbf{v}}{v^2} \quad (7.23)$$

Hence, if the particle has some specific direction attached to it (such as a spin vector), it will be observed from the laboratory system that this direction precesses with an angular velocity

$$\boldsymbol{\omega} = \frac{d\boldsymbol{\Omega}}{dt} = -(\gamma - 1) \frac{\mathbf{v} \times \mathbf{a}}{v^2} \quad (7.24)$$

where \mathbf{a} is the particle's acceleration as seen from S_1 . Equation (7.24) is frequency encountered in the form it takes when v is small enough that γ can be approximated (using $\gamma \approx 1 + \frac{1}{2}\beta^2$) as

$$\boldsymbol{\omega} = \frac{1}{2c^2} (\mathbf{a} \times \mathbf{v}). \quad (7.25)$$

In either form, $\boldsymbol{\omega}$ is known as the *Thomas precession frequency*.

7.4 ■ VECTORS AND THE METRIC TENSOR

We will use the notation that the coordinates, which need not be Cartesian, are written as x^μ where $x^0 = ct$ is the time coordinate, and x^1, x^2, x^3 are the space coordinates. This change in notation is needed to be consistent with the developments in the following sections.

Consider an arbitrary one-dimensional curve in 4-dimensional spacetime, \mathcal{P} , described by a parameter λ , where for a given λ the coordinates of a point of the curve can be written as $x^0(\lambda), x^1(\lambda), x^2(\lambda), x^3(\lambda)$. In introductory texts a 4-vector, v , is defined by this curve as an arrow whose tail is located at an event \mathcal{A} on the curve and whose head is at an event \mathcal{B} on the curve where $v_{AB} = \mathcal{P}_B - \mathcal{P}_A$. However, instead of defining the vector at two points, we can use the parameter λ , which is a measure of the length along the curve from \mathcal{A} to \mathcal{B} , by writing

$$v_{AB} = \left(\frac{d\mathcal{P}}{d\lambda} \right)_{\lambda=0}. \quad (7.26)$$

Such a 4-vector is a *tangent vector* to the curve. We adopt the notation that the components of vectors are written with superscripts such as v^0, v^1, v^2, v^3 . In spite of the way we draw tangent vectors, they do not have any extension in spacetime. The arrows we draw simply help us visualize the vector. At each point along the curve, the tangent vector has a direction and a magnitude. For curves that are timelike, the proper time, τ , is usually chosen as the parameter λ . The laboratory coordinates are then $x^0 = ct(\tau), x^1 = x(\tau), x^2 = y(\tau), x^3 = z(\tau)$, and the tangent to the curve is the *four-velocity*, u , of a particle traveling along the curve \mathcal{P} . Equation (7.26) becomes

$$u^0 = \frac{dct}{d\tau} = \gamma c, \quad u^i = \frac{dx^i}{d\tau} = \gamma v^i \quad (7.27)$$

where $v^i = dx^i/dt$ is the normal three-velocity with $v^2 = (v^x)^2 + (v^y)^2 + (v^z)^2$. We shall assume that Greek letters can take on the values 0–3 and Latin letters the values 1–3. Repeated indices are summed. Since the 4-velocity of a particle is defined over a range of the parameter λ , there is an infinite set of 4-velocities for the particle, one for each value of λ . Such a set of vectors is termed a *vector field*. Some common examples of vector fields are given in Table 7.1.

We assume that the components of any 4-vector can be expressed by the values of the vector's projections along a set of basis vectors, e_0, e_1, e_2, e_3 , and that the coordinates are measured along the direction given by the basis vectors. Such a system is called a *coordinates basis*.* Cartesian, spherical, and cylindrical coordinate systems, among many possible systems, can have such a basis set. The position of a point on the curve $\mathcal{P}(\tau)$ can be written as

$$\mathcal{P}(\tau) = x^\mu(\tau)e_\mu, \quad (7.28)$$

*The choice of a coordinate basis is arbitrary but avoids some complications. For this introductory chapter we will assume that each basis vector lies in the direction of its increasing coordinate.

TABLE 7.1 Examples of Vector Fields

Name	Time Portion	Space Portion	(Magnitude) ²	Type
Coordinate	ct	\mathbf{r}	$c^2 t^2 - \mathbf{r}^2$	spacelike, null, or timelike
Velocity	γc	$\gamma \mathbf{v}$	c^2	timelike
Momentum	$\frac{E}{c}$	\mathbf{p}	$m^2 c^2$	timelike
Force	$\frac{\gamma}{c} \frac{dE}{dt}$	$\gamma \frac{d\mathbf{p}}{dt} = \gamma \mathbf{F}$	$-(\mathbf{F}_{\text{Newtonian}})^2$	spacelike
Current density	$\gamma \rho c$	$\gamma \mathbf{j}$	$\rho^2 c^2$	timelike

where repeated Greek indices, one raised and one lowered, are summed from 0 to 3. In particular, the 4-velocity given in Eq. (7.27) becomes

$$u = \frac{d\mathcal{P}}{d\tau} = \frac{dx^\mu}{d\tau} \mathbf{e}_\mu = u^\mu \mathbf{e}_\mu. \quad (7.29)$$

The magnitude of the 4-velocity is a scalar whose values can vary as we change λ . This set of magnitudes is an example of a *scalar field*. To convert a 4-vector field to a scalar field, we need what is called a functional,* which can convert a pair of vectors into a scalar function at each point in spacetime. In other words, we wish to define the scalar product of two vectors or vector fields. This conversion of a 4-vector field (or two different vector fields) to a scalar field is an example of a *mapping*. If both the vectors are the same, then this scalar would be the square of the length of the vector, and when the vectors are different, it is called the scalar product of the vectors. Such a functional is called the *metric tensor*, g .† The metric tensor functional can be considered as a machine with two slots into which you can insert two vectors to produce a scalar (real-valued function). That is,

$$g(u, v) = g(v, u) = u \cdot v, \quad (7.30)$$

is the scalar product. In particular if the basis vectors are inserted into the metric,

$$g_{\alpha\beta} = g(\mathbf{e}_\alpha, \mathbf{e}_\beta) = \mathbf{e}_\alpha \cdot \mathbf{e}_\beta. \quad (7.31)$$

The $g_{\alpha\beta}$ are the components of the metric tensor associated with the basis vectors \mathbf{e}_α . For example, consider a two-dimensional Minkowski space with coordinates ct and x and a vector $v = (a, b)$. Then $g(v, v) = a^2 - b^2$ and $g_{00} = 1$, $g_{11} = -1$.

The form of the $g_{\alpha\beta}$ is defined by the form for the interval. This suggests that we consider small displacements. If the relative displacement vector between two

*A functional is a function whose arguments are themselves functions.

†We use the same notation for tensors in 4-space as we do for 4-vectors.

points is small, it can be written as

$$d\xi = \Delta x^\alpha e_\alpha. \quad (7.32)$$

Recasting Eq. (7.32) in the language of Eq. (7.4'), we see for Minkowski coordinates

$$\begin{aligned} (\Delta s)^2 &= d\xi \cdot d\xi = \Delta x^\alpha \Delta x^\beta e_\alpha \cdot e_\beta = g_{\alpha\beta} \Delta x^\alpha \Delta x^\beta \\ &= (c\Delta t)^2 - (\Delta x)^2 - (\Delta y)^2 - (\Delta z)^2. \end{aligned}$$

In the limit of infinitesimal displacements this can be written as

$$ds^2 = g_{\alpha\beta} dx^\alpha dx^\beta, \quad (7.32')$$

which holds for any metric tensor. The metric tensor for a Minkowski coordinate system, using the +--- sign convention, has the following tensor representation*

$$g = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}. \quad (7.33)$$

The scalar product of two vectors in this coordinate system is

$$u \cdot v = u^\alpha v^\beta g_{\alpha\beta} = u^0 v^0 - u^1 v^1 - u^2 v^2 - u^3 v^3. \quad (7.34)$$

It is straightforward to show that in any coordinate system, the square of the magnitude of the four-velocity is

$$u \cdot u = c^2. \quad (7.35)$$

The 4-momentum can be defined from Eq. (7.27)

$$p = mu, \quad (7.36)$$

where the mass, m , is a scalar. So the length squared of the four-momentum is

$$p \cdot p = m^2 c^2, \quad (7.37)$$

or from Eqs. (7.27) and (7.34),

$$p \cdot p = m^2 c^2 = m^2 c^2 \gamma^2 - m^2 v^2 \gamma^2 = \frac{E^2}{c^2} - \mathbf{p}^2 \quad (7.38)$$

*The notation used for the display of a matrix is [], while for tensors () will be used as it was in Chapter 5. Matrices are used for relating different coordinate frames while tensors are physical geometric objects.

where \mathbf{p} is the length of the 3-momentum. This last form of Eq. (7.38) is often written as

$$E^2 = m^2 c^4 + \mathbf{p}^2 c^2. \quad (7.38')$$

The relativistic kinetic energy, T , is defined as

$$T = E - mc^2 = mc^2(\gamma - 1) \quad (7.39)$$

$$= \sqrt{(mc^2)^2 + \mathbf{p}^2 c^2} - mc^2. \quad (7.39')$$

For $\beta \ll 1$, a power series expansion gives

$$T = \frac{1}{2}mv^2 + O(\beta^4). \quad (7.40)$$

Since $\mathbf{p} = m\gamma \mathbf{v}$, Eq. (7.39) shows that the kinetic energy of a body with finite rest mass tends to infinity as the speed approaches that of light (as $\beta \rightarrow 1$, $\gamma \rightarrow \infty$). In other words, it takes an infinite amount of energy to increase the speed of a mass particle (or a space ship) from any velocity less than c to c itself. This is another proof that it is impossible to attain or exceed the speed of light starting from any finite speed less than c .

7.5 ■ 1-FORMS AND TENSORS*

Suppose we insert only one 4-vector into the metric tensor in Eq. (7.30). We would produce an object that could be written as $u_\alpha = g_{\alpha\beta} u^\beta$. For example, in the two-dimensional Minkowski space, if \mathbf{u}^α has components (a, b) , then u_α has components $(a, -b)$. This geometric object, u_α , is called a *1-form* or, in an older notation, a *covariant vector*. In the older notation the vector itself was called a *contravariant vector*. If the vector is thought of as a directed line, the 1-form is a set of numbered surfaces through which the vector passes as is shown in Fig. 7.3. It is another functional (machine) similar to g , except it converts a vector to a linear real-valued scalar function. That is, if η is a 1-form (field) and v is some vector (field), the quantity denoted by $\langle \eta, v \rangle$ is a number that tells us how many surfaces of η are pierced by v . For each vector field V , there is an associated 1-form, V_η such that $\langle V_\eta, V \rangle = V \cdot V$ is the scalar contraction or the square of the magnitude of V .

The gradient is an example of a 1-form since, if we consider a curve \mathcal{P} , parameterized by λ , where $\lambda = 0$ at \mathcal{P}_0 and take a scalar function, f , defined along the curve,

$$\partial_v f = \frac{\partial}{\partial \lambda}_{\lambda=0} f(\mathcal{P}(\lambda)) = \frac{df}{d\lambda}_{\mathcal{P}_0} = v^\alpha \frac{\partial f}{\partial x^\alpha}. \quad (7.41)$$

*The material in Sections 7.5 and 7.6 is not needed for Section 7.7. The Section order has been chosen for continuity of ideas.

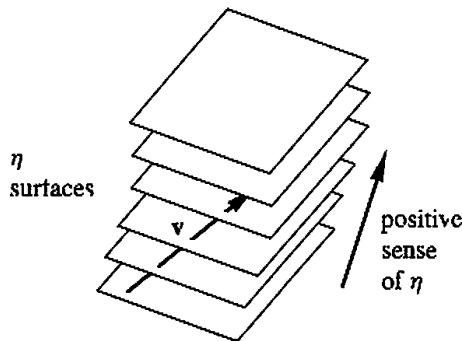


FIGURE 7.3 A vector v between two neighboring points and a 1-form η . The piercing of η by v produces a number given by $\langle \eta, v \rangle$, the number (including fractions) of surfaces pierced

So

$$\partial_\alpha = \partial_{e_\alpha} = \frac{\partial}{\partial x^\alpha}. \quad (7.42)$$

We often write either ∂_v or d to indicate the gradient of a scalar. Several examples of vectors, 1-forms, scalar products, and metrics from relativity and other areas of physics are given in Table 7.2.

The gradient of the coordinates, ω^α , defined as

$$\omega^\alpha = dx^\alpha, \quad (7.43)$$

provides a set of **basis 1-forms** since

$$\langle \omega^\alpha, e_\beta \rangle = \delta_\beta^\alpha, \quad (7.44)$$

TABLE 7.2 Examples of Vectors and 1-forms

SYSTEM	Vectors: (Contravariant Components)	1-forms. (Covariant Components)	Scalar Contraction	Metric
Euclidean Cartesian (x, y, z)	(dx, dy, dz)	(dx, dy, dz)	$dx^2 + dy^2 + dz^2$	$\begin{matrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{matrix}$
Euclidean Spherical	$(dr, d\theta, d\phi)$	$(dr, r^2 d\theta, r^2 \sin^2 \theta d\phi)$	$dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2$	$\begin{matrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2 \sin^2 \theta \end{matrix}$
Solid-state	r (lattice vector)	k (reciprocal vector)	$r \cdot k$	varies
Quantum theory	$ i\rangle$ (ket)	$\langle j $ (bra)	$\langle j i\rangle$	$ i\rangle\langle j $
Special theory of relativity (Minkowski)	$(c dt, d\mathbf{r})$	$(c dt, -d\mathbf{r})$	$c^2 dt^2 - d\mathbf{r}^2$	$\begin{matrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{matrix}$

and any 1-form η can be written as

$$\eta = \eta_\alpha \omega^\alpha. \quad (7.45)$$

It follows that

$$\langle \eta, e_\alpha \rangle = \eta_\alpha \quad (7.46)$$

and for any vector, v

$$\langle \eta, v \rangle = \eta_\alpha v^\alpha. \quad (7.47)$$

This gives us two ways to calculate the scalar product of two vectors v and u . If we define the inverse metric by

$$g^{\alpha\beta} g_{\beta\gamma} = \delta_\gamma^\alpha \quad (7.48)$$

or in index-free notation by

$$g^{-1}g = gg^{-1} = 1, \quad (7.48')$$

we can convert vectors (u^α) to 1-forms (u_α) and conversely as

$$u_\alpha = g_{\alpha\beta} u^\beta \quad \text{and} \quad u^\alpha = g^{\alpha\beta} u_\beta. \quad (7.49)$$

We can therefore write for two 4-vectors u and v (or they could be two 1-forms),

$$u \cdot v = g(u, v) = g_{\alpha\beta} u^\alpha v^\beta = u^\alpha v_\alpha = u_\alpha v_\beta g^{\alpha\beta}. \quad (7.34')$$

Since each 1-form has a unique associated vector, we could use the same symbol for both. The difference is important only when considering components.

In terms of the two-dimensional example that we previously considered (Minkowski spacetime) with ct and x as the coordinates, if the vector u has components (a, b) and the vector v has components (c, d) , the last three terms of the preceding equation can be written as

$$\begin{aligned} g_{\alpha\beta} u^\alpha v^\beta &= (1)(a)(c) + (-1)(b)(d) = ac - bd, \\ u^\alpha v_\alpha &= (a)(c) + (b)(-d) = ac - bd, \end{aligned}$$

and

$$u_\alpha v_\beta g^{\alpha\beta} = (a)(c)(1) + (b)(d)(-1) = ac - bd.$$

It may help to consider the relationship between a vector and a 1-form from a more general point of view using the Minkowski two-dimensional space as an