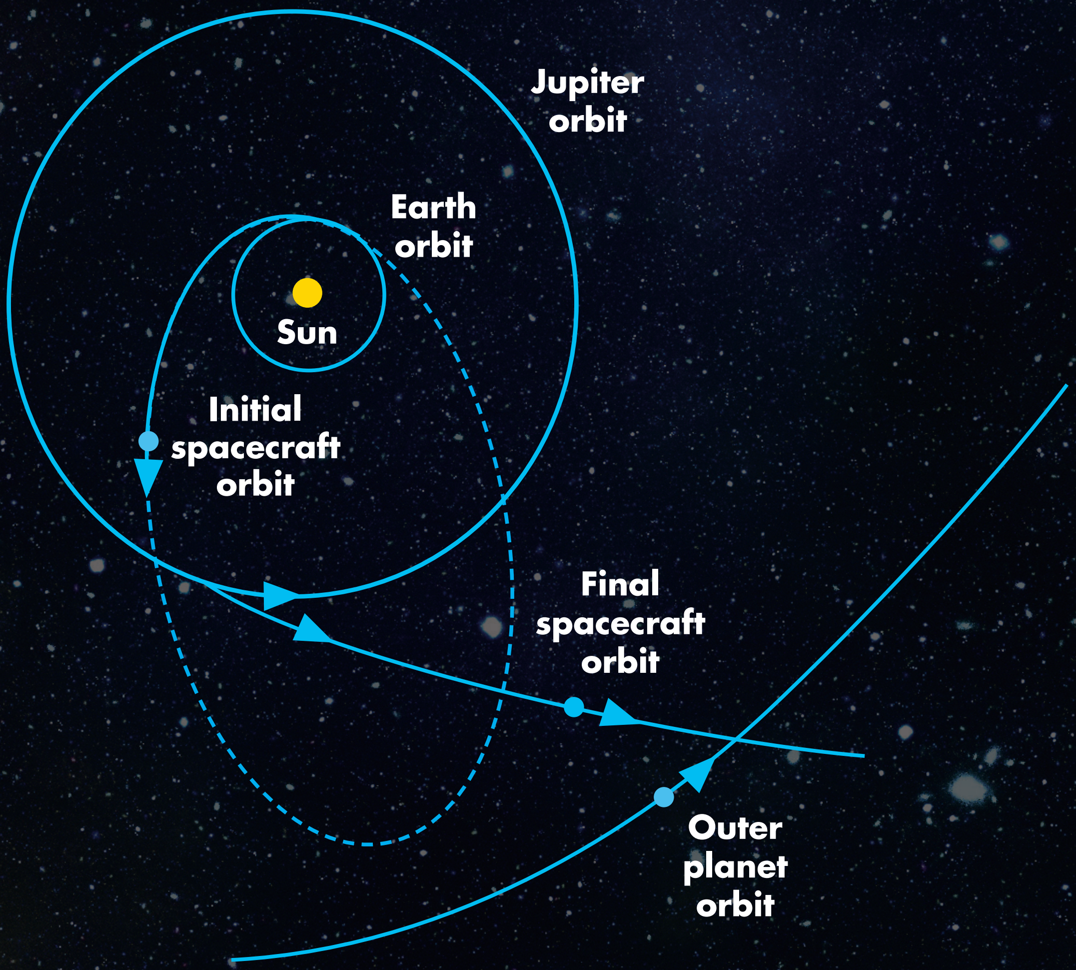
Classical Mechanics

Second Edition



Tai L. Chow

CRC Press
Taylor & Francis Group

Classical Mechanics

Second Edition

Classical Mechanics

Second Edition

Tai L. Chow

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Version Date: 20130227

International Standard Book Number-13: 978-1-4665-7000-9 (eBook - PDF)

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

To

Robert Youngme

and

David Lori

Contents

Preface xv
Author xvii

Chapter 1 Kinematics: Describing the Motion 1

1.1 Introduction 1
1.2 Space, Time, and Coordinate Systems 1
1.3 Change of Coordinate System (Transformation of Components of a Vector) 3
1.4 Displacement Vector 8
1.5 Speed and Velocity 8
1.6 Acceleration 10
1.6.1 Tangential and Normal Acceleration 11
1.7 Velocity and Acceleration in Polar Coordinates 14
1.7.1 Plane Polar Coordinates (r, θ) 14
1.7.2 Cylindrical Coordinates (ρ, ϕ, z) 15
1.7.3 Spherical Coordinates ($\vec{r}$, θ, ϕ) 16
1.8 Angular Velocity and Angular Acceleration 18
1.9 Infinitesimal Rotations and the Angular Velocity Vector 19

Chapter 2 Newtonian Mechanics 25

2.1 The First Law of Motion (Law of Inertia) 25
2.1.1 Inertial Frames of Reference 26
2.2 The Second Law of Motion; the Equations of Motion 27
2.2.1 The Concept of Force 28
2.3 The Third Law of Motion 32
2.3.1 The Concept of Mass 32
2.4 Galilean Transformations and Galilean Invariance 34
2.5 Newton's Laws of Rotational Motion 36
2.6 Work, Energy, and Conservation Laws 37
2.6.1 Work and Energy 38
2.6.2 Conservative Force and Potential Energy 39
2.6.3 Conservation of Energy 40
2.6.4 Conservation of Momentum 42
2.6.5 Conservation of Angular Momentum 42
2.7 Systems of Particles 46
2.7.1 Center of Mass 46
2.7.2 Motion of CM 48
2.7.3 Conservation Theorems 49
References 56

Chapter 3 Integration of Newton's Equation of Motion 57

3.1 Introduction 57
3.2 Motion Under Constant Force 58

3.3 Force Is a Function of Time 63
3.3.1 Impulsive Force and Green's Function Method 66
3.4 Force Is a Function of Velocity 67
3.4.1 Motion in a Uniform Magnetic Field 71
3.4.2 Motion in Nearly Uniform Magnetic Field 73
3.5 Force Is a Function of Position 74
3.5.1 Bounded and Unbounded Motion 75
3.5.2 Stable and Unstable Equilibrium 76
3.5.3 Critical and Neutral Equilibrium 78
3.6 Time-Varying Mass System (Rocket System) 79

Chapter 4 Lagrangian Formulation of Mechanics: Descriptions of Motion in Configuration Space 85

4.1 Generalized Coordinates and Constraints 85
4.1.1 Generalized Coordinates 85
4.1.2 Degrees of Freedom 85
4.1.3 Configuration Space 86
4.1.4 Constraints 86
4.1.4.1 Holonomic and Nonholonomic Constraints 86
4.1.4.2 Scleronomic and Rheonomic Constraints 88
4.2 Kinetic Energy in Generalized Coordinates 88
4.3 Generalized Momentum 90
4.4 Lagrangian Equations of Motion 91
4.4.1 Hamilton's Principle 91
4.4.2 Lagrange's Equations of Motion from Hamilton's Principle 92
4.5 Nonuniqueness of the Lagrangian 102
4.6 Integrals of Motion and Conservation Laws 104
4.6.1 Cyclic Coordinates and Conservation Theorems 104
4.6.2 Symmetries and Conservation Laws 106
4.6.2.1 Homogeneity of Time and Conservation of Energy 106
4.6.2.2 Spatial Homogeneity and Momentum Conservation 107
4.6.2.3 Isotropy of Space and Angular Momentum Conservation 108
4.6.2.4 Noether's Theorem 110
4.7 Scale Invariance 111
4.8 Nonconservative Systems and Generalized Potential 112
4.9 Charged Particle in Electromagnetic Field 112
4.10 Forces of Constraint and Lagrange's Multipliers 114
4.11 Lagrangian versus Newtonian Approach to Classical Mechanics 119
Reference 123

Chapter 5 Hamiltonian Formulation of Mechanics: Descriptions of Motion in Phase Spaces 125

5.1 The Hamiltonian of a Dynamic System 125
5.1.1 Phase Space 126
5.2 Hamilton's Equations of Motion 126
5.2.1 Hamilton's Equations from Lagrange's Equations 126
5.2.2 Hamilton's Equations from Hamilton's Principle 128

5.3 Integrals of Motion and Conservation Theorems 132
5.3.1 Energy Integrals 132
5.3.2 Cyclic Coordinates and Integrals of Motion 132
5.3.3 Conservation Theorems of Momentum and Angular Momentum 133
5.4 Canonical Transformations 135
5.5 Poisson Brackets 140
5.5.1 Fundamental Properties of Poisson Brackets 141
5.5.2 Fundamental Poisson Brackets 141
5.5.3 Poisson Brackets and Integrals of Motion 141
5.5.4 Equations of Motion in Poisson Bracket Form 144
5.5.5 Canonical Invariance of Poisson Brackets 144
5.6 Poisson Brackets and Quantum Mechanics 145
5.7 Phase Space and Liouville's Theorem 147
5.8 Time Reversal in Mechanics (Optional) 150
5.9 Passage from Hamiltonian to Lagrangian 151
References 154

Chapter 6 Motion Under a Central Force 155

6.1 Two-Body Problem and Reduced Mass 155
6.2 General Properties of Central Force Motion 157
6.3 Effective Potential and Classification of Orbits 159
6.4 General Solutions of Central Force Problem 163
6.4.1 Energy Method 163
6.4.2 Lagrangian Analysis 164
6.5 Inverse Square Law of Force 167
6.6 Kepler's Three Laws of Planetary Motion 172
6.7 Applications of Central Force Motion 174
6.7.1 Satellites and Spacecraft 174
6.7.2 Communication Satellites 178
6.7.3 Flyby Missions to Outer Planets 179
6.8 Newton's Law of Gravity from Kepler's Laws 182
6.9 Stability of Circular Orbits (Optional) 183
6.10 Apsides and Advance of Perihelion (Optional) 188
6.10.1 Advance of Perihelion and Inverse-Square Force 189
6.10.2 Method of Perturbation Expansion 190
6.11 Laplace–Runge–Lenz Vector and the Kepler Orbit (Optional) 192
References 198

Chapter 7 Harmonic Oscillator 199

7.1 Simple Harmonic Oscillator 199
7.1.1 Motion of Mass *m* on the End of a Spring 199
7.1.2 The Bob of Simple Pendulum Swinging through a Small Arc 200
7.1.3 Solution of Equation of Motion of SHM 201
7.1.4 Kinetic, Potential, Total, and Average Energies of Harmonic Oscillator 203
7.2 Adiabatic Invariants and Quantum Condition 206
7.3 Damped Harmonic Oscillator 209

7.4 Phase Diagram for Damped Oscillator 218
7.5 Relaxation Time Phenomena 220
7.6 Forced Oscillations without Damping 220
7.6.1 Periodic Driving Force 221
7.6.2 Arbitrary Driving Forces 223
7.7 Forced Oscillations with Damping 225
7.7.1 Resonance 227
7.7.2 Power Absorption 231
7.8 Oscillator Under Arbitrary Periodic Force 235
7.8.1 Fourier's Series Solution 236
7.9 Vibration Isolation 239
7.10 Parametric Excitation 241

Chapter 8 Coupled Oscillations and Normal Coordinates 249

8.1 Coupled Pendulum 249
8.1.1 Normal Coordinates 251
8.2 Coupled Oscillators and Normal Modes: General Analytic Approach 254
8.2.1 The Equation of Motion of a Coupled System 254
8.2.2 Normal Modes of Oscillation 255
8.2.3 Orthogonality of Eigenvectors 257
8.2.4 Normal Coordinates 259
8.3 Forced Oscillations of Coupled Oscillators 264
8.4 Coupled Electric Circuits 266

Chapter 9 Nonlinear Oscillations 273

9.1 Qualitative Analysis: Energy and Phase Diagrams 274
9.2 Elliptical Integrals and Nonlinear Oscillations 280
9.3 Fourier Series Expansions 283
9.3.1 Symmetrical Potential: $V(x) = V(-x)$ 284
9.3.2 Asymmetrical Potential: $V(-x) = -V(x)$ 287
9.4 The Method of Perturbation 288
9.4.1 Bogoliuboff–Kryloff Procedure and Removal of Secular Terms 292
9.5 Ritz Method 295
9.6 Method of Successive Approximation 297
9.7 Multiple Solutions and Jumps 299
9.8 Chaotic Oscillations 301
9.8.1 Some Helpful Tools for an Understanding of Chaos 301
9.8.2 Conditions for Chaos 306
9.8.3 Routes to Chaos 307
9.8.4 Lyapunov Exponentials 308
References 312

Chapter 10 Collisions and Scatterings 313

10.1 Direct Impact of Two Particles 313
10.2 Scattering Cross Sections and Rutherford Scattering 318
10.2.1 Scattering Cross Sections 319
10.2.2 Rutherford's α-Particle Scattering Experiment 320
10.2.3 Cross Section Is Lorentz Invariant 324

10.3 Laboratory and Center-of-Mass Frames of Reference 324
10.4 Nuclear Sizes 328
10.5 Small-Angle Scattering (Optional) 329
References 336

Chapter 11 Motion in Non-Inertial Systems 337

11.1 Accelerated Translational Coordinate System 337
11.2 Dynamics in Rotating Coordinate System 341
11.2.1 Centrifugal Force 345
11.2.2 The Coriolis Force 349
11.2.2.1 Trade Winds and Circulation of Ocean Currents 351
11.2.2.2 Weather Systems 352
11.2.2.3 Hurricanes 354
11.2.2.4 Bathtub Vortex and Earth Rotation 354
11.3 Motion of Particle Near the Surface of the Earth 355
11.4 Foucault Pendulum 361
11.5 Larmor's Theorem 364
11.6 Classical Zeeman Effect 365
11.7 Principle of Equivalence 368
11.7.1 Principle of Equivalence and Gravitational Red Shift 369

Chapter 12 Motion of Rigid Bodies 377

12.1 Independent Coordinates of Rigid Body 378
12.2 Eulerian Angles 379
12.3 Rate of Change of Vector 382
12.4 Rotational Kinetic Energy and Angular Momentum 384
12.5 Inertia Tensor 394
12.5.1 Diagonalization of a Symmetric Tensor 396
12.5.2 Moments and Products of Inertia 397
12.5.3 Parallel-Axis Theorem 398
12.5.4 Moments of Inertia about an Arbitrary Axis 401
12.5.5 Principal Axes of Inertia 403
12.6 Euler's Equations of Motion 407
12.7 Motion of a Torque-Free Symmetrical Top 409
12.8 Motion of Heavy Symmetrical Top with One Point Fixed 414
12.8.1 Precession without Nutation 417
12.8.2 Precession with Nutation 419
12.9 Stability of Rotational Motion 420
References 425

Chapter 13 Theory of Special Relativity 427

13.1 Historical Origin of Special Theory of Relativity 427
13.2 Michelson–Morley Experiment 430
13.3 Postulates of Special Theory of Relativity 433
13.3.1 Time Is Not Absolute 434
13.4 Lorentz Transformations 434
13.4.1 Relativity of Simultaneity, Causality 437
13.4.2 Time Dilation, Relativity of Co-Locality 438

13.4.3 Length Contraction 439
13.4.4 Visual Apparent Shape of Rapidly Moving Object 441
13.4.5 Relativistic Velocity Addition 441
13.5 Doppler Effect 445
13.6 Relativistic Space–Time (Minkowski Space) 446
13.6.1 Four-Velocity and Four-Acceleration 449
13.6.2 Four-Energy and Four-Momentum Vectors 450
13.6.3 Particles of Zero Rest Mass 452
13.7 Equivalence of Mass and Energy 453
13.8 Conservation Laws of Energy and Momentum 459
13.9 Generalization of Newton's Equation of Motion 459
13.9.1 Force Transformation 461
13.10 Relativistic Lagrangian and Hamiltonian Functions 463
13.11 Relativistic Kinematics of Collisions 467
13.12 Collision Threshold Energies 470
References 474

Chapter 14 Newtonian Gravity and Newtonian Cosmology 475

14.1 Newton's Law of Gravity 475
14.2 Gravitational Field and Gravitational Potential 477
14.3 Gravitational Field Equations: Poisson's and Laplace's Equations 479
14.4 Gravitational Field and Potential of Extended Body 480
14.5 Tides 481
14.6 General Theory of Relativity: Relativistic Theory of Gravitation 487
14.6.1 Gravitational Shift of Spectral Lines (Gravitational Red Shift) 488
14.6.2 Bending of Light Beam 489
14.7 Introduction to Cosmology 491
14.8 Brief History of Cosmological Ideas 492
14.8.1 Newton and Infinite Universe 493
14.8.2 Newton's Law of Gravity Predicts Nonstationary Universe 493
14.8.3 An Infinite Steady Universe Is an Empty Universe 495
14.8.4 Olbers' Paradox 496
14.9 Discovery of Expansion of the Universe, Hubble's Law 497
14.10 Big Bang 499
14.10.1 Age of the Universe 499
14.11 Formulating Dynamical Models of the Universe 499
14.12 Cosmological Red Shift and Hubble Constant H 503
14.13 Critical Mass Density and Future of the Universe 504
14.13.1 Density Parameter Ω 505
14.13.2 Deceleration Parameter q_0 505
14.13.3 An Accelerating Universe? 507
14.14 Microwave Background Radiation 507
14.15 Dark Matter 511
Reference 514

Chapter 15 Hamilton–Jacobi Theory of Dynamics 515

15.1 Canonical Transformation and H-J Equation 515
15.2 Action and Angle Variables 522

15.3 Infinitesimal Canonical Transformations and Time Development Operator 527
15.4 H-J Theory and Wave Mechanics........ 530
Reference........ 533

Chapter 16 Introduction to Lagrangian and Hamiltonian Formulations for Continuous Systems and Classical Fields........ 535

16.1 Vibration of Loaded String........ 535
16.2 Vibrating Strings and the Wave Equation 541
16.2.1 Wave Equation........ 541
16.2.2 Separation of Variables 543
16.2.3 Wave Number and Phase Velocity 543
16.2.4 Group Velocity and Wave Packets 544
16.3 Continuous Systems and Classical Fields........ 547
16.3.1 Lagrangian Formulation........ 547
16.3.2 Hamiltonian Formulation........ 550
16.3.3 Conservation Laws 552
16.4 Scalar and Vector of Fields........ 553
16.4.1 Scalar Fields 553
16.4.2 Vector Fields........ 554

Appendix 1: Vector Analysis and Ordinary Differential Equations........ 557

Appendix 2: D'Alembert's Principle and Lagrange's Equations 587

Appendix 3: Derivation of Hamilton's Principle from D'Alembert's Principle........ 595

Appendix 4: Noether's Theorem........ 599

Appendix 5: Conic Sections, Ellipse, Parabola, and Hyperbola 605

Preface

This book presents a reasonably complete account of the theoretical mechanics of particles and systems for physics students at the advanced undergraduate level. It is evolved from a set of lecture notes for a course on the subject, which I have taught at California State University, Stanislaus, for many years. We presume that the student has been exposed to a calculus-based general physics course (from a textbook such as that by Halliday and Resnick) and a course in calculus (including the handling of differentiations of field functions). No prior knowledge of differential equations is required. Differential equations and new mathematical methods are developed in the text as the occasion demands. Vectors are used from the start.

The book has 17 chapters, and with appropriate omission, the essential topics can be covered in a one-semester, four-hour course. We do not make any specific suggestions for a shorter course. We usually vary the topics to suit the ability and mathematical background of the students. We would encourage the more enthusiastic and able students to attempt to master on their own the material not covered in class (for extra credit).

A major departure of this book from the conventional approach is the introduction of the Lagrangian and Hamiltonian formulations of mechanics at an early stage. In the conventional approach to the subject, Lagrangian and Hamiltonian formulations are presented near the end of the course, and students rarely develop a reasonable familiarity with these essential methods.

The choice of topics and their treatment throughout the book are intended to emphasize the modern point of view. Applications to other branches of physics are made wherever possible. Special note is made of concepts that are important to the development of modern physics. Also, the relationship between symmetries and the laws of conservation—a subject directly relevant to the most modern developments of physics—is emphasized.

The student will find that a generous amount of detail has been given in mathematical manipulations, and that occurrences of "it may be shown that" have been kept to a minimum. However, to ensure that the student does not lose sight of the development underway, some of the more lengthy and tedious algebraic manipulations have been omitted when possible.

Each chapter contains a set of homework problems of varying degrees of difficulty. They are intended to supplement or amplify the material in the text and are arranged in the order in which the material is covered in the chapter. No effort has been made to trace the origins of the homework problems and examples in our book.

We have omitted a discussion of the historical development of the subject. This is because of the length of the book, not a lack of interest on the author's part. Also, references to the original literature have been omitted except for recent works to which the student may be expected to have access.

Many individuals have been very helpful in the preparation of this text. I wish to thank my colleagues at California State University, Stanislaus, for many helpful suggestions and discussions.

Tai L. Chow
Monterey Park, CA

Author

Dr. Tai Chow was born and raised in China. He received the Bachelor of Science degree in physics from National Taiwan University, a Master's degree in physics from Case Western Reserve University in Cleveland, and a Ph.D. degree in physics from the University of Rochester in New York. Since 1970, Dr. Chow has been in the Department of Physics at California State University, Stanislaus, and served as the department chairman for 18 years. He has also served as a visiting professor at the University of California at Davis and Berkeley and has worked as a summer faculty fellow at Stanford University and at NASA/Ames Center. Dr. Chow has published more than 40 articles in physics and astrophysics journals and is the author of four textbooks: *Classical Mechanics*, published in 1995 by John Wiley & Sons; *Mathematical Methods for Physicists*, published in 2000 by Cambridge University Press; *Introduction to Electromagnetic Theory*, published in 2005 by Jones & Bartlett Publishers; and *Gravity, Black Holes, and the Very Early Universe*, published in 2008 by Springer.

1 Kinematics
Describing the Motion

1.1 INTRODUCTION

Classical mechanics studies the motion of macroscopic bodies. It is the backbone of modern physics because it furnishes the basic conceptual framework of nearly all of physics. Its basic concepts, such as energy, momentum, and angular momentum, and the conservation laws associated with these basic concepts are of the greatest importance in all branches of physics.

Newton and Galileo first laid the foundations of classical mechanics in the 17th century. The essential physics of their mechanics, known as Newtonian mechanics, is contained in Newton's three laws of motion. Classical mechanics has since been reformulated in a few different forms: the Lagrange, the Hamilton, the Poisson brackets, and the Hamilton–Jacobi formalisms. These alternative formulations are equivalent to Newtonian mechanics. These different methods of description emphasize different aspects of the phenomena and, hence, serve different purposes. It is the aim of this book to present these various formulations and their applications to topics of present-day interest.

In the microscopic domain, classical mechanics has been superseded by quantum mechanics; for phenomena involving speeds approaching that of light, it has been modified by the special theory of relativity. Classical mechanics, therefore, lacks the glamour of being in the forefront of modern physics. But Hamiltonian and Lagrangian dynamics are the starting points for quantum mechanics, statistical mechanics, and quantum field theory. Also, the new frontiers of nonlinear behavior—chaos and stochastic motion—are analyzed using classical mechanics. A good grasp of classical mechanics is therefore a prerequisite for the study of the new theories of physics.

This chapter will survey a number of fundamental concepts, such as velocity and acceleration, which are basic to our presentation in the succeeding chapters. The branch of mechanics that describes motion that does not require knowledge of its cause is called "kinematics," and the part of mechanics that concerns the physical mechanisms that cause the motion to take place is termed "dynamics." We shall mainly be concerned with particle dynamics; the motion of rigid bodies will be addressed in Chapter 12. A body has both mass and extent, and a particle is a body whose dimensions may be neglected in describing its motion. Whether we can treat the motion of a given body as that of a particle depends not only on its size but also on the conditions of the physical problem concerned. The Earth may be regarded as a particle in the context of its motion around the sun but not in a discussion of its daily rotation on its axis.

1.2 SPACE, TIME, AND COORDINATE SYSTEMS

Motion involves the change in the body's position in space as time progresses. So we shall develop classical mechanics in terms of certain notions of space and time. We shall not go into a deep philosophical discussion on the basic concept of space and time that is a very difficult one to comprehend or described at our level. We are concerned with the motion of bodies, and so it is only necessary for a concept of space to provide a way in which the position of such bodies can be described. We may regard space, in classical mechanics, simply as the set of possible positions that the component points of a body may occupy. Because position of a point can only be defined relative to a set of other points, position is, therefore, a relative concept. This leads to the philosophical question as to whether space itself is a relative concept or whether it is in some sense absolute. Fortunately, however, it is not necessary here to discuss and answer this question. All that is required is to indicate how this problem is dealt with in practice.

Another question concerns whether space is continuous or not. If space were continuous, all positions would be possible. On the other hand, if only certain positions are allowed, then space is not continuous (or space is quantized). Space appears to be continuous down to a scale of at least 10^{-13} cm as we cannot determine any position with accuracy greater than this. In atomic and subatomic domains, classical dynamics breaks down, and quantum mechanics rules. We assume space is continuous for the theory of classical dynamics. We further assume that the geometry of space is Euclidean.

Time is a concept more difficult than space to describe. But we have an intuitive notion about time that is good enough in classical mechanics. Time is the way in which we order events as they happen. A body may be in the same place at different times but cannot be in different places at the same time. In classical mechanics, time is regarded as being totally independent of motion; we assume that there is a universal time scale in the sense that two observers who have synchronized their clocks will always agree about the time of any event. It should be noted that the assumptions of absolute time and of the geometry of space have been modified by the theory of relativity.

We need a reference system in space and time that will enable us to specify the position of a particle or a body as a function of time. We can use a body or a group of bodies as a frame of reference, relative to which the motion of the body can be measured. In physics, we use a coordinate system as a reference frame. For general discussion, we might take a Cartesian system of coordinates and a time scale t. The Cartesian coordinate system consists of a set of three mutually perpendicular axes (x_1, x_2, x_3) passing through a common origin O (Figure 1.1). The perpendicular distances from P to three mutually perpendicular planes intersecting at O may specify the position of a point P. The position of P may also be specified by the straight-line distance r from P to O and the direction of OP with respect to the coordinate axes as measured by the three angles α, β, and γ that OP makes with the x, y, and z axes, respectively. The four quantities r, α, β, and γ are not all independent because they satisfy the identical relation $\cos^2\alpha + \cos^2\beta + \cos^2\gamma = 1$. This follows at once from $x = r\cos\alpha$, $y = r\cos\beta$, and $z = r\cos\gamma$.

The line OP from O to P is a vector quantity that has both magnitude and direction, and it is called a position vector or a radius vector. Vector notation that does not refer explicitly to a particular coordinate system will be used freely.

In problems with particular symmetries, it is convenient to use nonrectangular coordinates. In the case of axial or spherical symmetry, we may use cylindrical or spherical polar coordinates. In cylindrical coordinates (Figure 1.2a), the position of P is given on the surface of a right circular

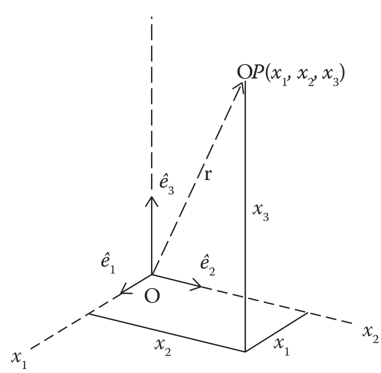
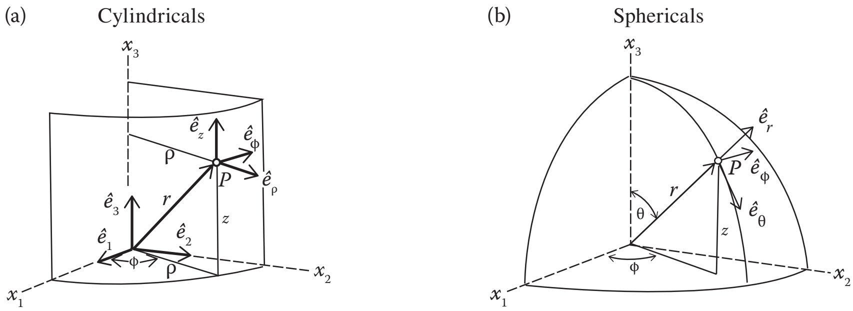


FIGURE 1.1 Position vector in Cartesian coordinates.



$x = \rho\cos\phi$

$y = \rho\sin\phi$

$z = z$

$x = r\ \sin\theta\ \cos\phi$

$y = r\ \sin\theta\ \sin\phi$

$z = r\ \cos\theta$

FIGURE 1.2 (a) Cylindrical coordinates and (b) spherical coordinates.

cylinder with the z (or x_3) axis as its axis and radius ρ. The coordinates are (ρ, ϕ, z). In the spherical coordinate system (Figure 1.2b), P is given on the surface of a sphere with the center at O and radius r. The longitude is given by ϕ and the colatitude by θ.

A reference frame can be chosen arbitrarily in an infinite number of ways, and the description of motion in different frames will, in general, be different. There are frames of reference relative to which a body that does not interact with other bodies moves with a constant speed in a straight line. Frames of reference satisfying this condition are called "inertial frames of reference." Any frame moving at constant velocity with respect to an inertial frame is also an inertial frame. We shall see in the following chapter that the laws of theoretical mechanics have the same form in every inertial reference frame but different forms in different non-inertial reference frames. Therefore, it is simple to study physical phenomena in inertial frames, and we shall do so except where otherwise stated.

1.3 CHANGE OF COORDINATE SYSTEM (TRANSFORMATION OF COMPONENTS OF A VECTOR)

Vector equations are independent of the coordinate system. But the components of a vector quantity are different in different coordinate systems. We now review how to represent a vector quantity in different coordinate systems. As the Cartesian coordinate system is the basic type of coordinate system, we shall limit our discussion to it. Now consider the vector $\vec{A}$ expressed in terms of the unit coordinate vectors $\left(\hat{e}_1, \hat{e}_2, \hat{e}_3\right)$:

$$\vec{A} = A_1\hat{e}_1 + A_2\hat{e}_2 + A\hat{e}_3 = \sum_{i=1}^{3} A_i\hat{e}_i. \tag{1.1}$$

Relative to a new system $\left(\hat{e}'_1, \hat{e}'_2, \hat{e}'_3\right)$ that is obtained from the old system $\left(\hat{e}_1, \hat{e}_2, \hat{e}_3\right)$ by a rotation about the origin, vector $\vec{A}$ is expressed as

$$\vec{A} = A'_1\hat{e}'_1 + A'_2\hat{e}'_2 + A'\hat{e}'_3 = \sum_{i=1}^{3} A'_i\hat{e}'_i. \tag{1.2}$$

Note that the dot product $\vec{A}\hat{e}'_1$ is equal to A'_1, the projection of $\vec{A}$ on the direction of $\hat{e}'_1$; $\vec{A}\hat{e}'_2$ is equal to A'_2, and $\vec{A}\hat{e}'_3$ is equal to A'_3. Thus we may write

$$\begin{aligned} A_1' &= (\hat{e}_1 \cdot \hat{e}_1')A_1 + (\hat{e}_2 \cdot \hat{e}_1')A_2 + (\hat{e}_3 \cdot \hat{e}_1')A_3 \\ A_2 &= (\hat{e}_1 \cdot \hat{e}_2)A_1 + (\hat{e}_2 \cdot \hat{e}_2)A_2 + (\hat{e}_3 \cdot \hat{e}_2)A_3. \\ A_3' &= (\hat{e}_1 \cdot \hat{e}_3')A_1 + (\hat{e}_2 \cdot \hat{e}_3')A_2 + (\hat{e}_3 \cdot \hat{e}_3')A_3 \end{aligned} \tag{1.3}$$

The dot products $(\hat{e}_i \cdot \hat{e}_j')$ are the direction cosines of the axes of the new coordinate system relative to the old system: $\hat{e}_i' \cdot \hat{e}_j = \cos(x_i', x_j)$; they are often called the coefficients of transformation. In matrix notation, we can write the above system of equations as

$$\begin{pmatrix} A_1' \\ A_2' \\ A_3' \end{pmatrix} = \begin{pmatrix} \hat{e}_1 \cdot \hat{e}_1' & \hat{e}_2 \cdot \hat{e}_1' & \hat{e}_3 \cdot \hat{e}_1' \\ \hat{e}_1 \cdot \hat{e}_2' & \hat{e}_2 \cdot \hat{e}_2' & \hat{e}_3 \cdot \hat{e}_2' \\ \hat{e}_1 \cdot \hat{e}_3' & \hat{e}_2 \cdot \hat{e}_3' & \hat{e}_3 \cdot \hat{e}_3' \end{pmatrix} \begin{pmatrix} A_1 \\ A_2 \\ A_3 \end{pmatrix}. \tag{1.4}$$

The 3 × 3 matrix in the above equation is called the rotation (or transformation) matrix; it is an orthogonal matrix. Successive transformations can be handled easily by means of matrix multiplication. Let us digress for a quick review of some basic matrix algebra.

A matrix is an ordered array of scalars that obeys prescribed rules of addition and multiplication. Its row number followed by its column number specifies a particular matrix element. Thus, a_{ij} is the matrix element in the ith row and jth column. We may also represent matrix $\tilde{A}$ by $[a_{ij}]$ or the entire array

$$\tilde{A} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}. \tag{1.5}$$

$\tilde{A}$ is an $m \times n$ matrix. A vector is represented in matrix form by writing its components as either a row or column array, such as

$$\tilde{B} = (b_{11}\ b_{12}\ b_{13}) \quad \text{or} \quad \tilde{C} = \begin{pmatrix} c_{11} \\ c_{21} \\ c_{31} \end{pmatrix} \tag{1.6}$$

where $b_{11} = b_x$, $b_{12} = b_y$, $b_{13} = b_z$, and $c_{11} = c_x$, $c_{21} = c_y$, $c_{31} = c_z$.

The multiplication of two matrices $\tilde{A}$ and $\tilde{B}$ is defined only when the number of columns of $\tilde{A}$ is equal to the number of rows of $\tilde{B}$ and is performed in the same way as the multiplication of two determinants:

$$\tilde{C} = \tilde{A}\tilde{B}, \quad c_{ij} = \sum_k a_{ik} b_{kl}. \tag{1.7}$$

We illustrate the multiplication rule for the case of the 3 × 3 matrix $\tilde{A}$ multiplied by the 3 × 3 matrix $\tilde{B}$:

$$a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} = c_{12}$$

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}\begin{pmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \\ b_{31} & b_{32} & b_{33} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & c_{13} \\ c_{21} & c_{22} & c_{23} \\ c_{31} & c_{32} & c_{33} \end{pmatrix}.$$

If we denote the direction cosines $\hat{e}'_i \cdot \hat{e}_j$ by λ_{ij}, then Equation 1.3 can be written as

$$A'_i = \sum_{j=1}^{3} \hat{e}'_i \cdot \hat{e}_j \; A_j = \sum_{j=1}^{3} \lambda_{ij} A_j. \tag{1.8}$$

Not all of the nine quantities λ_{ij} are all independent; there are six relationships that exist among the λ_{ij}, so only three of them are independent. These six relationships are given by

$$\sum_{i=1}^{3} \lambda_{ij}\lambda_{ik} = \delta_{jk}, \quad (j,k = 1,2,3). \tag{1.9}$$

Any linear transformation, such as Equation 1.8, that has the properties required by Equation 1.9 is called an orthogonal transformation, and Equation 1.9 is known as the orthogonal condition.

Equation 1.9 can be found by using the fact that the magnitude of the vector must be the same in both systems. The invariance of the magnitude is expressed as

$$\sum_{i=1}^{3} (A'_i)^2 = \sum_{i=1}^{3} (A_i)^2. \tag{1.10}$$

The left side can be rewritten as

$$\sum_{i=1}^{3} (A'_i)^2 = \sum_{i=1}^{3}\left(\sum_{j=1}^{3} \lambda_{ij} A_j\right)\left(\sum_{k=1}^{3} \lambda_{ik} A_k\right) = \sum_{j=1}^{3}\sum_{k=1}^{3}\left(\sum_{i=1}^{3} \lambda_{ij}\lambda_{ik}\right) A_j A_k,$$

which will be reduced to the right-hand side of Equation 2.10 if, and only if, Equation 1.9 is true.

Example 1.1

Find the transformation matrix for coordinate change by a rotation through an angle ϕ about the x_3 axis (Figure 1.3).

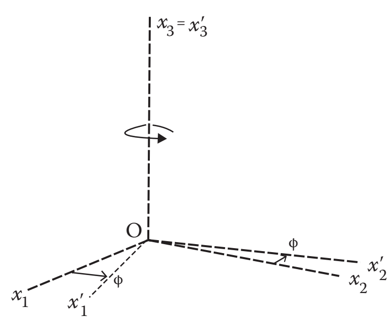


FIGURE 1.3 Coordinate changes by a rotation.

Solution:

We have

$$x_i' = \sum_{j=1}^{3} \lambda_{ij} x_j$$

with, for such a rotation,

$$\lambda_{11} = \hat{e}_1 \cdot \hat{e}_1' = \cos\phi, \quad \lambda_{22} = \hat{e}_2 \cdot \hat{e}_2' = \cos\phi, \quad \lambda_{33} = \hat{e}_3 \cdot \hat{e}_3' = 1,$$
$$\lambda_{12} = \hat{e}_2 \cdot \hat{e}_1' = \sin\phi, \quad \lambda_{21} = \hat{e}_1 \cdot \hat{e}_2' = -\sin\phi.$$

All other components of λ_{ij} are zero. The transformation equations take the form

$$x_1' = x_1 \cos\phi + x_2 \sin\phi, \quad x_2' = -x_1 \sin\phi + x_2 \cos\phi, \quad x_3' = x_3$$

and the transformation matrix is

$$\tilde{\lambda} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The orthogonal conditions are obviously satisfied:

$$\lambda_{11}\lambda_{11} + \lambda_{21}\lambda_{21} = \cos^2\phi + \sin^2\phi = 1$$
$$\lambda_{22}\lambda_{22} + \lambda_{12}\lambda_{12} = \cos^2\phi + \sin^2\phi = 1$$
$$\lambda_{11}\lambda_{12} + \lambda_{21}\lambda_{22} = \cos\phi\sin\phi - \sin\phi\cos\phi = 0.$$

It was mentioned earlier that one advantage of using the matrix notation is that successive transformations can be easily handled by matrix multiplication. We now demonstrate this by showing that successive application of two orthogonal transformations is equivalent to a third orthogonal transformation. We denote the first transformation from the unprimed coordinate set r to the primed coordinate set r' by $\tilde{\alpha} = [\alpha_{ij}]$:

$$x_k' = \sum_{j=1}^{3} \alpha_{kj} x_j \quad (k = 1, 2, 3) \tag{1.11a}$$

and the succeeding transformation from r' to r'' by $\tilde{\beta} = [\beta_{ij}]$:

$$x_i'' = \sum_{k=1}^{3} \beta_{ik} x_k' \quad (i = 1, 2, 3). \tag{1.11b}$$

Inserting Equation 1.11a into Equation 1.11b, we obtain

$$x_i'' = \sum_{k,j=1}^{3} \beta_{ik}\alpha_{kj} x_j = \sum_{j=1}^{3} \gamma_{ij} x_j \quad (i = 1, 2, 3) \tag{1.12}$$

where

$$\gamma_{ij} = \sum_{k=1}^{3} \beta_{ik}\alpha_{kj} \quad (i, j = 1, 2, 3). \tag{1.12a}$$

It is easy to show that γ_{ij} satisfies the orthogonality condition

$$\sum_{i=1}^{3} \gamma_{ij}\gamma_{in} = \sum_{i,k,m=1}^{3} \beta_{ik}\alpha_{kj}\beta_{im}\alpha_{mn} = \sum_{k,m=1}^{3}\left(\sum_{i=1}^{3}\beta_{ik}\beta_{im}\right)\alpha_{kj}\alpha_{mn} = \sum_{k,m=1}^{3}\delta_{km}\alpha_{kj}\alpha_{mn}$$
$$= \sum_{m=1}^{3}\alpha_{mj}\alpha_{mn} = \delta_{jn} \quad (j, n = 1, 2, 3).$$

Example 1.2

Let us revisit Example 1.1. Suppose now a succeeding transformation in which the double-primed system is generated by a rotation through an angle θ about the x_1' axis:

$$x_k'' = \sum_{i=1}^{3} \beta_{ki} x_i' \quad (k = 1, 2, 3)$$

where

$$\tilde{\beta} = [\beta_{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}.$$

The transformation matrix $\tilde{\gamma} = [\gamma_{ij}]$ for the succeeding transformation of rotation ϕ followed by rotation θ is given by

$$\tilde{\gamma} = \tilde{\beta}\tilde{\alpha}$$

or

$$\begin{pmatrix} \gamma_{11} & \gamma_{12} & \gamma_{13} \\ \gamma_{21} & \gamma_{22} & \gamma_{23} \\ \gamma_{31} & \gamma_{32} & \gamma_{33} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}\begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
$$= \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\cos\theta\sin\phi & \cos\theta\cos\phi & \sin\theta \\ \sin\theta\sin\phi & -\sin\theta\cos\phi & \cos\theta \end{pmatrix}$$

from which we obtain

$$\gamma_{11} = \cos\phi, \qquad \gamma_{12} = \sin\phi, \qquad \gamma_{13} = 0$$
$$\gamma_{21} = -\cos\theta\sin\phi, \qquad \gamma_{22} = \cos\theta\cos\phi, \qquad \gamma_{23} = \sin\theta$$
$$\gamma_{31} = \sin\theta\sin\phi, \qquad \gamma_{32} = -\sin\theta\cos\phi, \qquad \gamma_{33} = \cos\theta.$$

The orthogonal condition is obviously satisfied. We leave this for the reader to verify.

Before turning our discussion to kinematics, we would like to make two remarks:

1. In the above discussion, the vector is considered fixed, and we rotate the coordinate axes. Thus, the transformation matrix can be thought of as an operator that acts on the unprimed system and transforms it into the primed system. This is often called the passive view of rotation. We could equally keep the coordinate axes fixed and rotate the vector by an equal angle but in the opposite direction. Then, the transformation matrix would be considered as an operator acting on the vector, say $\vec{A}$, and changing it into $\vec{A}'$ of equal magnitude. This latter procedure is known as the active view of rotation.
2. Scalar and vector quantities can be defined in terms of transformation properties. A scalar is a quantity that is unaffected by an orthogonal transformation of coordinates; a vector is a set of three quantities, say, (A_1, A_2, A_3), whose transformation properties are the same as those of the coordinates:

$$A_j = \sum_{k=1}^{3} \lambda_{jk} A_k, \quad j = 1,2,3$$

where the coefficients λ_{jk} are exactly those appearing in Equation 1.8.

1.4 DISPLACEMENT VECTOR

When a particle changes its position in space, it is said to undergo a displacement. Suppose the particle is at time t at P in which its position is given by the position vector $\vec{r}$; at the later moment $t + \Delta t$ is at P', its position being given by $\vec{r}'$. The displacement vector $\Delta\vec{r}$ between P and P' is the straight-line distance between these two points, and its direction is from P to P':

$$\Delta\vec{r} = \vec{r}' - \vec{r}.$$

As depicted in Figure 1.4, a butterfly and a bee fly from point P to point Q. The bee takes a direct path from P to Q, and the butterfly takes a complicated route, but they have the same displacement vector. Although the butterfly wanders over a large region, its net displacement vector is the vector sum of many small elementary displacements.

1.5 SPEED AND VELOCITY

Speed v is defined as the rate of change of distance with time. It is the distance an object moves during a certain time interval divided by the interval of time:

$$v = \Delta d/\Delta t. \tag{1.13}$$

More precisely, the speed given by Equation 1.13 is only the average speed of the motion. The speed of a moving object is not, by itself, a complete description of its state of motion. The direction of

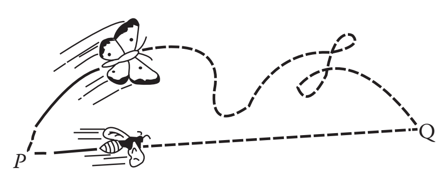


FIGURE 1.4 Displacement vector.

motion is also important. Without knowing the direction of motion, we cannot determine the path of a moving object. To specify motion completely, then, we must include the direction of motion as well as the speed. Such a total description is called velocity, which is a vector quantity, and speed is its magnitude.

The velocity $\vec{v}$ of a moving particle is defined as the rate of change of displacement with time. The ratio $\Delta\vec{r}/\Delta t$ is the average velocity of the particle during the displacement. This ratio has a unique limit as $\Delta t \to 0$, provided $\vec{r}$ is a continuous function of t. This limit $\mathrm{d}\vec{r}/\mathrm{d}t$ is called the instantaneous velocity $\vec{v}$ of the particle at the instant in question and is a vector tangential to the particle path at P (Figure 1.5):

$$\vec{v}(t) = \frac{\mathrm{d}\vec{r}(t)}{\mathrm{d}t} = \dot{\vec{r}} = \lim_{\Delta t\to 0}\frac{\Delta\vec{r}}{\Delta t} = \lim_{\Delta t\to 0}\frac{\vec{r}(t+\Delta t) - r(t)}{\Delta t}. \tag{1.14}$$

We can write

$$\vec{v} = v\hat{e}_t$$

where $\hat{e}_t = \vec{v}/v$, a unit vector in the direction of $\vec{v}$. For many purposes, it is convenient to represent $\vec{r}$ in component form:

$$\vec{r}(t) = x(t)\hat{e}_1 + y(t)\hat{e}_2 + z(t)\hat{e}_3.$$

Then, Equation 1.14 becomes

$$\vec{v} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}t} = \hat{e}_1\frac{\mathrm{d}x}{\mathrm{d}t} + \hat{e}_2\frac{\mathrm{d}y}{\mathrm{d}t} + \hat{e}_3\frac{\mathrm{d}z}{\mathrm{d}t}. \tag{1.15}$$

The magnitude of the velocity (the speed) is given by

$$v = |\vec{v}| = \sqrt{\vec{v}\cdot\vec{v}} = \sqrt{(\mathrm{d}x/\mathrm{d}t)^2 + (\mathrm{d}y/\mathrm{d}t)^2 + (\mathrm{d}z/\mathrm{d}t)^2} = \mathrm{d}s/\mathrm{d}t, \tag{1.16}$$

where s is the arc length along the particle path measured from some initial point to P.

If $\vec{v}$ and $\vec{v}'$ are velocities of a moving particle in two different frames of reference O and O′, respectively, and frame O moves with a velocity $\vec{V}$ relative to the frame O′, then

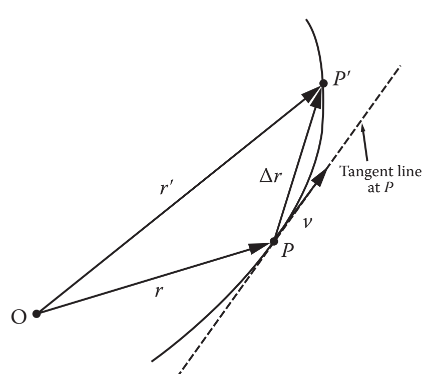


FIGURE 1.5 Instantaneous velocity is a vector tangent to the path at P.

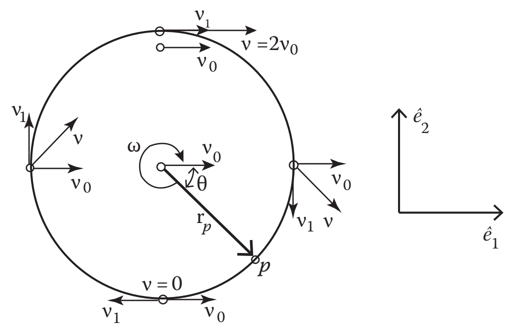


FIGURE 1.6 Velocity vectors for various points on a turning wheel.

$$\vec{v}' = \vec{v} + \vec{V}.$$

This expression is not valid at very high velocities compared to that of light. In getting the expression, we have assumed tacitly that the passage of time is the same in both frames. This assumption has been radically modified by the special theory of relativity.

Example 1.3

A wheel of radius b rolls without sliding along the ground with a forward speed of v_o. Calculate the velocity, relative to a person standing on the ground, of a point on the rim.

Solution:

As shown in Figure 1.6, p is the arbitrary point on the rim, and its radius vector from the center of the wheel is $\boldsymbol{r}_p$, which can be expressed in terms of t and θ:

$$\vec{r}_p = b\cos\theta\hat{e}_1 - b\sin\theta\hat{e}_2$$

where $\theta = \omega t$, and $\omega = v_o/b$. The wheel turns clockwise about its center.

The velocity of p relative to the center of the wheel is given by

$$\vec{v}_1 = \mathrm{d}\vec{r}_p/\mathrm{d}t = -b\omega\sin\theta\hat{e}_1 - b\omega\cos\theta\hat{e}_2$$

as the velocity of the center of the wheel relative to the ground is $v_o\hat{e}_1$; hence, the velocity of p relative to the ground is

$$\vec{v} = v_o\hat{e}_1 + \vec{v}_1 = v_o(1-\sin\theta)\hat{e}_1 - v_o\cos\theta\hat{e}_2.$$

1.6 ACCELERATION

When the velocity changes with time, we introduce the concept of acceleration $\vec{a}$, which is defined as the time derivative of the velocity:

$$\vec{a} = \lim_{\Delta t\to 0}\frac{\Delta\vec{v}}{\Delta t} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2} = \ddot{\vec{r}}. \quad (1.17)$$

In terms of $\vec{r}(t) = x(t)\hat{e}_1 + y(t)\hat{e}_2 + z(t)\hat{e}_3$, the acceleration $\vec{a}$ is given by

$$\vec{a} = \frac{d^2\vec{r}}{dt^2} = \frac{d^2x}{dt^2}\hat{e}_1 + \frac{d^2y}{dt^2}\hat{e}_2 + \frac{d^2z}{dt^2}\hat{e}_3,$$

and its magnitude is

$$a = |\vec{a}| = \sqrt{\vec{a}\cdot\vec{a}} = \sqrt{(d^2x/dt^2)^2 + (d^2y/dt^2)^2 + (d^2z/dt^2)^2}\,.$$

Whenever the words "velocity" and "acceleration" are used hereafter, they shall be taken to mean the instantaneous values unless otherwise specified.

It should be noted that $|d\vec{r}/dt|$ is not at all the same thing as the derivative dr/dt of the magnitude of $\boldsymbol{r}$ with respect to t. Consider, for example, a particle moving with constant speed in a circle the center of which is at the origin. As r is constant, dr/dt is zero, but $|d\vec{r}/dt|$ is equal to the scalar magnitude of the velocity.

1.6.1 Tangential and Normal Acceleration

When a particle moves along a curve, its velocity may vary in both magnitude and direction, so its acceleration vector has two components: one parallel to the velocity and one perpendicular to it. Occasionally, it is desirable to resolve the acceleration into these two components. Let $\hat{e}_t$ (Figure 1.7) be a unit vector along the tangent taken in the sense in which the motion is described, and $\hat{e}_n$ is a unit vector directed toward the center of curvature O of the path. Then

$$\vec{v} = v\hat{e}_t$$

and

$$\vec{a} = \frac{d\vec{v}}{dt} = \hat{e}_t\frac{dv}{dt} + \frac{d\hat{e}_t}{dt}v.$$

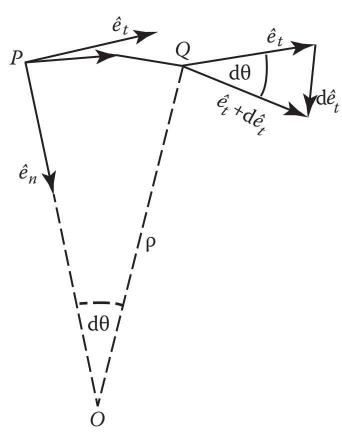


FIGURE 1.7 Tangential and normal acceleration.

Clearly, the first term on the right is the tangential component. To see the nature of the second term, we note that $\mathrm{d}\hat{e}_t$ has the direction of $\hat{e}_n$ and the magnitude dθ. So

$$\vec{a} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \hat{e}_t \frac{\mathrm{d}v}{\mathrm{d}t} + \hat{e}_n v \frac{\mathrm{d}\theta}{\mathrm{d}t}. \tag{1.18}$$

Now as the distance *PQ* that the particle has traveled in the time Δt is equal to $v\mathrm{d}t$, it follows that $\mathrm{d}\theta = PQ/\rho = (v/\rho)\mathrm{d}t$, where ρ is the radius of curvature of the path. And Equation 1.18 becomes

$$\vec{a} = \frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \hat{e}_t \frac{\mathrm{d}v}{\mathrm{d}t} + \hat{e}_n \frac{v^2}{\rho} = a_t\hat{e}_t + a_n\hat{e}_n \tag{1.19}$$

where $a_t = \mathrm{d}v/\mathrm{d}t = \dot{v}$, and $a_n = v^2/\rho$. It is sometimes very convenient to use a dot placed over a symbol to denote a derivative with respect to time *t*: one dot for a first derivative, two dots for a second derivative, and so forth. The normal component a_n is always directed toward the center of curvature on the concave side of the curve of motion; it is therefore called the "centripetal acceleration." It may be written in three forms:

$$\frac{v^2}{\rho} = v\frac{\mathrm{d}\theta}{\mathrm{d}t} = \rho\left(\frac{\mathrm{d}\theta}{\mathrm{d}t}\right)^2.$$

Example 1.4

A particle moves along a space curve *C*, whose position vector is given by

$$\vec{r}(t) = 3\cos 2t\hat{e}_1 + 3\sin 2t\hat{e}_2 + (8t-4)\hat{e}_3.$$

(a) Find a unit tangent vector $\hat{e}_t$ to the curve and verify that $\boldsymbol{v} = v\hat{e}_t$.
(b) Find the curvature, radius of curvature, and unit normal vector $\hat{e}_n$ to any point of the space curve *C*.

Solution:

(a)

$$\mathrm{d}r/\mathrm{d}t = -6\sin 2t\hat{e}_1 + 6\cos 2t\hat{e}_2 + 8\hat{e}_3,$$

and

$$\left|\mathrm{d}\vec{r}/\mathrm{d}t\right| = \mathrm{d}s/\mathrm{d}t = \sqrt{(-6\sin 2t)^2 + (6\cos 2t)^2 + 8^2 = 10};$$

hence

$$\hat{e}_t = \frac{\mathrm{d}\vec{r}/\mathrm{d}t}{\left|\mathrm{d}\vec{r}/\mathrm{d}t\right|} = \frac{\mathrm{d}\vec{r}/\mathrm{d}t}{\mathrm{d}s/\mathrm{d}t} = \frac{\mathrm{d}\vec{r}}{\mathrm{d}s} = -\frac{3}{5}\sin 2t\hat{e}_1 + \frac{3}{5}\cos 2t\hat{e}_2 + \frac{4}{5}\hat{e}_3$$

$$\begin{aligned}\mathrm{d}\vec{r}/\mathrm{d}t &= -6\sin 2t\hat{e}_1 + 6\cos 2t\hat{e}_2 + 8\hat{e}_3 \\ &= (10)\left(-\frac{3}{5}\sin 2t\hat{e}_1 + \frac{3}{5}\cos 2t\hat{e}_2 + \frac{4}{5}\hat{e}_3\right) \\ &= v\hat{e}_t\end{aligned}$$

(b)

$$\frac{d\hat{e}_t}{ds} = \frac{d\hat{e}_t/dt}{ds/dt} = \frac{(-6/5)\cos 2t\hat{e}_1 - (6/5)\sin 2t\hat{e}_2}{10}$$
$$= -\frac{3}{25}\cos 2t\hat{e}_1 - \frac{3}{25}\sin 2t\hat{e}_2.$$

The curvature κ is

$$\kappa = |d\hat{e}_t/ds| = 3/25,$$

and the radius of curvature R is

$$R = 1/\kappa = 25/3.$$

$$\hat{e}_n = \frac{d\hat{e}_t/ds}{|d\hat{e}_t/ds|} = R\frac{d\hat{e}_t}{ds} = -\cos 2t\hat{e}_1 - \sin 2t\hat{e}_2.$$

Example 1.5

A car moves uniformly on a circular horizontal track of radius b and completes a circle in time T. Find (a) its position vector $\boldsymbol{r}$, (b) its velocity $\boldsymbol{v}$, and (c) its acceleration $\boldsymbol{a}$ at time t (Figure 1.8).

Solution

(a) The position vector of P is

$$\vec{r}(t) = b(\cos\theta\hat{e}_1 + \sin\theta\hat{e}_2).$$

Now

$$\frac{\theta}{2\pi} = \frac{t}{T}, \quad \text{or} \quad \theta = \frac{2\pi t}{T}$$

and so $\boldsymbol{r}(t)$ can be rewritten as

$$\vec{r}(t) = b\left(\cos\frac{2\pi t}{T}\hat{e}_1 + \sin\frac{2\pi t}{T}\hat{e}_2\right),$$

where we have chosen the orientation of the x-axis so that, initially (t = 0), the car is at (b, 0, 0).

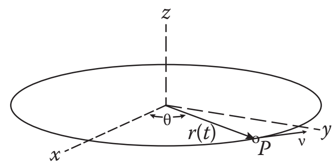


FIGURE 1.8 A car moves on a circular horizontal track.

(b) The velocity vector $\boldsymbol{v}(t)$ is given by

$$\vec{v}(t) = \frac{d\vec{r}(t)}{dt} = \frac{2\pi b}{T}\left(\sin\frac{2\pi t}{T}\hat{e}_1 - \cos\frac{2\pi t}{T}\hat{e}_2\right),$$

and its magnitude is

$$v = \sqrt{\vec{v}\cdot\vec{v}} = \frac{2\pi b}{T}\left[\left(\sin\frac{2\pi t}{T}\hat{e}_1 - \cos\frac{2\pi t}{T}\hat{e}_2\right)\cdot\left(\sin\frac{2\pi t}{T}\hat{e}_1 - \cos\frac{2\pi t}{T}\hat{e}_2\right)\right]^{1/2}.$$

Using the Kronecker delta symbol $\hat{e}_i \cdot \hat{e}_j = \delta_{ij}$, the last expression can be simplified to

$$v = 2\pi b/T,$$

a constant.

(c) The acceleration vector $\boldsymbol{a}(t)$ is given by

$$\vec{a}(t) = \frac{d\vec{v}(t)}{dt} = -\left(\frac{2\pi b}{T}\right)^2\left(\cos\frac{2\pi b}{T}\hat{e}_1 + \sin\frac{2\pi b}{T}\hat{e}_2\right)$$

and because its sign is opposite that of the position vector, it points toward the center of the circle, and so it is perpendicular to the velocity vector. This latter point can be verified directly by computing $\boldsymbol{a}(t)\cdot\boldsymbol{v}(t)$. The tangential acceleration vanishes because v is constant.

1.7 VELOCITY AND ACCELERATION IN POLAR COORDINATES

Calculation of the velocity and acceleration in Cartesian coordinates is a simple matter because the unit vectors $\hat{e}_i (i = 1, 2, 3)$ are constant in time. In polar coordinates, the unit vectors move with the particle. This makes calculation of velocity and acceleration more tedious.

1.7.1 Plane Polar Coordinates (r, θ)

It is often convenient to study the motion of a particle in a plane in terms of plane polar coordinates as shown in Figure 1.9; $\hat{e}_r$ and $\hat{e}_\theta$ are unit vectors in the directions of increasing r and θ, respectively. As $\hat{e}_r$ and $\hat{e}_\theta$ move with the particle, their directions change in space.

The position vector $\boldsymbol{r}$ of the particle can be written as the product of the radial distance r and the unit vector $\hat{e}_r$:

$$\boldsymbol{r} = r\hat{e}_r \tag{1.20}$$

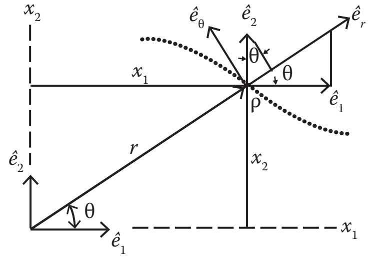


FIGURE 1.9 Plane polar coordinates.

bearing in mind that both r and $\hat{e}_r$ vary and are, therefore, both functions of time. Differentiating Equation 1.20 with respect to time, we get the velocity of the particle:

$$\vec{v} = \frac{d\vec{r}}{dt} = \dot{r}\hat{e}_r + r\frac{d\hat{e}_r}{dt}. \tag{1.21}$$

A study of Figure 1.9 shows that we can express $\hat{e}_r$ and $\hat{e}_\theta$ in terms of rectangular unit vectors $\hat{e}_1$ and $\hat{e}_2$ as follows:

$$\hat{e}_r = \hat{e}_1\cos\theta + \hat{e}_2\sin\theta, \quad \hat{e}_\theta = -\hat{e}_1\sin\theta + \hat{e}_2\cos\theta$$

from which we find

$$\begin{aligned}\dot{\hat{e}}_r &= -\hat{e}_1\dot{\theta}\sin\theta + \hat{e}_2\dot{\theta}\cos\theta \\ &= \dot{\theta}\,\hat{e}_\theta.\end{aligned}$$

With this, Equation 1.21 becomes

$$\vec{v} = \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta. \tag{1.22}$$

In a manner similar to that for obtaining $\dot{\hat{e}}_r$, we find

$$\dot{\hat{e}}_\theta = -\dot{\theta}\hat{e}_r$$

and after some straightforward manipulations, we can finally write the equation for the acceleration in plane polar coordinates as

$$\vec{a} = a_r\hat{e}_r + a_\theta\hat{e}_\theta = (\ddot{r} - r\dot{\theta}^2)\hat{e}_r + (r\ddot{\theta} + 2\dot{r}\dot{\theta})\hat{e}_\theta. \tag{1.23}$$

1.7.2 Cylindrical Coordinates (ρ, ϕ, z)

The cylindrical coordinates are defined by their relationships to the Cartesian coordinates:

$$\left.\begin{aligned} x_1 &= \rho\cos\theta \\ x_2 &= \rho\sin\theta \\ x_3 &= z \end{aligned}\right\} \quad \text{or} \quad \left\{\begin{aligned} \rho^2 &= x_2^2 + x_1^2 \\ \phi &= \tan^{-1}(x_2/x_1). \\ z &= x_3 \end{aligned}\right.$$

A study of Figure 1.2a shows that ρ is the projection of the radius vector ***r*** onto the x_1x_2-plane, and ϕ is the angle that meets with the x_1-axis. The unit vectors $\hat{e}_\rho$ and $\hat{e}_z$ are in the direction of increasing ρ and z, respectively. The unit vector $\hat{e}_\phi$ is parallel to the x_1x_2-plane and perpendicular to $\hat{e}_\rho$ in the direction of increasing ϕ. Both $\hat{e}_\rho$ and $\hat{e}_\phi$, but not $\hat{e}_z$, change orientations in space as P moves on the cylindrical surface.

The position vector ***r*** now takes the form

$$\vec{r} = \rho\hat{e}_\rho + z\hat{e}_z. \tag{1.24}$$

Differentiating once with respect to time t yields the velocity

$$\vec{v} = \mathrm{d}\vec{r}/\mathrm{d}t = \dot{\rho}\hat{e}_\rho + \rho\dot{\hat{e}}_\rho + \dot{z}\hat{e}_z. \quad (1.25)$$

In order to calculate the derivatives of the unit vectors $\hat{e}_\rho$ and $\hat{e}_\phi$ with respect to time, we need to first establish relationships connecting them to $\hat{e}_1, \hat{e}_2$, and $\hat{e}_3$. Such connecting formulas can be established from a study of Figure 1.2a as we did for the plane polar coordinates. However, it is not a trivial matter anymore. An alternative way is to first find the vectors that are tangential to the ρ-, ϕ-, and z-curves. These vectors are given, respectively, by $\partial \boldsymbol{r}/\partial\rho$, $\partial \boldsymbol{r}/\partial\phi$, and $\partial \boldsymbol{r}/\partial z$. Then, the unit vectors $\hat{e}_\rho$, $\hat{e}_\phi$, and $\hat{e}_z$are given by

$$\hat{e}_\rho = \frac{\partial\vec{r}/\partial\rho}{\left|\partial\vec{r}/\partial\rho\right|} = \frac{\cos\phi\hat{e}_1 + \sin\phi\hat{e}_2}{\left[\cos^2\phi + \sin^2\phi\right]^{1/2}} = \cos\phi\hat{e}_1 + \sin\phi\hat{e}_2$$

$$\hat{e}_\phi = \frac{\partial\vec{r}/\partial\phi}{\left|\partial\vec{r}/\partial\phi\right|} = -\sin\phi\hat{e}_1 + \cos\phi\hat{e}_2$$

$$\hat{e}_z = \hat{e}_3$$

from which we find

$$\dot{\hat{e}}_\rho = \dot{\phi}\hat{e}_\phi$$

and

$$\dot{\hat{e}}_\phi = -\dot{\phi}\hat{e}_\rho.$$

Substituting these into Equation 1.25, we obtain the velocity vector in cylindrical coordinates:

$$\vec{v} = \dot{\rho}\hat{e}_\rho + \rho\dot{\phi} + \dot{z}\hat{e}_z. \quad (1.26)$$

Similarly, the acceleration vector in cylindrical coordinates is given by

$$\vec{a} = \mathrm{d}\vec{v}/\mathrm{d}t = \left(\ddot{\rho} - \rho\dot{\phi}^2\right)\hat{e}_\rho + \left(\rho\ddot{\phi} + 2\dot{\rho}\dot{\phi}\right)\hat{e}_\phi + \ddot{z}\hat{e}_z. \quad (1.27)$$

1.7.3 Spherical Coordinates ($\vec{r}$, θ, ϕ)

The spherical coordinates are defined by the following relationships:

$$\left.\begin{aligned} x_1 &= r\sin\theta\cos\phi \\ x_2 &= r\sin\theta\sin\phi \\ x_3 &= r\cos\theta \end{aligned}\right\} \quad \text{or} \quad \left\{\begin{aligned} r &= \left[x_1^2 + x_2^2 + x_3^2\right]^{1/2} \\ \theta &= \tan^{-1}\left(x_1^2 + x_2^2\right)^{1/2}/x_3. \\ \phi &= \tan^{-1}(x_2/x_1) \end{aligned}\right.$$

It is easy to see from Figure 1.2b that θ is the angle that the radius vector $\boldsymbol{r}$ makes with the x_3 axis and that ϕ is the angle that the projection of the position vector $\boldsymbol{r}$ onto the x_1x_2-plane makes with the x_1 axis.

The three unit vectors that we will employ in spherical coordinates are defined by the equation

$$\hat{e}_r = \frac{\partial\vec{r}/\partial r}{|\partial\vec{r}/\partial r|}, \quad \hat{e}_\phi = \frac{\partial\vec{r}/\partial\phi}{|\partial\vec{r}/\partial\phi|}, \quad \hat{e}_\theta = \frac{\partial\vec{r}/\partial\theta}{|\partial\vec{r}/\partial\theta|} \tag{1.28}$$

and their directions are defined in the following manner:

$\hat{e}_r$ is directed along increasing.
$\hat{e}_\theta$ is perpendicular to r in the direction of increasing θ in the plane containing r and z.
$\hat{e}_\phi$ completes the right-handed triad. That is, $\hat{e}_r$ turned into $\hat{e}_\theta$, according to the right-hand rule going in the direction of $\hat{e}_\phi$.

Thus, $\hat{e}_r$ and $\hat{e}_\theta$ are in the same vertical plane with $\hat{e}_3$, and $\hat{e}_\phi$ is parallel to the horizontal x_1x_2-plane.

The position vector ***r*** can be expressed as

$$\vec{r} = r\hat{e}_r$$

from which we obtain

$$\vec{v} = \dot{r}\hat{e}_r + r\dot{\hat{e}}_r \tag{1.29}$$

and

$$\vec{a} = \ddot{r}\hat{e}_r + 2\dot{r}\dot{\hat{e}}_r + r\ddot{\hat{e}}_r. \tag{1.30}$$

From

$$\begin{aligned}\vec{r} &= x_1\hat{e}_1 + x_2\hat{e}_2 + x_3\hat{e}_3 \\ &= r(\sin\theta\cos\phi\hat{e}_1 + \sin\theta\sin\phi\hat{e}_2 + \cos\theta\hat{e}_3)\end{aligned}$$

and through the use of Equation 1.28, we obtain the connecting formula between the two sets of unit vectors:

$$\begin{aligned}\hat{e}_r &= \sin\theta\cos\phi\hat{e}_1 + \sin\theta\sin\phi\hat{e}_2 + \cos\theta\hat{e}_3 \\ \hat{e}_\theta &= \cos\theta\cos\phi\hat{e}_1 + \cos\theta\sin\phi\hat{e}_2 - \sin\theta\hat{e}_3 \\ \hat{e}_\phi &= -\sin\phi\hat{e}_1 + \cos\phi\hat{e}_2.\end{aligned}$$

It is easy to verify that these unit vectors are mutually orthogonal and that they define a right-hand coordinate system:

$$\hat{e}_r \times \hat{e}_\theta = \hat{e}_\phi, \quad \hat{e}_\theta \times \hat{e}_\phi = \hat{e}_r, \quad \hat{e}_\phi \times \hat{e}_r = \hat{e}_\theta.$$

The time derivatives of these three unit vectors are

$$\begin{aligned}\dot{\hat{e}}_r &= \dot{\theta}\hat{e}_\theta + \sin\theta\dot{\phi}\hat{e}_\phi, \\ \dot{\hat{e}}_\theta &= -\dot{\theta}\hat{e}_r + \cos\theta\dot{\phi}\hat{e}_\theta, \\ \dot{\hat{e}}_\phi &= -\sin\theta\dot{\phi}\hat{e}_r - \cos\theta\dot{\phi}\hat{e}_\theta\end{aligned}$$

$$\ddot{\hat{e}}_r = \left(-\sin^2\theta\dot{\phi}^2 - \dot{\theta}^2\right)\hat{e}_r + \left(\ddot{\theta} - \sin\theta\cos\theta\dot{\phi}^2\right)\hat{e}_\theta + \left(2\cos\theta\dot{\theta}\dot{\phi} + \sin\theta\ddot{\phi}\right)\hat{e}_\theta.$$

With these relationships, we finally find the expressions for the velocity and acceleration vectors in spherical coordinates:

$$\vec{v} = \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta + r\sin\theta\dot{\phi}\hat{e}_\phi \tag{1.31}$$

$$\begin{aligned}\vec{a} = &\left(\ddot{r} - r\sin^2\theta\dot{\phi}^2 - r\dot{\theta}^2\right)\hat{e}_r \\ &+ \left(r\ddot{\theta} + 2\dot{r}\dot{\theta} - r\sin\theta\cos\theta\dot{\phi}^2\right)\hat{e}_\theta \\ &+ \left(r\sin\theta\ddot{\phi} + 2\dot{r}\dot{\phi}\sin\theta + 2r\cos\theta\dot{\theta}\dot{\phi}\right)\hat{e}_\phi.\end{aligned} \tag{1.32}$$

1.8 ANGULAR VELOCITY AND ANGULAR ACCELERATION

The basic quantifiers in linear motion are displacement, velocity, and acceleration. The corresponding quantifiers in rotational motion are angular displacement, angular velocity, and angular acceleration.

Angular displacement is the angle traversed by a particle moving in a plane, a circular path about an axis that is perpendicular to the instantaneous direction of motion. For a rotating rigid body with a fixed point, each point of this body, at any given instant of time, moves along a circular path. The center of these circular paths lies along a line passing through the fixed point. This line is the instantaneous axis of rotation. Angular displacement can be measured in degrees, but the equations of rotational motion become much simpler and easier if it is measured in radians. By definition, the angular displacement θ (in radians) and the arc length s of a circle of radius r are connected by the simple relation $\theta = s/r$.

An infinitesimal angular displacement is a vector, which can be proved by showing that two infinitesimal angular displacements obey the commutative law in addition in the same way as linear displacement or velocity vectors. But finite angular displacement does not commute, and hence, we shall not consider them as vectors.

As the particle moves in a circular path, the time rate of change of the angular displacement is taken as the angular velocity ω whose magnitude is

$$\omega = \lim_{\Delta t\to 0}\left|\frac{\Delta\vec{\theta}}{\Delta t}\right| = \left|\frac{d\vec{\theta}}{dt}\right|. \tag{1.33}$$

The angular velocity is a vector, and its magnitude is given by Equation 1.33. We can associate the two possible directions of rotation with the two opposite directions of the rotational axis. The conventional rule is this: The direction of the angular velocity vector is that of the advance of a right-hand screw turning with the rotation along the axis of rotation.

If the angular velocity changes by an amount $\Delta\vec{\omega}$ in a time Δt, the angular acceleration, denoted by $\boldsymbol{\alpha}$, is taken as the angular velocity change per unit time:

$$\vec{\alpha} = \lim_{\Delta t\to 0}\frac{\Delta\vec{\omega}}{\Delta t} = \frac{d\vec{\omega}}{dt}. \tag{1.34}$$

In general, the angular acceleration vector is not necessarily in the direction of the instantaneous angular velocity vector, and the rotational motion is complicated.

1.9 INFINITESIMAL ROTATIONS AND THE ANGULAR VELOCITY VECTOR

The noncommutative property of finite rotations about different axes can be proved with orthogonal transformations. The proof, however, is obvious from an examination of Figure 1.10, which shows the different orientations of a parallelepiped that rotates successively in different orders. We see that the order of operations affects the result, a property incompatible with laws of vector addition. Therefore, we cannot associate a vector with a finite rotation.

The situation is different for infinitesimal rotations that are commutative. We can demonstrate this difference explicitly by using the fact that an infinitesimal rotation is an infinitesimal orthogonal transformation. A vector changes only by infinitesimally small amounts under an infinitesimal orthogonal transformation. This can be expressed as

$$x_i' = x_i + \sum_{j=1}^{3} \varepsilon_{ij} x_j = \sum_{j=1}^{3} (\delta_{ij} + \varepsilon_{ij}) x_j, \quad j = 1,2,3$$

where ε_{ij} are infinitesimal, and δ_{ij} is the Kronecker delta symbol. In matrix notation, the last equation becomes

$$\tilde{x}' = (\tilde{I} + \tilde{\varepsilon})\tilde{x} \tag{1.35}$$

where

$$\tilde{I} = [\delta_{ij}] = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \varepsilon = [\varepsilon_{ij}] = \begin{pmatrix} \varepsilon_{11} & \varepsilon_{12} & \varepsilon_{13} \\ \varepsilon_{21} & \varepsilon_{22} & \varepsilon_{23} \\ \varepsilon_{31} & \varepsilon_{32} & \varepsilon_{33} \end{pmatrix}, \quad \tilde{x} = \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Now, if $\tilde{I} + \tilde{\varepsilon}_1(\delta\theta_1)$ and $\tilde{I} + \tilde{\varepsilon}_2(\delta\theta_2)$ are two infinitesimal orthogonal transformations corresponding, respectively, to $\delta\theta_1$ and $\delta\theta_2$, then

$$(\tilde{I} + \tilde{\varepsilon}_1)(\tilde{I} + \tilde{\varepsilon}_2) = \tilde{I}^2 + \tilde{I}(\tilde{\varepsilon}_1 + \tilde{\varepsilon}_2) + O(\tilde{\varepsilon}_1\tilde{\varepsilon}_2) = \tilde{I} + \tilde{\varepsilon}_1 + \tilde{\varepsilon}_2 + O(\tilde{\varepsilon}_1\tilde{\varepsilon}_2)$$

where $O(\tilde{\varepsilon}_1\tilde{\varepsilon}_2)$ is the usual higher-order-of-magnitude symbol. If the order of rotations was reversed, we would have

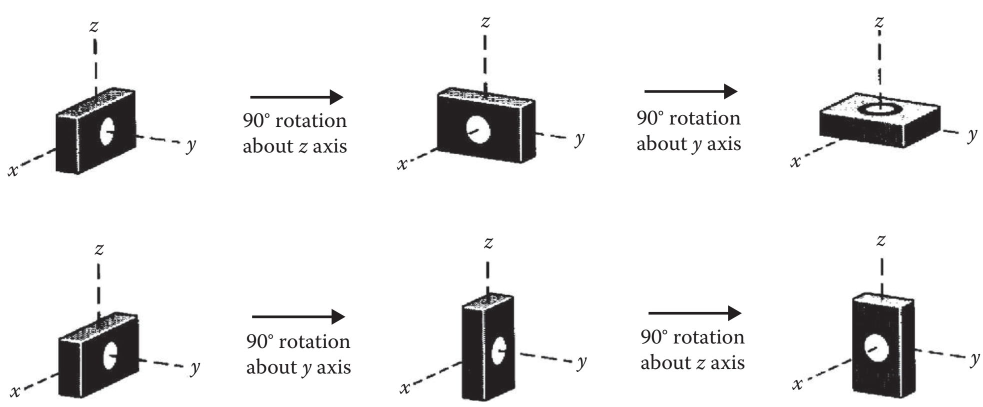


FIGURE 1.10 Rotation of a parallelepiped about coordinates axes.

$$(\tilde{I}+\tilde{\varepsilon}_2)(\tilde{I}+\tilde{\varepsilon}_1)=\tilde{I}^2+\tilde{I}(\tilde{\varepsilon}_2+\tilde{\varepsilon}_1)+O(\tilde{\varepsilon}_1\tilde{\varepsilon}_2)=\tilde{I}+\tilde{\varepsilon}_2+\tilde{\varepsilon}_1+O(\tilde{\varepsilon}_1\tilde{\varepsilon}_2).$$

We see that the two infinitesimal rotations commute to the first order, so we can associate a vector (actually an axial vector) with an infinitesimal rotation. An "axial vector" is a quantity with three components that, relative to a given coordinate system, do not change sign when all three coordinate axes are reflected. On the other hand, a "polar vector" changes sign under a coordinate reflection. Examples of polar vectors are displacement, velocity, and force. Alternatively, we can show the commutative property of infinitesimal rotations using a straight but lengthy method. To this end, consider a particle moving in a circular path of radius b about an axis perpendicular to the plane of motion. Vector $\vec{r}$ is the position vector of the particle relative to an origin located at an arbitrary point O on the axis of rotation $\hat{e}$ (Figure 1.11). Now, an infinitesimal rotation $\delta\theta_1$ about the axis $\hat{e}$ carries the particle from P to Q through an infinitesimal displacement $\delta\vec{r}_1$ whose magnitude is given by

$$\delta r_1 = b\delta\theta_1 = (r \sin \phi)\delta\theta_1.$$

Furthermore, $\delta\vec{r}_1$ is perpendicular to both $\boldsymbol{r}$ and $\hat{e}$; this geometrical situation is correctly represented by

$$\delta\vec{r}_1 = \delta\vec{\theta}_1 \times \vec{r} \tag{1.36}$$

where

$$\delta\vec{\theta}_1 = \delta\theta_1\hat{e}.$$

The position vector of point Q is

$$\vec{r}_1 = \vec{r} + \delta\vec{r}_1. \tag{1.37}$$

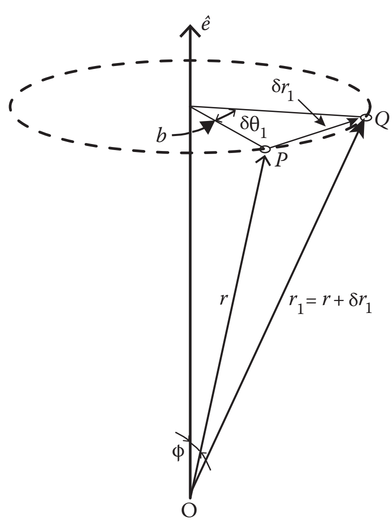


FIGURE 1.11 Infinitesimal rotation.

If this is followed by a second infinitesimal rotation $\delta\theta_2$ about a different axis $\hat{e}'$ through O, the final radius vector $\boldsymbol{r}_{12}$ of the particle for $\delta\theta_1$ followed by $\delta\theta_2$ is

$$\begin{aligned}\vec{r}_{12} &= \vec{r}_1 + \delta\vec{r}_2\\ &= \vec{r}_1 + (\delta\vec{\theta}_2 \times \vec{r}_1)\\ &= (\vec{r} + \delta\vec{\theta}_1 \times \vec{r}) + \delta\vec{\theta}_2 \times (\vec{r} + \delta\vec{\theta}_1 \times \vec{r})\\ &= \vec{r} + \delta\vec{\theta}_1 \times \vec{r} + \delta\vec{\theta}_2 \times \vec{r} + \delta\vec{\theta}_2 \times (\delta\vec{\theta}_1 \times \vec{r}).\end{aligned}$$

We can write $\boldsymbol{r}_{12}$ as

$$\vec{r}_{12} = \vec{r} + \delta\vec{r}_{12} \tag{1.38}$$

with

$$\delta\vec{r}_{12} = \delta\vec{\theta}_1 \times \vec{r} + \delta\vec{\theta}_2 \times \vec{r} + \delta\vec{\theta}_2 \times (\delta\vec{\theta}_1 \times \vec{r}). \tag{1.39}$$

If the orders of rotations were reversed, that is, if $\delta\theta_2$ was followed by $\delta\theta_1$, then we would have

$$\vec{r}_{21} = \vec{r} + \delta\vec{r}_{21} \tag{1.40}$$

with

$$\delta\vec{r}_{21} = \delta\vec{\theta}_{21} \times \vec{r} + \delta\vec{\theta}_1 \times \vec{r} + \delta\vec{\theta}_1 \times (\delta\vec{\theta}_2 \times \vec{r}). \tag{1.41}$$

For infinitesimal rotations, we can neglect the higher-order terms; then Equations 1.39 and 1.41 become identical, and $\vec{r}_{12} = \vec{r}_{21}$. We get the same result: the two infinitesimal rotations commutative.

Some simple relations between linear and rotational motions can be established with the help of Equation 1.39 with the higher-order terms neglected. If we divide both sides by Δt and take the limit, we obtain the particle's velocity at the final position:

$$\vec{v} = (\vec{\omega}_1 + \vec{\omega}_2) \times \vec{r} = \vec{\omega} \times \vec{r} \tag{1.42}$$

where

$$\vec{\omega} = \vec{\omega}_1 + \vec{\omega}_2.$$

That is, the motion of the particle can be described by a single rotation with angular velocity $\vec{\omega}$.

If we differentiate Equation 1.42 with respect to time t, we obtain a simple relationship that connects the angular and linear accelerations:

$$\vec{a} = \vec{\alpha} \times \vec{r} + \omega \times \vec{v} = \vec{\alpha} \times \vec{r} + \omega \times (\vec{\omega} \times \vec{r}) \tag{1.43}$$

where the first term, $\vec{\alpha} \times \vec{r}$, is perpendicular to the radius vector $\boldsymbol{r}$, and the second term can be simplified to

$$\omega \times (\vec{\omega} \times \vec{r}) = \vec{\omega}(\vec{\omega} \cdot \vec{r}) - \vec{r}\omega^2$$

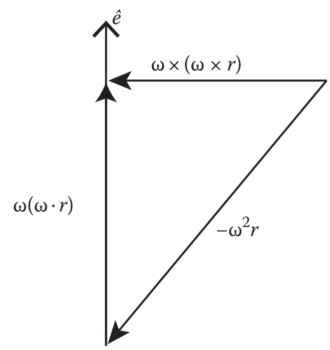


FIGURE 1.12 Centripetal acceleration.

which points to the center of rotation and is called the centripetal acceleration (Figure 1.12). Further insight into the nature of centripetal and centrifugal accelerations can be found in a later chapter on motion in noninertial systems (Figure 1.12).

PROBLEMS

1. A particle moves in space along a path whose parametric equations are given by
   $x_1 = b\sin(\omega t), \quad x_2 = b\cos(\omega t), \quad x_3 = c$
   where b and c are constants.
   (a) Find its position at vector $\vec{r}$, velocity $\vec{v}$, and acceleration $\vec{a}$ at any time t.
   (b) Show that the particle traverses its path with constant speed and that its distance from the origin remains constant.
   (c) Show that the acceleration is perpendicular to the velocity and the x_3-axis.
   (d) Determine the trajectory of the particle, that is, the path described in terms of spatial coordinates only. Sketch the trajectory.
2. A ladder of length b rests against a vertical wall (Figure 1.13). The base of the ladder is pulled away with a constant speed v_0.
   (a) Show that the midpoint M of the ladder describes the arc of a circle of radius $b/2$ with the center at O.
   (b) Find the velocity of the midpoint M at the moment where the base of the ladder is at a distance a $< b$ from the wall.

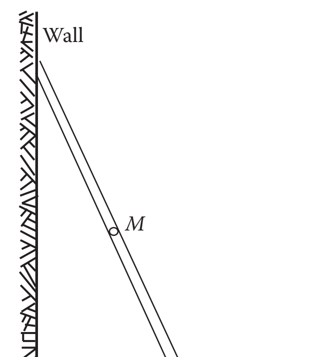


FIGURE 1.13 A ladder rests against a vertical wall.

3. The speed of a train increases at a constant rate A from 0 to V, then remains constant for an interval, and finally decreases to 0 at a constant rate B. If S is the total distance traveled, prove that the total time elapsed is

$$\frac{S}{V}+\frac{1}{2}V\left(\frac{1}{A}+\frac{1}{B}\right).$$

4. Given two matrices,

$$\tilde{A}=\begin{vmatrix} 3 & 4 \\ -4 & 3 \end{vmatrix}, \quad \tilde{B}=\begin{vmatrix} 3 & 1 \\ 1 & 3 \end{vmatrix}$$

determine $\tilde{A}\tilde{B}$ and $\tilde{B}\tilde{A}$.
Is $\tilde{A}\tilde{B}=\tilde{B}\tilde{A}$?

5. A particle moves along a path such that its radius vector is given by

$$\vec{r}(t)=a\sin(kt)\hat{e}_1+b\cos(kt)\hat{e}_2$$

where a, b, and k are constants. Show that the path is an ellipse of axes $2a$ and $2b$.
Show also, by expressing the acceleration in terms of the vector $\vec{r}$, that the acceleration is directed toward the origin at all times.

6. The position of a particle is given by

$$\vec{r}=A(e^{\alpha t}\hat{e}_1+e^{-\alpha t}\hat{e}_2)$$

where α is a constant. Find the velocity and sketch the trajectory. (Hint: In sketching the motion of a particle, it is usually helpful to look at limiting cases as $t \to 0$ and as $t \to \infty$.)

7. The position of a particle at any time t is given by $x = a$, $y = bt^2$, where a and b are constants. Find the rectangular and polar components of its velocity and acceleration (a) at any time t, (b) when $a = 12$ cm, $b = \mathrm{I}$ cm/s^2, and when $t = 3$ s.

8. Two particles have position vectors given by

$$\vec{r}_1=2t\hat{e}_1-t^2\hat{e}_2+(3t^2-4t)\hat{e}_3, \quad \vec{r}_2=(5t^2-12t+4)\hat{e}_1+t^2\hat{e}_2-3t\hat{e}_3.$$

Find (a) the relative velocity and (b) the relative acceleration of the second particle with respect to the first at time $t = 2$ s.

9. A particle moves along a space curve C with a position vector given by

$$\vec{r}=3\cos 2t\hat{e}_1+3\sin 2t\hat{e}_2+(8t-4)\hat{e}_3.$$

(a) Find a unit tangent vector $\hat{T}$ to the curve C.
(*b*) Show that $\vec{v}=v\hat{T}$.

10. Find the (a) curvature, (b) radius of curvature, and (c) unit normal $\hat{e}_n$ to any point of the space curve C in Problem 9.

11. The runway of an airport faces north. A light airplane is flying south at a speed of 150 km/h at a height of 3 km and is on a straight path that passes over the runway. At the time the airplane is 3 km from the end of the runway, a jet takes off from the airport and climbs at a constant angle α to the horizontal at a constant speed of 600 km/h. Find the angle α that results in a head-on collision between the two aircraft.

12. A particle moves at a constant acceleration along the curve $x_2 = x_1/100$ from point A to point B. The velocity of the particle at A is 10 m/s, and 10 s later, at point B it is traveling at 50 m/s. Determine the total acceleration of the particle at point B.
13. Prove that the velocity of a particle in three-dimensional space is independent of the fixed origin, to which its position vector is referred. (Hint: Let O′ be a second fixed origin with position vector $\boldsymbol{a}$ relative to origin O.)
14. The position of a particle moving in a plane is given by

$$x = \sin(\omega t) \; y = \cos(\alpha\omega t).$$

    Show that the trajectory repeats itself periodically only if α is a rational number.
15. Show that finite rotation is not a vector quantity.

2 Newtonian Mechanics

The foundation of Newtonian mechanics is the three laws of motion that were postulated by Isaac Newton (1642–1727) as a result of the combination of experimental evidence and a great deal of intuition.

2.1 THE FIRST LAW OF MOTION (LAW OF INERTIA)

Newton's first law of motion describes the behavior of a body that has no net outside force acting on it. This law may be stated as follows:

> A body remains in a state of rest or in uniform motion in a straight line unless it is compelled to change that state by an applied force.

This contradicts the view of motion held by Aristotle, the Greek philosopher who lived in 340 B.C. He said that the natural state of an object was to be at rest and that it moved only if driven by a force. Was Aristotle right? We all observe that moving objects tend to slow down and come to rest. But this is because of friction. If friction is negligible, a body in motion will remain in motion. Galileo Galilee (1564–1642) first observed this. Rather than rely solely on intuition and general observation as Aristotle did, Galileo tested his ideas with carefully designed experiments and started the modern chain of development in the study of motion. Galileo used two smooth inclined planes set end-to-end, one tilted down and the other up, and released balls of different weights, from rest, down the first inclined plane. (It is similar to that of balls falling vertically, but it is easier to observe because the speeds are slower.) When a ball was released and moved down the first inclined plane, it swept past the bottom and ascended the second inclined plane to about the same height (a little less—friction cannot be eliminated completely) regardless of the tilt of the second plane. He also observed that when the angle of the incline of the second plane was made less steep, the ball slid farther in the horizontal direction (Figure 2.1). Galileo then asked the following question: What would happen if the second incline were made frictionless and horizontal? The ball would continuously slide forever with a constant speed, provided the plane was made infinitely long.

Galileo did not extend his idea beyond motion under the gravitational force of the Earth. It was Isaac Newton who made the grand generalization of Galileo's result as his first law of motion.

The first law of motion is so familiar that we sometimes tend to lose sight of its true significance and complain of its lack of content. In fact, the insight into natural processes that the first law expresses is important. In contrast to the view held by his contemporary scientists or those before them, Newton equates the state of a body at rest with that of a body in uniform motion in a straight line. No external force is necessary in order to maintain uniform motion; it continues without change because of an intrinsic property of all bodies that we call inertia. Because of this property, the first law is also known as the "law of inertia."

Newton made the first law more precise by introducing definitions of "quantity of motion" and "amount of matter," which we now call momentum and mass, respectively. The momentum of a body is simply proportional to its velocity $\vec{v}$. The coefficient of proportionality is a constant for any given body and is called its mass. Denoting the body's momentum vector by $\vec{p}$ and its mass by m, we can write

$$\vec{p} = m\vec{v}. \tag{2.1}$$

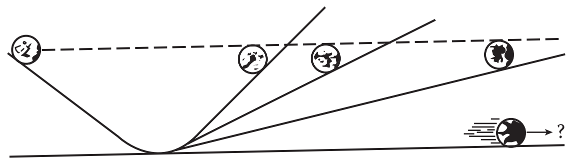


FIGURE 2.1 Galileo's two inclined planes experiment.

The first law now can be expressed mathematically by the equation

$$\vec{p} = m\vec{v} = \text{constant} \tag{2.2}$$

in the absence of an external force acting on the body. This is also a law of conservation of momentum.

For a group of particles that interact with one another but not with anything outside the group, the individual particle momentum may vary with time. However, the total momentum of the whole group does not vary with time: That is, the total momentum P of a closed system (sometimes called an isolated system) is conserved:

$$\vec{P} = \vec{p}_1 + \vec{p}_2 + ... + \vec{p}_n = \sum_i \vec{p}_i = \text{constant}. \tag{2.3}$$

We must note here that, in the theory of special relativity, the expression for the momentum is more complicated than Equation 2.1:

$$\vec{p} = \frac{m_0\vec{v}}{\sqrt{1 - v^2/c^2}}$$

where m_0 is the so-called rest mass of a particle (its mass at rest), and c is the speed of light in a vacuum. The relativistic expression for the momentum can still be written in a form similar to Equation 2.1 if we set

$$m(v) = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

$m(v)$ is known as the relativistic mass of the particle. We should not, however, attribute the relativistic increase in momentum to an increase in mass. A detailed discussion can be found in Chapter 13 on special relativity.

2.1.1 Inertial Frames of Reference

At first sight, the first law of motion is a clear statement that can be easily put to a test. But upon further thought, it is not so clear. With respect to what reference does the body remain at rest or in uniform motion? When two bodies fall side by side, one of the two bodies remains at rest relative to the other while, at the same time, it is subject to the force of gravity. Such cases would contradict the first law. Thus, the first law cannot generally hold relative to all reference frames. The first law is correct only in regard to inertial reference frames, which are not subject to acceleration and in which the free motion of the body occurs at a constant velocity. Some physicists suspect that Newton separated the laws of motion with the intention of having the first law serve to define the inertial reference. The remaining laws are then understood to be valid only in such a frame of reference.

The question then arises: How is it possible to determine whether or not a given frame of reference is an inertial frame? The answer is not quite as trivial as it might seem at first sight for, in order to eliminate all forces on a body, it would be necessary to isolate the body completely. In practice, this is impossible because there are at least some forces of gravity acting on it unless the body is removed infinitely far away from all other matter. Therefore, in practice, we merely specify an approximate inertial frame in accordance with the needs of the problem under investigation. For elementary applications in the laboratory, a frame attached to the Earth usually suffices. This frame is, of course, an approximate inertial frame resulting from the daily rotation of the Earth on its axis and its revolution around the sun. The fact that the Earth does not provide an inertial reference frame is not important for the operation of something like a particle accelerator. The total time required to accelerate, say, an electron from rest to its final energy is less than 1 s, so the change in velocity of the machine resulting from the acceleration of the Earth is very small during this time. What is worth noticing is that while during any 1 s the accelerator itself can be regarded as an inertial reference frame, it is a very different inertial frame at one time than it was 6 months earlier because of the orbital motion of the Earth. If all inertial reference frames were not equivalent, the accelerator ought to have very different characteristics in the spring and fall and less different, but still distinguishable, characteristics in the morning and evening. The complete absence of such effects is a strong indication that all inertial reference frames are equivalent.

2.2 THE SECOND LAW OF MOTION; THE EQUATIONS OF MOTION

Newton's second law of motion describes the behavior of a body when an external force acts upon it. Galileo's pioneer work on inclined planes paved the way for Newton to formulate the second law of motion. Before Galileo, velocity was the main feature of motion; then Galileo made a giant step when he realized that acceleration, the rate at which the velocity of a body changes with time, was the quantity characterizing the motion of a body when an external force acts upon it. Acceleration is much harder to identify than velocity when simply observing the motion of a body. In the inclined-plane experiment, Galileo observed that the same force (weight) acted upon on the ball as it rolled down the inclined plane. And the effect of the force was to make the ball constantly speed up. This showed that the real effect of a force is to change the velocity of a body rather than just to set it moving as claimed by Aristotle. But Galileo did not extend his observation beyond motion under the influence of gravity. Newton made a big step in realizing that many different external forces, such as friction or the act of being thrown, could produce acceleration. He also recognized that the mass of a body was also an important factor in describing the motion of the body. The same force acting on a body of twice the mass will produce half the acceleration. Newton's second law of motion states the following:

> Change of motion is proportional to the external applied force and takes place in the direction of the straight line in which the force acts.

The term "motion" used here by Newton really means "quantity of motion," which we now commonly refer to as momentum. Thus, "change of motion" may be translated into "change of momentum," and the second law of motion is now written mathematically as

$$\vec{F} = \frac{\mathrm{d}\vec{p}}{\mathrm{d}t} = \frac{\mathrm{d}(m\vec{v})}{\mathrm{d}t} \tag{2.4a}$$

where $\vec{F}$ is the force, and m is the mass. Equation 2.4a reduces, if m is constant, to

$$\vec{F} = m\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = m\vec{a}, \tag{2.4b}$$

which is a familiar result from basic physics. The differential equation 2.4a or 2.4b is known as the equation of motion of the particle of mass m.

As we pointed out earlier, when a body is moving at relativistic speed, its mass is not constant but is a function of the speed of the body; Equations 2.4a and 2.4b are then not equivalent.

We may write the second law in terms of position vector $\vec{r}\,(t)$:

$$m\frac{\mathrm{d}^2\vec{r}(t)}{\mathrm{d}t^2}=\vec{F}. \tag{2.5}$$

It is called the equation of motion of the particle, a second-order differential equation. If the force $\vec{F}$ is known, and if the position and velocity of the particle at some instant are given, the second law enables us to calculate the state of the particle at all other times, past or future. Thus, Newtonian mechanics is deterministic. This deterministic picture of the world was abandoned in quantum mechanics. Even in the classical case, doubts about the determinism have recently arisen in connection with the study of chaotic systems. Such system's future motion depends so sensitively on the initial conditions that it can be impossible to predict that future motion even qualitatively for large but finite times from initial conditions determined with only finite precision. More discussion can be found in Chapter 9 on nonlinear oscillations.

The solution of the second-order differential equation 2.5 for any given force $\vec{F}$ is a fairly formidable task. If we restrict ourselves to seeking analytic solutions, we are able to deal only with a small number of interesting physical problems. Approximating the second law by a sufficiently accurate approximation form is one of the arts in physics that we all should learn well. With the help of computers, we can also use numerical analysis to handle incredibly complicated physics problems. However, we should remember that a purely numerical solution cannot shed light on the mechanisms operating in a physical situation. Therefore, before seeking a numerical solution, try the approximation method.

2.2.1 The Concept of Force

To date, there is no generally accepted interpretation of the second law of motion. The difficulty is that there are no independent ways to determine the quantity $\vec{F}$. Newton gave a somewhat ambiguous discussion in the *Principia* on which conceptual interpretation to adopt. It has been suggested that the second law itself may be taken as the definition of force as well as a law of motion. But the consequence of this approach is that the laws of motion then become merely a convention and not an assertion about the workings of nature; the whole formalism of Newtonian mechanics becomes axiomatic in the sense that all results follow from this definition rather than from an experimentally deduced law of nature.

We must stress the fact that the second law, like the other two laws of motion, is an experimental one. It was formulated from generalizations of the experimental data and observations. We share the opinion of Arnold Sommerfeld, Richard Feynman, and many other physicists that we should not use the second law as a definition of force and, thus, make mechanics a pure mathematical theory. No prediction can be made from a definition! The second law acquires real significance only when the force is independently defined. But completely describing the specific independent properties of force is not a trivial task. Everyone has an intuitive feeling for the concept of a force. We can take force, like length, as a primitive concept and define it operationally in terms of, for example, the compression or expansion of a "standard" spring by some specific amount. Then Newton's laws of motion are laws, and so are the laws of theories of special forces.

Most of the forces that are found in nature are now well understood. There are only four basic forces: strong nuclear, electromagnetic, gravitational, and weak nuclear. Of these four, only the electromagnetic and gravitational forces properly belong to the domain of classical mechanics. All other apparently different forces in the classical domain are manifestations of either electromagnetic or gravitational force. As an example, the friction force between two bodies ultimately results from an electrostatic force between the charged particles that make up the atoms of the bodies.

Gravitational interaction is a property of all bodies whether they are electrically charged or neutral and is determined only by the masses of the bodies. In Chapter 6 on central force motion, we shall learn how Newton deduced the law of gravitation from Kepler's three laws of planetary motion. He found that the gravitational force between two particles is inversely proportional to the square of the distance between them and proportional to the product of the masses of the two particles. As shown in Figure 2.2, if we denote the masses of the particles by m_1 and m_2 and the distance between them by r_{12}, and $\vec{F}_{12}$ is the force on m_1 by m_2, then Newton's law of gravity may be written as

$$\vec{F}_{12} = -\frac{Gm_1m_2}{r_{12}^2}\hat{r}_{12} \tag{2.6}$$

in which the origin is set at m_2, and $\hat{r}_{12} = (r_{12} / 1r_{12}1)$ is a unit vector pointing along the radius vector from the origin. The minus sign indicates that the force is always attractive. The quantity G is a universal coefficient of proportionality independent of the nature of the interacting bodies. It is called the gravitational constant, and its value in the CGS system is 6.67×10^{-8} cm^3/g s^2.

The extremely small value of G shows that the force of gravitational attraction becomes considerable only for very large masses. Thus, gravitational force plays no great part in the mechanics of atoms and molecules, where the Coulomb forces are much larger in comparison. The static Coulomb force between two point charges q_1 and q_2 is similar in mathematical form to the gravitational force law:

$$F_{12} = q_1q_2/r_{12}. \tag{2.7}$$

This force is attractive if the charges have opposite signs and is repulsive if the charges are of the same sign. The proportionality coefficient in Coulomb's law has been made equal to unity by the appropriate choice of the unit of charge. A charge q moving in an electromagnetic field E and B experiences the Lorentz force:

$$\vec{F} = q(\vec{E} + \vec{v} \times \vec{B}/c) \tag{2.8}$$

where $\vec{v}$ is the velocity of the charge q.

The gravitational and electromagnetic forces are fundamental forces; they cannot be reduced to other, simpler forces. Elastic and friction forces, for example, are not fundamental, and we can obtain only approximate empirical formulas for them.

The elastic force has a wide range of applications. Any body changes its dimensions and shape under the action of an externally applied force. A body that regains its original dimensions and shape when the externally applied force is removed is an elastic body. An elastic body has a definite elastic limit beyond which the deformation will become permanent. Within the elastic limit, the elastic cord or spring obeys Hooke's law.

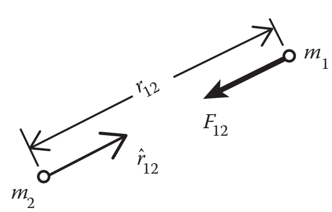


FIGURE 2.2 Two gravitational interaction particles.

$$F = -k\Delta x. \tag{2.9}$$

The constant of proportionality k is called the spring constant and is dependent on the properties of an elastic body, and Δx is the extension or compression of the spring from its relaxed position.

Friction forces play a crucial role in damping or retarding motion initiated by other forces. The friction force between two contact bodies results from the interaction between the surface molecules of the two bodies and involves a very large number of such interactions. The phenomenon is therefore complex and depends on factors such as the condition and nature of the surfaces and their relative velocity. These factors can be determined collectively through experimentation and are represented approximately by the coefficient of friction. If N is the normal force pressing one body against the other (Figure 2.3) and f is the coefficient of friction, the magnitude of the friction force F_f is found experimentally:

$$F_f = fN. \tag{2.10}$$

In general, there are two kinds of coefficients of friction: the coefficient of static friction f_s and the coefficient of kinetic friction f_k. The coefficient of static friction multiplied by the normal force gives the minimum force required to set in relative motion two bodies initially in contact and at relative rest. The coefficient of kinetic friction multiplied by the normal force gives the force required to maintain the two bodies in uniform relative motion. For most materials, the coefficient of kinetic friction is, in general, smaller than the coefficient of static friction.

It is often more convenient to make use of the angle of friction, ε, which is defined as

$$\tan\varepsilon = \frac{f_s N}{N} = f_s. \tag{2.11}$$

Example 2.1

A block of mass m rests on a rough plane inclined at an angle α to the horizontal. A force $\vec{F}$ is exerted horizontally on the block, and its line of action passes through the geometrical center of the block. Find the minimum value of $\vec{F}$ at which the block is at the point of slipping down the plane and the maximum value of $\vec{F}$ at which the block is at the point of sliding up the plane.

Solution:

(a) Minimum force $\vec{F}_1$

The forces acting on the block are $\vec{F}_1$ (to be found), the weight mg, and the reaction force $\vec{N}$ of the plane on the block (not shown). Resolve all forces into components that are parallel

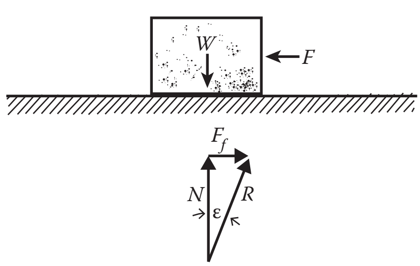


FIGURE 2.3 Force of friction opposes the motion.

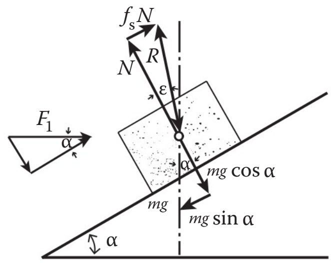


FIGURE 2.4 A block rests on a rough inclined plane: minimum force F_1.

and perpendicular to the plane (Figure 2.4). Because there is no motion along the perpendicular direction, we have

$$F_1 \sin \alpha + mg \cos \alpha = N. \tag{2.12}$$

$\vec{F}_1$ is the minimum value of $\vec{F}$ at which the block is just about to slip down the plane, so that

$$F_1 \cos \alpha + f_s N = mg \sin \alpha. \tag{2.13}$$

Eliminating N between the above two equations, we obtain

$$F_1 (\cos \alpha + f_s \sin \alpha) = mg (\sin \alpha - f_s \cos \alpha). \tag{2.14}$$

In terms of the angle of friction, $f_s = \tan \varepsilon = \sin \varepsilon / \cos \varepsilon$. Equation 2.14 becomes

$$F_1 = \frac{mg(\cos\varepsilon \sin\alpha - \sin\varepsilon \cos\alpha)}{\cos\alpha \cos\varepsilon + \sin\varepsilon \sin\alpha} = mg \tan(\alpha - \varepsilon).$$

(b) Maximum force $\vec{F}_2$

The maximum value $\vec{F}_2$ of the force $\vec{F}$ can be obtained in a similar manner. But in this case, because the block is at the point of sliding up the plane, the friction force is therefore directed down the plane (Figure 2.5). Again taking components parallel and perpendicular to the plane, we obtain

$$\begin{aligned} F_2 \sin\alpha + mg \cos\alpha &= N' \\ F_2 \cos\alpha &= f_s N' + mg \sin\alpha \end{aligned}.$$

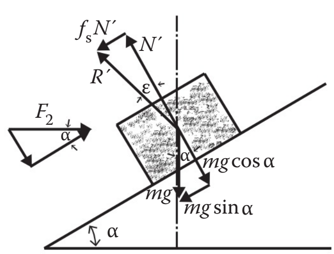


FIGURE 2.5 A block rests on a rough inclined plane: maximum force F_2.

In the same manner as (a), we finally obtain

$$F_2 = mg \tan(\alpha + \varepsilon).$$

2.3 THE THIRD LAW OF MOTION

Newton's second law of motion is an assertion about the workings of nature, and the force is supposed to have some independent properties that Newton did not completely describe. But Newton did give us one rule about force: the third law. He observed that forces always occur paired in nature. Friction requires the presence of something to rub against; the falling apple attracts the Earth, and the Earth also attracts the apple. This sort of observation led Newton to formulate the third law, which states that the forces that two bodies exert on each other are always equal in magnitude and oppositely directed along the same straight line. The third law is often simply stated as "For every action, there is always an equal and opposite reaction." It should be emphasized that action and reaction never act on the same object.

According to the second law, force is the rate of change of momentum with time, so the third law can be written as

$$\frac{\mathrm{d}\vec{p}_1}{\mathrm{d}t} = -\frac{\mathrm{d}\vec{p}_2}{\mathrm{d}t} \tag{2.15}$$

where $\vec{p}_1$ is the momentum of the first body, similar to $\vec{p}_2$, and the minus sign on the right-hand side indicates that the two forces are oppositely directed. If m_1 and m_2 are constant, Equation 2.15 becomes

$$\frac{m_1\mathrm{d}\vec{v}_1}{\mathrm{d}t} = -\frac{m_2\mathrm{d}\vec{v}_2}{\mathrm{d}t}$$

from which we have

$$\frac{m_2}{m_1} = -\frac{|\vec{a}_1|}{|\vec{a}_2|}. \tag{2.16}$$

Equation 2.16 is a very useful relationship, which can be used to determine the ratio of the masses of two bodies from the measurement of their accelerations. If m_1 is selected as the unit mass, then the mass of the other body can be determined.

The third law breaks down when the interaction between bodies is velocity dependent. In general, a velocity-dependent interaction is noncentral. As an example, consider charges in motion. Each moving charge is equivalent to a current element, and the force between them involves a vector product and so is not directed toward contrary parts, that is, along the line joining them.

2.3.1 The Concept of Mass

The laws of motion involve not only the concept of velocity and acceleration but also the concept of mass and force. Thus far, we have avoided discussing the concept of mass. Now, it is time to answer the question: What is the meaning of mass? This is not quite as trivial as it might seem at first sight. Even Newton did not address the meaning of mass clearly, only stating that the mass of a body is the product of its volume and its density. To us, accustomed to defining density as mass per unit of volume, this definition is circular. One often finds that, in some textbooks on mechanics and general physics, mass is defined as the amount of matter in a body or meets the definition of mass as inertia. What is meant by the words “amount of matter”? It conveys nothing unless a method for ascertaining the amount can be specified. Similarly, “mass as inertia” only provides a qualitative notion.

It is a general principle of physics that any quantity introduced into a theory should be, at least in principle, measurable. Thus, we are interested in giving an operational definition of mass. In this regard, Equation 2.16 is a very useful relationship. If we isolate two bodies effectively and compare their mutually induced acceleration, the accelerations are oppositely directed and inversely proportional to the masses as given by Equation 2.16. If the mass of one body, say, m_1, is selected as the unit mass, then the mass of the other body can be determined. This is, in fact, one of the commonly used methods for comparing masses. For example, we can allow two small bodies or particles to collide; during the collision, the effects of more remote bodies on the two colliding bodies are generally negligible, and we may treat the two colliding bodies as an effectively isolated system. The mass ratio can then be determined by Equation 2.16 or its immediate consequence, the law of conservation of momentum:

$$m_1\vec{v}_1 + m_2\vec{v}_2 = \text{constant}.$$

The definition of mass given above is essentially that of Ernst Mach. And the mass of a body as determined this way is termed its inertial mass because it characterizes the inertial properties of bodies. Mass also comes up in Newton’s law of gravity. In this case, it describes the gravitational properties—the ability of bodies to attract one another—and is called the gravitational mass of the body. The question naturally arises: Is the inertial mass of a body equal to its gravitational mass? Galileo and Newton tried to answer this question by conducting experiments. Finally, Roll et al. (1967) and Braginski and Panov (1971) showed that these masses are equal to within a few parts of 10^{12}. The assertion of the equivalence of the gravitational mass and the inertia mass of a body is known as the “principle of equivalence.” Pursuit of the consequences of the principle of equivalence eventually led Einstein to formulate his general theory of relativity between 1911 and 1916.

Example 2.2

A train has four boxcars; each boxcar has a mass M, and the locomotive applies a net force $\vec{F}$ on the train (Figure 2.6). Find the force on each boxcar.

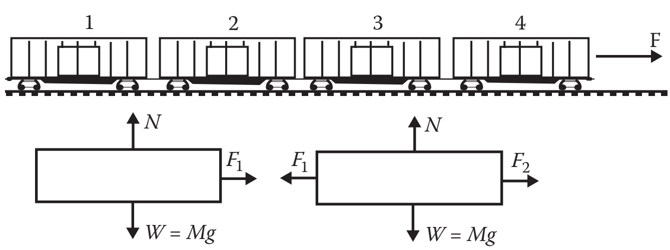


FIGURE 2.6 A net force F acts on a train of four box cars.

Solution:

The acceleration of the train is given by

$$a = \frac{F}{4M}$$

which is also the acceleration acting on the boxcar. Now consider the motion of boxcar 1.

The forces acting upon it are the weight $W = Mg$, the normal reaction of the track upon the boxcar $\vec{N}$, and the force $\vec{F}_1$ exerted by boxcar 2 on boxcar 1 (the force exerted by boxcar 1 on boxcar 2 is $\vec{F}_1'$; $\vec{F}_1$ and $\vec{F}_1'$ are action–reaction pairs). The second law gives

$$N - W = 0 \text{ (along the } y\text{-direction)}$$

and $F_1 = Ma = M\left(\frac{F}{4M}\right) = \frac{F}{4}$ (along the x-direction).

Next, consider the motion of boxcar 2. The forces acting on it are F_1', N, W, and F_2 (the force exerted on boxcar 2 by boxcar 3). The second law gives

$$N - W = 0$$

$$F_2 - F_1' = Ma.$$

But F_1 and F_1' are action–reaction pairs, and $F_1' = F_1 = F/4$, so

$$F_2 = F_1' + Ma = F/4 + M(F/4M) = 2F/4.$$

Similarly, we can show that $F_3 = 3F/4$ and $F_4 = 4F/4$. In general, if there are n boxcars, the acceleration a would be

$$a = F/nM \text{ to the right,}$$

and the forward force on the kth car would be

$$F = kF/n.$$

The net force acting on each car is simply F/n.

2.4 GALILEAN TRANSFORMATIONS AND GALILEAN INVARIANCE

In the use of the equation of motion, a question arises naturally: Does the equation retain its form invariant with respect to changes in the coordinate system? The equation does retain its form invariant under the so-called Galilean transformations. To see this, let us consider two reference frames O and O' with O' moving with a constant velocity V relative to O, which is assumed to be an inertial frame. An inspection of Figure 2.7 gives

$$\vec{r}' = \vec{r} - \vec{V}t. \tag{2.17}$$

We also have $t' = t$.

These transformations are called Galilean transformations. The first of these relationships is commonly called a boost; we take the second $t' = t$ for granted just as Newton did. This latter relationship asserts that time is not affected by relative motion; that is, it expresses the universality, or absolute, of time.

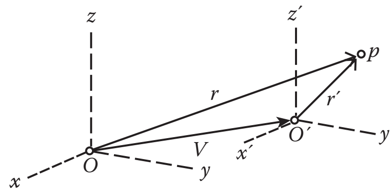


FIGURE 2.7 O' moving with constant velocity V relative to O.

Differentiating Equation 2.17 once, we obtain

$$\dot{\vec{r}}' = \dot{\vec{r}} - \vec{V}. \tag{2.18}$$

Thus, if the velocity $\dot{\vec{r}}$ of a body P in the frame O is a constant, then its velocity $\dot{\vec{r}}$ in the frame O' is also a constant as shown by Equation 2.18, and the body will obey the law of inertia as observed in the frame O'.

Differentiating Equation 2.18 once with time, we get

$$\ddot{\vec{r}} = \ddot{\vec{r}}. \tag{2.19}$$

Thus, the acceleration of the body does not change under Galilean transformations, and we say that it is a Galilean invariant.

The inertial mass m is unchanged under a Galilean transformation, so

$$m\,\ddot{\vec{r}}' = m\,\ddot{\vec{r}} \tag{2.20}$$

that is, the *ma* term of Newton's second law is invariant under Galilean transformations. If the force is velocity independent, then the force on the body P in the frame O' is the same as the force in the frame O, and the second is preserved:

$$\vec{F}' = m\ddot{\vec{r}}' = m\ddot{\vec{r}} = \vec{F}. \tag{2.21}$$

This is known as *Galilean invariance* or the principle of Newtonian relativity. Although the individual components may not be invariant, they transform according to the same scheme and are said to be covariant. Thus, two observers who are in uniform motion relative to one another observe the same laws of mechanics.

We have tacitly made a number of plausible assumptions in the above argument:

1. We have assumed that both observers use the same scale for measuring distance. This is indeed the case if there is no motion between the two systems. However, it is not generally true. We shall see in Chapter 13 on the theory of special relativity that length is contracted along the direction of motion.
2. Time is assumed to be the same in both systems. This assumption breaks down at high velocities.

   The reason these two results are so unexpected is that our notion of space and time come mainly from immediate contact with the world around us, and this does not involve velocities close to the velocity of light.
3. We also made an assumption that the observers agree on the value of the mass. This assumption must also be examined with the special theory of relativity.

Maxwell's equations, which form the basis of electrodynamics, are not invariant under Galilean transformations. To show this, it is sufficient to consider the consequence of Maxwell's equations: the constancy of the speed of light in all inertial frames. However, the Galilean transformation predicts that the speed of light should be different in two reference frames moving with constant velocity relative to each other. Let us look back at Figure 2.7. Suppose there is a light source emitting light waves at O, and P is a point on some given wave surface. In the frame O, the velocity of the point P is $\dot{\vec{r}} = c\hat{n}$, where $\vec{r}$ is the position vector of P, and $\hat{n}$ is the unit vector along $\vec{r}$. Then, according to Galilean transformation, the velocity of P in the frame O' is $\dot{\vec{r}} = c\hat{n} - \vec{v}$. Thus, in the frame moving with respect to the light source, the speed of light is, in general, no longer c; furthermore, because it depends on direction, the wave is also no longer spherical. Experiments carried out by Michelson-Morley and others have demonstrated that the velocity of light is the same in all directions and is independent of the relative uniform motion of the emitting source and the observer. This conflicts with the Galilean transformations. A number of attempts were made to resolve the conflict, and all failed. The solution was finally given by the theory of special relativity, in which the Galilean transformation was replaced by the Lorentz transformation.

2.5 NEWTON'S LAWS OF ROTATIONAL MOTION

We now consider how to apply Newton's laws to rotational motion. Torque and angular momentum are two important quantities in rotational motion. A force causes a linear acceleration; a torque causes an angular acceleration. Force is proportional to the time rate of change of momentum, and we shall see that torque is proportional to the time rate of change of angular momentum. To demonstrate this, we take the vector product of the position vector $\vec{r}$ of the particle by both sides of the second law:

$$\vec{r} \times (m\,\ddot{\vec{r}}) = \vec{r} \times \vec{F} \tag{2.22}$$

where

$$\vec{r} \times \vec{F} = \vec{N}_{\rm O} \tag{2.23}$$

is known as the torque (or moment of force) about the origin O (Figure 2.8), and $\vec{F}$ is the force acting on the particle. The subscript O of $\vec{N}_{\rm O}$ indicates that the torque is measured relative to the point O. The torque of the force $\vec{F}$ about point A, whose position vector relative to O is a, would be

$$\vec{N}_A = (\vec{r} - \vec{a}) \times \vec{F}.$$

This is different from $\vec{N}_{\rm O}$. The torque of force is also referred to as the moment of force.

Vector $\vec{N}_{\rm O}$ is perpendicular to the plane defined by $\vec{r}$ and $\vec{F}$ and points in the direction of the advance of a right-hand screw turned from $\vec{r}$ to $\vec{F}$. Its magnitude is $N_{\rm O}\, rF \sin\theta$, where $0 < \theta < \pi$ is the

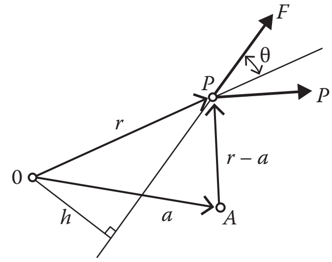


FIGURE 2.8 Torque (moment of force F).

angle between r and $\vec{F}$. As $h = r\sin\theta$, so $N_O = hF$, which is the product of $\vec{F}$ and the length of the perpendicular drawn from O to the line of action of the force $\vec{F}$ (it is often called the momentum arm). Now,

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\vec{r}\times\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right)=\vec{r}\times\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2}+\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\times\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}=\vec{r}\times\frac{\mathrm{d}^2\vec{r}}{\mathrm{d}t^2}.$$

Combining this result with Equations 2.22 and 2.23, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\vec{r}\times m\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right)=\vec{N}_O \tag{2.24}$$

but $\vec{r}\times m\vec{v}=\vec{r}\times\vec{p}$ is the angular momentum $\vec{L}_o$ of the particle about the origin O:

$$\vec{L}_o=\vec{r}\times m\vec{v}=\vec{r}\times\vec{p}.$$

Thus, Equation 2.24 can be put in a simple but interesting form:

$$\vec{N}_o=\frac{\mathrm{d}\vec{L}_o}{\mathrm{d}t}. \tag{2.25}$$

That is, the time rate of change of angular momentum about a fixed point O is equal to the torque about the same point O. It further follows that if the torque of the force $\vec{N}_O$ is zero, then the angular momentum $\vec{L}_O$ is constant.

The important concept of angular momentum will be found to be particularly useful in the study of central force motion as well as in the study of systems of particles and rigid bodies. We now apply the preceding result to a simple system: a particle of mass m moving on a circular path of radius r about the center O. In this case, we have

$$v=|\vec{v}|=r\dot{\theta}$$

and

$$L=|\vec{L}|=|\vec{r}\times\vec{p}|=mr^2\dot{\theta}=I\dot{\theta}$$

where $I = mr^2$ is the moment of inertia of m about the axis of rotation. Newton's second law for rotational motion now takes the simple form

$$N=I\frac{\mathrm{d}^2\theta}{\mathrm{d}t^2}=I\alpha,$$

which is a familiar result from basic physics.

2.6 WORK, ENERGY, AND CONSERVATION LAWS

In the preceding sections, we looked at the three laws of Newtonian mechanics. In principle, once the forces are known, every detail of the motion can be predicted by the use of Newton's second law of motion. Thus, the power of Newtonian mechanics lies in the possibility of finding force laws by which objects interact. For example, the force of air drag on an object can be obtained from a formula involving the size of the object and how fast it is moving; the force exerted by a spring depends on how much it is stretched or compressed. However, sometimes we are unable to discover in advance what forces will come into play when objects interact. Moreover, most problems in mechanics cannot be solved completely in terms of known force, or it is too tedious to solve them. In these circumstances, it is unusually helpful to identify some simple quantities, such as energy, momentum, and angular momentum, which are conserved (i.e., remain the same) during the entire motion. Very often, certain restricted information about a physical system that is of greater interest or importance than complete knowledge of a path can be obtained relatively easily from these conservation laws. The French philosopher Rene Descartes was the first person to make a good start on the construction of conservation laws. Newton built mechanics from the ground up with his three laws of motion, and Descartes tried to work from the top down on some general conservation principles of mechanics.

We first discuss the concepts of the work done by a force and of potential and kinetic energies. A discussion of work and energy naturally leads us to the concept of energy conservation. This and the other two conservation laws—the momentum and angular momentum conservation laws—have long been an important part of physics, and their importance will continue to increase because of their usefulness in many areas of physics.

2.6.1 Work and Energy

If a particle is subject to a force at any point in a region of space, it is said to be in a field of force in that region. Suppose the particle is initially at point A and is transported along some path to another point B. Now, the work done $\mathrm{d}W$ as the particle is traversing an element $\mathrm{d}\vec{r}$ along the path is $\mathrm{d}W = \vec{F}\cdot \mathrm{d}\vec{r}$. It is easy to show that the work done on the particle is expended in increasing its kinetic energy. We first take the scalar product of the second law with $\vec{v}\mathrm{d}t$:

$$m\dot{\vec{v}}\cdot\vec{v}\mathrm{d}t = \vec{F}\cdot\vec{v}\mathrm{d}t = \vec{F}\cdot\mathrm{d}\vec{r} = \mathrm{d}W.$$

The quantity $m\dot{\vec{v}}\cdot\vec{v}\mathrm{d}t$ can be simplified to a familiar form:

$$m\dot{\vec{v}}\cdot\vec{v}\mathrm{d}t = \frac{m}{2}\left[\frac{\mathrm{d}(\vec{v}\cdot\vec{v})}{\mathrm{d}t}\right]\mathrm{d}t = \mathrm{d}\left(\frac{1}{2}mv^2\right) = \mathrm{d}T.$$

Here, $T = mv^2/2$ is the kinetic energy of the particle, so we now have

$$\mathrm{d}W = \vec{F}\cdot\mathrm{d}\vec{r} = \mathrm{d}T. \tag{2.26}$$

This says that work done on the particle is equal to the increment in the kinetic energy. When the element of work done $\mathrm{d}W = \vec{F}\cdot\mathrm{d}\vec{r}$ is negative—as when the momentum of the particle carries it in a direction opposite that of the force—the particle works against the force, resulting in a reduction in its kinetic energy.

The work done by the force $\vec{F}$ in a finite displacement of the particle from point A to nearby point B on its path is given by the line integral

$$W = \int_A^B \mathrm{d}W = \int_A^B \vec{F}\cdot\mathrm{d}\vec{r} = \int_A^B \mathrm{d}T = T_B - T_A. \tag{2.27}$$

In terms of the components of the force along the coordinate axes, we have

$$\mathrm{d}W = \left(\sum_{j=1}^{3}\hat{e}_jF_j\right)\left(\sum_{k=1}^{3}\hat{e}_k\mathrm{d}x_k\right) = \sum_{k=1}^{3}F_j\mathrm{d}x_k\delta_{jk} = \sum_{j=1}^{3}F_j\mathrm{d}x_j = \mathrm{d}W_1 + \mathrm{d}W_2 + \mathrm{d}W_3$$

and

$$W = \int_A^B (F_1\,dx_1 + F_2\,dx_2 + F_3\,dx_3) = W_1 + W_2 + W_3.$$

Similarly, we can obtain the work done by a torque acting through a small angle dθ by considering a particle moving along a circular path of radius r. For simplicity, we assume the force F acting in a direction tangential to the path at all points of the path. Then

$$dW = Fr d\theta = N d\theta,$$

and

$$W = \int_{\theta_1}^{\theta_2} N d\theta.$$

2.6.2 Conservative Force and Potential Energy

In general, the work done by a force depends on the path along which the work is performed. As an example, consider the force

$$\vec{F} = (3x^2 + 6yz)\hat{e}_1 + (2y + 3xz)\hat{e}_2 + (1 + 5xyz^2)\hat{e}_3.$$

We shall see that the work done by this force along two different paths that join the points (0, 0, 0) and (1, 1, 1) depends on the path.

The work along the straight line $x = y = z$, which joins (0, 0, 0) and (1, 1, 1), is found to be

$$\begin{aligned} W &= \int_0^1 (3x^2 + 6yz)\,dx + \int_0^1 (2y + 3xz)\,dy + \int_0^1 (1 + 5xyz^2)\,dz \\ &= \int_0^1 \left\{9x^2 dx + (2y + 3y^2)dy + (1 + 5z^4)dz\right\} \\ &= 7 \text{ units of work.} \end{aligned}$$

On the other hand, the work along the path $x^2 = y = z$, which joins (0, 0, 0) and (1, 1, 1), is

$$\begin{aligned} W &= \int_0^1 (3x^2 + 6yz)\,dx + \int_0^1 (2y + 3xz)\,dy + \int_0^1 (1 + 5xyz^2)\,dz \\ &= \int_0^1 \left\{(3x^2 + 6x^4)dx + (2y + 3y^{3/2})dy + (1 + 5z^{7/2})dz\right\} \\ &= 6.4 \text{ units of work.} \end{aligned}$$

The work done along these paths is obviously not the same.

In dealing with a force whose work depends on the path, the work principle Equation 2.27 is of no use to us unless the path followed by the particle is known. However, if $\vec{F} \cdot d\vec{r}$ is an exact differential of a certain function of the integration variable r, say $V(r)$,

$$\vec{F} \cdot d\vec{r} = -dV(r). \tag{2.28}$$

The line integral for the work then becomes independent of the path of integration and depends only on the initial and final positions of the particle:

$$W = \int_A^B \vec{F} \cdot \mathrm{d}\vec{r} = -\int_A^B \mathrm{d}V = V_A - V_B \tag{2.29}$$

where V_A is the potential energy of the particle at point A, similarly to V_B.

Forces whose work is independent of the path are called conservative forces, and the motion under the action of such forces is known as conservative motion. Equation 2.28 is the necessary condition for a conservative force that is a function of position. It can be shown that it is also a sufficient condition. From Equation 2.28, we have

$$\vec{F} = -\text{grad } V(r) = -\nabla V(r) \tag{2.30}$$

where

$$\nabla = \sum_{i=1}^{3} \hat{e}_i \frac{\partial}{\partial x_i}.$$

That is, a conservative force can be expressed as a gradient of a certain scalar function $V(r)$. Because the curl of the gradient of any scalar function is zero, the curl of a conservative force vanishes:

$$\nabla \times \vec{F} = 0. \tag{2.30a}$$

Furthermore, in a conservative force field, the work integral depends only on the initial and final positions of the particle; thus, the work integral around any closed path is zero:

$$\oint_s \vec{F} \cdot \mathrm{d}\vec{s} = 0 \text{ for a closed path } s. \tag{2.30b}$$

Equation 2.30b also follows from the disappearance of the curl of $\boldsymbol{F}$ and Stokes' theorem.

2.6.3 Conservation of Energy

Combining Equation 2.29 with Equation 2.27, we obtain

$$T_B - T_A = V_A - V_B$$

or

$$T_B + V_B = T_A + V_A. \tag{2.31}$$

That is, the total energy of the particle, which is defined as the sum of the kinetic and potential energies, is conserved (i.e., kept constant in time) in a conservative force field. This is known as the principle of energy conservation.

It should be noted that the potential energy $V(r)$ is defined only to within an additive constant. This is related to the arbitrariness of the choice of the initial point from which the work done is measured. It is usual to choose the arbitrary constant so that the potential energy of the particle vanishes at an infinite distance from other bodies.

Example 2.3: Gravitational Escape

As an illustration of energy-conservation methods, we consider the problem of the gravitational escape of a particle from the Earth. The gravitational potential energy resulting from the Earth's attraction on a particle of mass m at a distance $r > r_e$ (Earth's radius) from the Earth's center is

$$V(r) = -\int_{\infty}^{r} \frac{-GmM}{r^2} \mathrm{d}r = -\frac{GmM}{r}$$

where M is the mass of the Earth, and we have chosen the convention that the potential energy vanishes at infinite distance, so by the principle of energy conservation, we have

$$\frac{1}{2}mv^2 + \frac{-GmM}{r} = E = \text{constant}.$$

Evaluating the constant at the initial point, we see that

$$\frac{1}{2}mv^2 + \frac{-GmM}{r} = \frac{1}{2}mv_0^2 + \frac{-GmM}{r_e}$$

from which it follows that

$$v^2 = 2GM\left(\frac{1}{r} - \frac{1}{r_e}\right) + v_0^2.$$

If the particle just reaches infinity (i.e., $r \to \infty$) as $v \to 0$, we see that $V_0^2 = 2GM/r_e$. Thus, if $V_0^2 \geq 2GM/r_e$, the particle will escape to infinity. The critical speed

$$v_0 = \sqrt{2GM/r_e}$$

is called the escape velocity. The minimum escape velocity from the Earth's surface is 11.2 km/s.

Example 2.4: Simple Harmonic Motion

Simple harmonic motion is an oscillatory motion that occurs whenever a force acts on a body in the opposite direction to its displacement from its equilibrium position, with the magnitude of the force proportional to the magnitude of the displacement. Thus, the force on a body in a simple harmonic motion always tends to return to its equilibrium position. This restoring force is of the form $F = -kx$, where k is the spring constant. The equation of motion is

$$m\ddot{x} = -kx$$

and the potential energy function is

$$V(x) = -\int_0^x F\,\mathrm{d}x = \int_0^x kx\,\mathrm{d}x = \frac{1}{2}kx^2.$$

If we take $x = x_0$ at $t = 0$ and $\dot{x} = 0$ at the maximum displacement A as our boundary conditions, then the energy equation, $E = T + V$, becomes

$$\frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 = 0 + \frac{1}{2}kA^2.$$

Solving for $\dot{x}$, we have

$$\dot{x} = \sqrt{\frac{k}{m}(A^2 - x^2)}$$

from which we obtain

$$\frac{\mathrm{d}x}{\sqrt{\frac{k}{m}(A^2 - x^2)}} = \mathrm{d}t.$$

Integration gives

$$\sin^{-1}\left(\frac{x}{A}\right) = \sqrt{\frac{k}{m}} + C,$$

where C is the integration constant. Because $x = x_0$ at $t = 0$, we find

$$C = \sin^{-1}\left(\frac{x_0}{A}\right) \equiv -\beta.$$

And, finally, we obtain

$$x = A\sin\left[\sqrt{\frac{k}{m}}t - \beta\right].$$

Small oscillations in a complicated system can often be approximately treated in terms of simple harmonic motion. More detailed treatment of linear oscillations is given in Chapter 8.

2.6.4 Conservation of Momentum

Multiplying Equation 2.4 by dt and integrating from t_1 to t_2, we obtain

$$\vec{P}_2 - \vec{P}_1 = m\vec{v}_2 - m\vec{v}_1 = \int_{t_1}^{t_2} \vec{F}\,\mathrm{d}t \tag{2.32}$$

in which $\vec{p}_2$ is the momentum of the particle at the end of the interval, at which time the force ceases to act. The integral on the right is the impulse delivered by the force $\vec{F}$. It is the momentum imparted to the system (a particle in this case) by the force during the time interval in which the force acts. Equation 2.32 is very useful when the time interval $\Delta t(= t_2 - t_1)$ is so small that the system does not move appreciably while the force is acting. Thus, in the present case, the particle has its momentum instantaneously increased from $\vec{p}_1$ to $\vec{p}_2$. A force that acts in this manner for a very short time duration is called an impulsive force.

When the impulse is zero, Equation 2.32 reduces to $\vec{p}_2 = \vec{p}_1$, or

$$\vec{p} = \text{constant (a constant vector in time)}.$$

That is, the linear momentum of the particle is conserved.

2.6.5 Conservation of Angular Momentum

The quantity, which in angular motion corresponds to the linear impulse, is the angular impulse. From Equation 2.25, we have

$$\vec{L}_2 - \vec{L}_1 = \int_{t_1}^{t_2} \vec{N}\mathrm{d}t = \vec{r} \times \int_{t_1}^{t_2} \vec{F}\mathrm{d}t. \tag{2.33}$$

The quantity on the right is the moment of the linear impulse and is called the angular impulse. When the angular impulse is zero, the angular momentum is conserved:

$$\vec{L} = \text{constant (a constant vector in time).}$$

As an example, let us consider the motion of a particle in an inverse square force field $\vec{F} = k\hat{r}/r^2$, where k is a constant and $\hat{r}$ is a unit vector in the direction of $\vec{r}$. The torque on the particle is zero:

$$\vec{N} = \vec{r} \times \vec{F} = \frac{k}{r^2}(\vec{r} \times \vec{r}) = 0,$$

and so its angular momentum is conserved. That is, the angular momentum of a particle in an inverse square force field is a constant of motion in time. Further, the total energy of the particle is also conserved:

$$\vec{F} = -\nabla V(r) = \frac{k}{r^2}\hat{r}$$

from which we get

$$V(r) = \frac{k}{r},$$

that is, the inverse square force is a conservative force. Thus, the total energy of a particle is conserved. However, because the force (and therefore the impulse) is not zero, the linear momentum is not conserved.

Example 2.5

Two particles, P and Q, each of mass m, are connected by a light, inextensible string of length $2l$, which passes through a small smooth hole O in a smooth horizontal table. Particle P is free to slide on the table, and Q hangs freely below it. Initially OQ is of length l, and P is projected from rest at right angles to OP with velocity $(8gl/3)^{1/2}$. Show that in the ensuing motion, particle Q will just reach the table.

Solution:

We can solve this problem by making use of the conservation of energy and conservation of angular momentum. Let (r, θ) be the polar coordinates of P referred to some initial line through O in the table (Figure 2.9). Particle P has radial and transverse velocity components $(\dot{r}, r\dot{\theta})$. The distance of Q below the table is $2l - r$, and its velocity is $\mathrm{d}(2l - r)/\mathrm{d}t = -\dot{r}$ (downward). Hence, the kinetic energy of the system is

$$\frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) + \frac{1}{2}m(-\dot{r})^2 = m\left(\dot{r}^2 + \frac{1}{2}r^2\dot{\theta}^2\right).$$

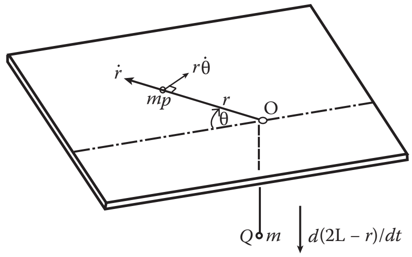


FIGURE 2.9 Example of energy conservation principle.

The potential energy is referring to the table as zero level $-mg(2l-r)$. The principle of energy conservation gives

$$m\left(\dot{r}^2+\frac{1}{2}r^2\dot{\theta}^2\right)-mg(2l-r)=C \text{ (constant)}$$

or

$$\dot{r}^2+\frac{1}{2}r^2\dot{\theta}^2+gr=K, \quad (K=C/m+2gl).$$

Applying the initial conditions, $\dot{r}=0$, $r+l=2l$ (or $r=l$), and $r\dot{\theta}=(8gl/3)^{1/2}$, we find $K=7gl/3$. Hence, we have

$$\dot{r}^2+\frac{1}{2}r^2\dot{\theta}^2+gr=\frac{7gl}{3}. \tag{2.34}$$

Further, the initial angular momentum is $L_0=ml(8gl/3)^{1/2}$, and in general, we have

$$|\vec{L}|=|\vec{r}\times\vec{p}|=|r\hat{e}_r\times m(\dot{r}\hat{e}_r+r\dot{\theta}\,\hat{e}_\theta)|=mr^2\dot{\theta}.$$

The angular momentum conservation then gives

$$mr^2\dot{\theta}=ml(8gl/3)^{1/2}. \tag{2.35}$$

Elimination of $\dot{\theta}$ from Equations 2.34 and 2.35 gives

$$\dot{r}^2=\frac{7gl}{3}-\frac{4gl^3}{3r^2}-gr$$

from which $\dot{r}=0$ when

$$3r^3-7r^2l+4l^3=0$$

or

$$(r-l)\,(3r+2l)\,(r-2l)=0.$$

Thus, the system comes to instantaneous rest at $r=l$ or $r=2l$. When $r=2l$, particle Q just reaches the hole.

Example 2.6

A charged particle $+e_1$ moving with a very high velocity v_0 along a straight line passes another charge $+e_2$ of mass m at a distance b. Assuming a central law of force of magnitude e_1e_2/r^2 between e_1 and e_2, find the energy Q transferred from e_1 to e_2 during the encounter.

Solution:

Because the charged particle e_1 moves at a very high speed, we make the assumption that e_2 does not have time to change position during the encounter, so we place e_2 on the x-axis at a distance b from the origin (Figure 2.10). Let us now consider the impulse. From symmetry, we see that the y component is zero, and the x-component is given by

$$\mathrm{d}P_x = F_x\mathrm{d}t = -\frac{e_1e_2}{r^2}\cos\theta\,\mathrm{d}t = -\frac{e_1e_2}{r^2}\cos\theta\frac{\mathrm{d}t}{\mathrm{d}\theta}d\theta.$$

But $r\dot{\theta} = -v_0\cos\theta$, where $r = b/\cos(\pi - \theta) = -b/\cos\theta$, so the last equation becomes

$$\mathrm{d}P_x = -\frac{e_1e_2}{bv_0}\cos\theta\,\mathrm{d}\theta.$$

Integrating from $\theta = \pi/2$ to $\theta = 3\pi/2$ ($y = \infty$ to $y = -\infty$)

$$P_x = -\frac{e_1e_2}{bv_0}\int_{\pi/2}^{3\pi/2}\cos\theta\,\mathrm{d}\theta = \frac{2e_1e_2}{bv_0}.$$

This the momentum $m\dot{x}$ imparted to e_2; the energy Q transferred from e_1 to e_2 is

$$Q = \frac{1}{2}m\dot{x}^2 = \frac{2e_1^2e_2^2}{mb^2v_0}.$$

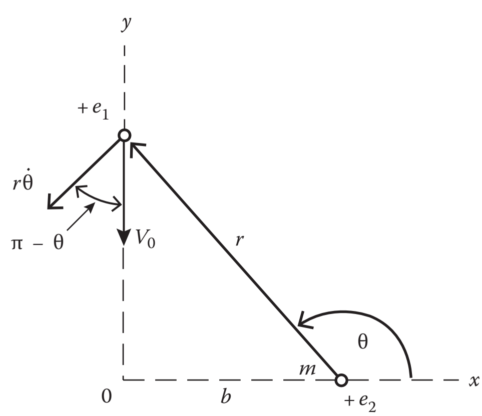


FIGURE 2.10 Two interacting charged particles.

Example 2.7

A rubber ball of mass 0.2 kg hits the floor with a speed of 8 m/s and rebounds with approximately the same speed. If the ball is in contact with the floor for 10^{-3} s (high-speed photographs can show this), what is the force exerted on the ball by the floor?

Solution:

The momentum of the ball just before it hits the floor is

$$P_1 = 0.2 \text{ kg} \times 8 \text{ m/s} = -1.6 \text{ kg m/s}.$$

The minus sign here indicates that P_1 is downward. The momentum of the rebound ball is $P_2 = +1.6$ kg·m/s. Thus, the impulse is

$$\int_{t_1}^{t_2} \vec{F}\,\mathrm{d}t = P_2 - P_1 = 3.2 \text{ kg}\cdot\text{m/s}.$$

The exact variation of F with time is not known, but we can find the average force exerted by the floor on the ball. If the collision is $\Delta t = t_2 - t_1$, the average force F_{av} acting during the collision is

$$\vec{F}_{av}\Delta t = \int_{t_1}^{t_1+\Delta t} \vec{F}\,\mathrm{d}t.$$

Now $\Delta t = 10^{-3}$ s, so $F_{av} = 3.2 \text{ kg} \cdot \text{m/s}/10^{-3} \text{ s} = 3200$ N.

The gravitational effect has been neglected. This is justified. The impulse resulting from the gravitational force is

$$-\int_0^{10^{-3}} Mg\,\mathrm{d}t = -1.9\times 10^{-3}\text{kg}\cdot\text{m/s},$$

where the minus sign indicates that the gravitational force is downward. It is less than one-thousandth of the total impulse, 3.2 kg · m/s. Thus, we can neglect the gravitational effect without encountering appreciable error. However, over a long period of time, gravity can produce a large change in the ball's momentum.

2.7 SYSTEMS OF PARTICLES

We now expand the discussion to include a system of interacting particles, assuming that the individual masses are constant and that the laws of motion remain valid. We shall find that Newton's third law plays a very important role when more than one particle is involved. For our study of the motion of a system of interacting particles, we can place the origin of the coordinate system either at a fixed point or at the center of mass (CM) of the system. Very often, it is more convenient to place the origin at the CM. It is, therefore, worthwhile to make the definition of the CM more general than was stated in basic physics.

2.7.1 Center of Mass

To start with, consider a system of two particles that can be thought of as point particles and label them 1 and 2. The total momentum $\vec{P}$ of the system is the sum of the momenta of the two constituent particles:

$$\vec{P} = \vec{P}_1 + \vec{P}_2 = m_1\vec{v}_1 + m_2\vec{v}_2.$$

The mass of the system is $m = m_1 + m_2$.

It is then reasonable to define the velocity $\vec{v}$ of the two-particle system so that the momentum of the system is its mass times its velocity. Therefore,

$$\vec{P} = m_1\vec{v}_1 + m_2\vec{v}_2 = m\vec{v} = (m_1 + m_2)\vec{v}$$

and solving for v, we find that

$$\vec{v} = (m_1\vec{v}_1 + m_2\vec{v}_2)/(m_1 + m_2). \tag{2.36}$$

The velocity $\vec{v}$ is called the velocity of the CM of the pair of particles. Equation 2.36 suggests the most convenient definition for the position of the CM of the two-particle system. Denoting the position vector of this position by $\vec{R}$, then the choice

$$\vec{R} = \frac{m_1\vec{r}_1 + m_2\vec{r}_2}{m_1 + m_2} \tag{2.37}$$

leads directly to Equation 2.36. The CM of a pair of particles lies on the line joining them and nearer the more massive one.

The definition of the CM is easily extended to any number of particles (Figure 2.11). It is given by

$$\vec{R} = \frac{\sum_{i=1}^{n} m_i\vec{r}_i}{\sum_{i=1}^{n} m_i} = \frac{\sum_{i=1}^{n} m_i\vec{r}_i}{M}. \tag{2.38}$$

In elementary physics, the CM of a body is defined as a point of balance. (Note that this is true only in a uniform gravitational field.) We now consider a simple example of two particles connected by a massless rod that can rotate about a point C (Figure 2.12a). If C is a point of balancing, we have

$$m_1 a = m_2 b.$$

We now introduce a coordinate system, the x-axis (Figure 2.12b), to describe these two particles:

$$a = x_{CM} - x_1$$

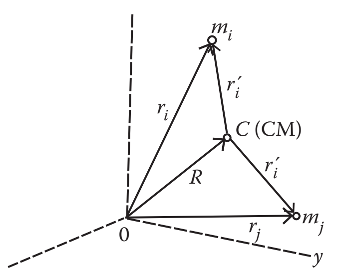


FIGURE 2.11 Position vectors of a particle system.

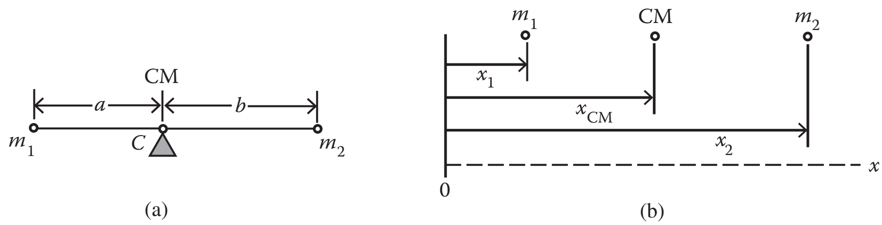


FIGURE 2.12 Center of mass of two particles. (a) C is a balancing point, (b) coordinate system of the two particles.

and

$$b = x_2 - x_{CM}.$$

Accordingly, we have

$$m_1(x_{CM} - x_1) = m_2(x_2 - x_{CM}).$$

Solving for x_{CM}

$$x_{CM} = \frac{m_1x_1 + m_2x_2}{m_1 + m_2}$$

which is a special case of Equation 2.38.

2.7.2 Motion of CM

The forces that act on the particles of the system can be divided into two classes: (1) the external force $\vec{F}_j^{(e)}$ whose origin lies outside the system and (2) the internal force, which originates within the system and describes the interactions among the particles. If $\vec{F}_{jk}$ denotes the internal force on the jth particle from the kth particle, the total force on the jth particle $\vec{F}_{jt}$ can be written as

$$\vec{F}_{jt} = \vec{F}_j^{(e)} + \sum_{k \neq j}^{n} \vec{F}_{jk}, j = 1,2,\ldots n$$

and the equation of motion of the jth particle is

$$m_j\ddot{\vec{r}}_j = \vec{F}_j^{(e)} + \sum_{k \neq j}^{n} \vec{F}_{jk}, j = 1,2,\ldots n \tag{2.39}$$

where m_j is the mass of the jth particle, and r_j is its radius vector relative to a fixed origin. There is a total of n equations, such as Equation 2.39, one for each constituent particle.

Summing Equation 2.39 over index j, we obtain

$$\sum_{j=1}^{n} m_j\ddot{\vec{r}}_j = \sum_{j=1}^{n} \vec{F}_j^{(e)} + \sum_{j,k=1}^{n} \vec{F}_{jk} (j \neq k). \tag{2.40}$$

If the internal forces obey Newton's third law, that is, $\vec{F}_{jk} = -\vec{F}_{kj}$, then

$$\sum_{j,k=1}^{n} \vec{F}_{jk} = \frac{1}{2}\sum_{j,k=1}^{n} (\vec{F}_{jk} + \vec{F}_{kj}) = 0, \quad (k \neq j)$$

and Equation 2.40 becomes

$$\sum_{j=1}^{n} m_j \ddot{\vec{r}}_j = \sum_{j=1}^{n} \vec{F}_j^{(e)} = \vec{F}^{(e)} \tag{2.41}$$

where $\vec{F}^{(e)}$ is the total external force acting on the whole system.

Equation 2.41 can be put into a simple but interesting form in the CM system. The position vector $\vec{R}$ of the CM system is given by Equation 2.38. If we differentiate Equation 2.38 twice, we find

$$M\ddot{\vec{R}} = \sum_{j=1}^{n} m_j \ddot{\vec{r}}_j.$$

Combining this with Equation 2.41, we have

$$M\ddot{\vec{R}} = \vec{F}^{(e)}. \tag{2.42}$$

This is a very important and useful result, which states that the CM of the system moves as if it were a single particle of mass M (the total mass of the system) acted upon by the external force, as long as the internal forces follow Newton's third law.

If $\vec{F}^{(e)} = 0$, then $\dot{\vec{R}} = \text{constant}$, that is, the CM moves with constant velocity, and so the CM system is an inertial frame.

2.7.3 Conservation Theorems

We have seen that there are three important conservation laws for a single particle: energy, momentum, and angular momentum. These laws are still valid for a system composed of n particles as long as the internal forces follow Newton's third law.

The momentum $\vec{P}$ of the system is the vector sum of the momentum of the constituent particles. Thus,

$$\vec{P} = \sum_{j=1}^{n} m_j \dot{\vec{r}}_j.$$

If the mass m_j is constant, and with the help of Equation 2.38, we can write $\vec{P}$ as

$$\vec{P} = \sum_{j=1}^{n} m_j \dot{\vec{r}}_j = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{j=1}^{n} m_j \vec{r}_j\right) = \frac{\mathrm{d}}{\mathrm{d}t}(M\vec{R}) = M\dot{\vec{R}}. \tag{2.43}$$

That is, the momentum $\vec{P}$ of a system composed of n particles is the same as if a single particle of mass M were located at the position of the CM and moving in the manner in which the CM moves (i.e., moving with $\dot{\vec{R}}$). It is thus sufficient to know the velocity of the CM of the system and the mass of all particles to be able to obtain the resultant or total momentum of the system.

Differentiating Equation 2.43 with respect to time, we find the second law of motion for the system as a whole:

$$\dot{\vec{P}} = M\ddot{\vec{R}} = \vec{F}^{(e)}. \tag{2.44}$$

That is, the total momentum $\vec{P}$ of a system of particles changes only as a result of the action of external forces. Hence, if $\vec{F}^{(e)} = 0$, we get the momentum conservation law for a system of particles:

$$\vec{P} = \text{a constant vector} \tag{2.45}$$

Because the velocity of the CM of a closed system of particles is constant in time, the CM frame is an inertial frame. The description of phenomena in this frame of reference eliminates complications arising from the motion of the system as a whole and demonstrates more clearly the properties of the internal processes occurring within the system. For this reason, the CM frame is frequently used in mechanics.

Our next concern is the rotational properties of such a system of particles. The angular momentum of the jth particle relative to the origin is

$$\vec{L}_j = \vec{r}_j \times \vec{p}_j.$$

The total angular momentum of the system $\boldsymbol{L}$ is the vector sum of the angular momentum of the individual particles:

$$\vec{L} = \sum_{j=1}^{n} \vec{L}_j = \sum_{j=1}^{n} \left(\vec{r}_j \times \vec{p}_j\right) = \sum_{j=1}^{n} \left(\vec{r}_j \times m_j \dot{\vec{r}}_j\right). \tag{2.46}$$

Now, as shown in Figure 2.11,

$$\vec{r}_j = \vec{R} + \vec{r}_j^{\,\prime}.$$

Substituting this into Equation 2.46, we have

$$\begin{aligned}\vec{L} &= \sum_{j=1}^{n} \left[\left(\vec{r}_j^{\,\prime} + \vec{R}\right) \times m_j \left(\dot{\vec{r}}_j^{\,\prime} + \dot{\vec{R}}\right)\right.\\ &= \sum_{j=1}^{n} m_j \left[\left(\vec{R} \times \dot{\vec{R}}\right) + \left(\vec{R} \times \dot{\vec{r}}_j^{\,\prime}\right) + \left(\vec{r}_j^{\,\prime} \times \dot{\vec{R}}\right) + \left(\vec{r}_j^{\,\prime} \times \dot{\vec{r}}_j^{\,\prime}\right)\right].\end{aligned}$$

We can write the second and third terms on the right as

$$\left(\sum_{j=1}^{n} m_j \vec{r}_j^{\,\prime}\right) \times \dot{\vec{R}} + \vec{R} \times \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{j=1}^{n} m_j \vec{r}_j^{\,\prime}\right) = 0$$

where we have made use of the fact that $\sum_j m_j \vec{r}_j^{\,\prime}$ is a null vector because it specifies the position of the CM in the CM frame:

$$\sum_j m_j \vec{r}_j^{\,\prime} = \sum_j m_j (\vec{r}_j - \vec{R}) = \sum_j m_j \vec{r}_j - \sum_j m_j \vec{R} = M\vec{R} - M\vec{R} = 0.$$

The total angular momentum $\boldsymbol{L}$ of the system now takes a simple form:

$$\vec{L} = \vec{R} \times M\dot{\vec{R}} + \sum_{j=1}^{n} \vec{r}_j' \times m_j \dot{\vec{r}}_j' = \vec{L}_{CM} + \vec{L}' \tag{2.47}$$

where the last term is the vector sum of the angular momentum of the particles about the CM. Hence, we find that the angular momentum of a system of particles about a fixed origin O is equal to that of a single particle whose mass is the entire mass M of the system located at and moving with the CM, $\vec{L}_{\text{CM}}$, plus the angular momentum of the particle system about the CM, $\vec{L}'$.

We now consider the conservation law for angular momentum. Differentiating $\vec{L}_j$ with respect to time, we obtain

$$\frac{d\vec{L}_j}{dt} = \vec{r}_j \times \dot{\vec{p}}_j + \dot{\vec{r}}_j \times \vec{p}_j,$$

where the second term on the right vanishes, and the factor $\dot{\vec{p}}_j$ in the first term can be written as $\vec{F}_j^{(e)} + \sum_j \vec{F}_{jk}$. Thus, we have

$$\frac{\mathrm{d}\vec{L}_j}{\mathrm{d}t} = \vec{r}_j \times \dot{\vec{p}}_j = \vec{r}_j \times \left[\vec{F}^{(e)} + \sum_{j=1}^{n} \vec{F}_{jk} \right], \; (j \neq k)$$

and after summing over index j

$$\begin{aligned}
\dot{\vec{L}} = \sum_{j=1}^{n} \dot{\vec{L}}_j &= \sum_{j=1}^{n} \left(\vec{r}_j \times \vec{F}_j^{(e)} \right) + \sum_{j,k=1}^{n} \left(\vec{r}_j \times \vec{F}_{jk} \right) (j \neq k) \\
&= \sum_{j=1}^{n} \left(\vec{r}_j \times \vec{F}_j^{(e)} \right) + \sum_{j<k}^{n} \left(\vec{r}_j \times \vec{F}_{jk} \right) + \sum_{j>k}^{n} \left(\vec{r}_j \times \vec{F}_{jk} \right) \\
&= \sum_{j=1}^{n} \left(\vec{r}_j \times \vec{F}_j^{(e)} \right) + \sum_{j<k}^{n} \left[\left(\vec{r}_j \times \vec{F}_{jk} \right) + \left(\vec{r}_k \times \vec{F}_{kj} \right) \right] \\
&= \sum_{j=1}^{n} \left(\vec{r}_j \times \vec{F}_j^{(e)} \right) + \sum_{j<k}^{n} \left(\vec{r}_{jk} \times \vec{F}_{jk} \right)
\end{aligned}$$

where $\vec{r}_{jk} = \vec{r}_j - \vec{r}_k$, and it is parallel to $\vec{F}_{jk}$ (Figure 2.13). Thus, the last term on the right vanishes, and we finally have

$$\frac{\mathrm{d}\vec{L}}{\mathrm{d}t} = \sum_{j=1}^{n} \left(\vec{r}_j \times \vec{F}_j^{(e)} \right) = \sum_{j=1}^{n} \vec{N}_j^{(e)} = \vec{N}^{(e)}. \tag{2.48}$$

If the external torque $N^{(e)}$ about a given point is zero, then the total angular momentum of the system about the same point is a constant vector in time. Thus, we see that the law of conservation of angular momentum is valid for a system of particles. It is important to note that because of the vector nature, Equation 2.48 holds independently for each component.

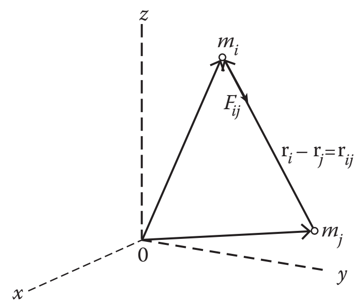


FIGURE 2.13 Two interacting particles.

It is often preferable to transfer to an internal coordinate system located at the moving CM. Thus, we now differentiate both sides of Equation 2.47 with respect to time and obtain

$$\frac{d\vec{L}}{dt} = \frac{d\vec{L}_{CM}}{dt} + \frac{d\vec{L}'}{dt}$$

but

$$\frac{d\vec{L}_{CM}}{dt} = \frac{d}{dt}\left(\vec{R}\times M\dot{\vec{R}}\right) = \vec{R}\times M\ddot{\vec{R}} = \vec{R}\times\vec{F}^{(e)}.$$

Combining this with the preceding relationship, we obtain

$$\begin{aligned}\frac{d\vec{L}'}{dt} = \frac{d\vec{L}}{dt} - \frac{d\vec{L}_{CM}}{dt} &= \sum_{j=1}^{n}\vec{r}_j\times\vec{F}^{(e)} - \vec{R}\times\sum_{j=1}^{n}\vec{F}^{(e)} \\ &= \sum_{j=1}^{n}\left(\vec{r}_j - \vec{R}\right)\times\vec{F}_j^{(e)} = \vec{N}'\end{aligned} \tag{2.49}$$

where $\vec{N}'$ is the torque about the CM. Thus, for motion relative to the CM, the time rate of change of the angular momentum equals the external torque about the CM. This is true even if the CM has a translational motion relative to a fixed-coordinate system. This is a very important result. It means that when studying the rotational motion of a system of particles or a rigid body, we may ignore the translational motion of the CM of the system provided we refer all torques and angular momenta to the CM as the origin.

Similarly, we can show that the kinetic energy of a system of particles is equal to the kinetic energy of a single particle of mass M moving with the CM plus the kinetic energy of the system of particles relative to the CM:

$$T = \frac{1}{2}MV^2 + \sum_{j=1}^{n}\frac{1}{2}m_j\left(v_j'\right)^2, \tag{2.50}$$

and the potential energy of the system is

$$V = \sum_{j=1}^{n}V_j + \sum_{j<k}^{n}V_{jk}. \tag{2.51}$$

The law of energy conservation is also valid:

$$T_2 + V_2 = T_1 + V_1 = E \text{ (constant).} \tag{2.52}$$

The internal potential energy can be neglected in the discussion of the motion of a rigid body because V_{jk} is always constant for rigid bodies. A rigid body is one for which the distances between the constituent particles are constant; that is, $r_{jk}^2 = \vec{r}_{jk} \cdot \vec{r}_{jk} = \text{constant}$, so $\vec{r}_{jk} \cdot \mathrm{d}\vec{r}_{jk} = 0$, and $\vec{r}_{jk} \perp \mathrm{d}\vec{r}_{jk}$. Because the internal forces $\vec{F}_{jk}$ are parallel to $\vec{r}_{jk}$, the $\vec{F}_{jk}$ is perpendicular to $\mathrm{d}\vec{r}_{jk}$, and $\vec{F}_{jk} \cdot d\vec{r}_{jk} = 0$. That is, the internal forces do not work, so the internal potential energy is always constant. We shall see later that the concept of "rigid" is modified in the special theory of relativity.

Example 2.8

A cannon shell of mass M moves along a parabolic trajectory. An internal explosion, generating an amount of energy E, blows the shell into two parts. One part of mass kM with $k < 1$ continues in the original direction, and the other part is reduced to rest. Find the velocity of the mass kM immediately after the explosion.

Solution:

It is important to notice that the energy E is available only for the motion relative to the CM. Immediately after the explosion, the CM of the two parts continues along the trajectory with the same velocity that it had originally. Now, if v is the speed of the shell just before the explosion and v' is the speed of kM immediately after the explosion, then we have

$$\frac{1}{2}Mv^2 + E = \frac{1}{2}kMv'^2 \text{ (energy conservation)}$$

and

$$Mv = kMv' \text{ (momentum conservation).}$$

Eliminating v between these two equations, we obtain

$$v' = \sqrt{2E/\{k(1-k)M\}}.$$

Example 2.9

Two masses, m_1 and m_2, are connected by a linear spring of constant k, and the system is initially at rest on a frictionless surface (Figure 2.14a). The spring is compressed an amount δ and released. Find the velocity of each mass when the spring returns to its natural non-deformed length L.

Solution:

Figure 2.14b shows the system of particles after the spring is released. The internal forces (spring forces F_s) cancel in pairs, and the sum of the external forces (weights and normal reactions) is zero. Hence, the impulse is zero, and the momentum is conserved. At any time, the sum of the

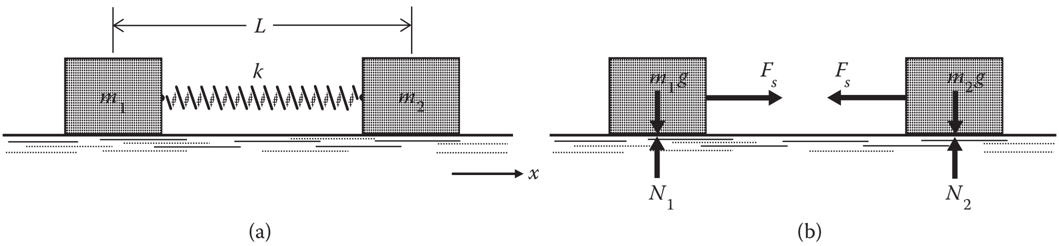


FIGURE 2.14 (a) Two masses connected by a linear spring, (b) force diagram of the two-mass system.

momentum must be equal to a constant, which, in this case, is zero (because the initial system is at rest):

$$m_1\dot{x}_1 + m_2\dot{x}_2 = 0.$$

When there is no friction, energy is also conserved. Thus, we have

$$E = \frac{1}{2}k\delta^2 = \frac{1}{2}m_1\dot{x}_1^2 + \frac{1}{2}m_2\dot{x}_2^2$$

where $\dot{x}_1$ and $\dot{x}_2$ represent the velocities of the non-deformed state of the spring. Solving the above two equations, we obtain

$$\dot{x}_2 = \sqrt{\frac{k\delta^2}{(m_1m_2 + m_2 2)/m_1}}, \quad \dot{x}_1 = \sqrt{\frac{k\delta^2}{(m_1m_2 + m_1 2)/m_2}}.$$

PROBLEMS

1. A particle of mass m moves in the XY-plane, and its position vector is given by

$$\vec{r} = a\cos(\omega t)\hat{i} + b\sin(\omega t)\hat{j}$$

where a, b, and ω are positive constants, and $a > b$. Show that
(a) The particle moves in an ellipse.
(b) The force acting on the particle is always directed toward the origin.
(c) The total work done by the force in moving the particle once around the ellipse is zero.
(d) The force is conservative.

2. A constant force $\vec{F}$ acting on a particle of mass m changes the velocity from $\vec{v}_1$ to $\vec{v}_2$ in time τ. (a) Prove that $\vec{F} = m(\vec{v}_2 - \vec{v}_1)/\pi$. (b) Does the result in (a) hold if the $\vec{v}_2$ force is variable? Explain.

3. Find the work done in moving an object along a path given by

$$\vec{r} = 3\hat{i} + 2\hat{j} - 5\hat{k}$$

if the applied force is

$$\vec{F} = 2\hat{i} - \hat{j} - \hat{k}.$$

4. (a) Show that

$$\bar{F} = (2xy + z^3)\hat{i} + x^2\hat{j} + 3xz^2\hat{k}$$

is a conservative force field.

(b) Find the potential energy V.

(c) Find the work done in moving an object in this force field from (1, –2, 1) to (3, 1, 4).

5. A particle of mass m moves along the x-axis under the influence of a conservative force field having potential $V(x)$. If the particle is located at positions x_1 and x_2 at respective times t_1 and t_2, prove that if E is the total energy,

$$t_2 - t_1 = \sqrt{m/2}\int_{x_1}^{x_2} \frac{\mathrm{d}x}{\sqrt{E - V(x)}}.$$

6. A particle moves in a force field given by $\vec{F} = r^2\vec{r}$, where $\vec{r}$ is the position vector of the particle. Show that the angular momentum of the particle is conserved.
7. A particle of mass m moves along the path given by
$x = x_0 + at^2 \quad y = bt^3 \quad z = ct$
where x_0, a, b, and c are constants. Find the following quantities at any later time t: angular momentum $\vec{L}$, force $\vec{F}$, and torque $\vec{N}$ on the particle. Verify that they satisfy

$$\mathrm{d}\vec{L}/\mathrm{d}t = \vec{r} \times \vec{F} = \vec{N}.$$

8. Two astronauts, A and B, initially at rest in free space, pull on either end of a rope. The maximum force with which A can pull, F_A, is larger than the maximum force with which B can pull, F_B. Their masses are M_A and M_B. The mass of the rope M_r is negligible. Find their motion if each pulls on his rope end as hard as he can.
9. A block of mass M, resting on a smooth table, is pulled by a string of mass m. If a force $\vec{F}$ is applied to the string, what is the force that the string transmits to the block?
10. Two particles of masses m_1 and m_2 are connected by a rigid rod lying on a smooth horizontal table. If an impulse I is applied at m_1 in the plane of the table and perpendicular to the rod, find the initial velocities of m_1 and m_2.
11. A wheel of radius b is rolling along a muddy road with a speed v_0. Particles of mud attached to the wheel are being continuously thrown off from all points of the wheel. If $v_0^2 > bg$, where g is the gravitational acceleration, show that the maximum height above the road attained by the mud will be

$$b + \frac{v_0^2}{2g} + \frac{b^2 g}{2v_0^2}.$$

12. A particle of mass m_1 is free to slide on the inclined face of a smooth wedge of mass m_2 and angle α. The wedge is itself free to slide on a smooth horizontal plane. Find the acceleration of both the particle and the wedge.
13. A wooden block of mass M is resting on a horizontal surface. The coefficient of friction is f. One end of a spring with a spring constant k is attached to the block; the other end is attached to a solid wall. The spring is unstretched. A bullet of mass m hits the block and embeds in it. Find the velocity of the bullet before impact in terms of the maximum compression x of the spring and M, k, g, and f.
14. Two massless springs S_1 and S_2 with spring constants k_1 and k_2, respectively, are arranged to support a weight A. In case I, the springs are coupled in series, and in case II, they are in parallel. Determine the extensions of the individual springs in these two cases as a result of the force of gravity on A. Also determine the equivalent spring constant in the two cases.
15. Consider a rod of length L. The mass per unit length of the rod, ρ, varies as $\rho = \rho_0 s/L$, where ρ_0 is a constant, and s is the distance from the end marked 0. Find the center of mass.

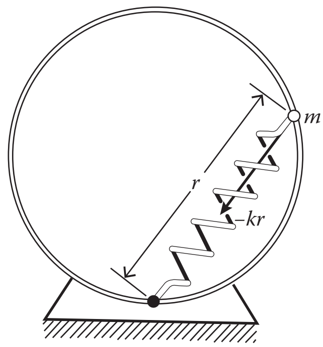


FIGURE 2.15 A bead slides on a vertical hoop.

16. A loaded spring gun, initially at rest on a horizontal frictionless surface, fires a marble at an angle of elevation θ. The mass of the gun is M, the mass of the marble is m, and the muzzle velocity of the marble is v_0. Find the final motion of the gun.
17. Two men of weights W_1 and W_2 are seated in the bow and stern, respectively, of a boat of weight W at a distance L from each other. Ignoring the water resistance, determine the direction and size of the displacement of the boat if the men change places.
18. A bar of mass m is placed on a plank of mass M, which rests on a smooth horizontal plane. The coefficient of friction between the surfaces of the bar and the plank is equal to f. The plank is subject to a horizontal force F of the form $F = at$, where a is a constant. Find
    (a) The moment of time t_0 at which the plank starts sliding from under the bar
    (b) The acceleration of the bar and of the plank in the process of their motion
19. A uniformly straight rigid bar of mass m and length b is placed in a horizontal position across the top of two identical cylindrical rollers. Axes of the two rollers are a distance $2d$ apart. If f is the frictional coefficient between the cylinder surface and the bar, show that if the bar is displaced a distance x from its central position, then the net horizontal force on the bar is $F = -fmgx/d$, and the bar will execute simple harmonic motion with a period of $2\pi\sqrt{d/fg}$.
20. A particle of mass m is attached to the end of a string and moves in a circle of radius r on a frictionless horizontal table. The string passes through a frictionless hole in the table and, initially, the other end is fixed.
    (a) If the string is pulled so that the radius of the circular orbit decreases, how does the angular velocity change if it is ω_0 when $r = r_0$?
    (b) What work is done when the particle is pulled slowly in from a radius r_0 to a radius $r_0/2$?
21. A bead of mass m slides without friction on a vertical hoop of radius R. The bead moves under the combined action of gravity and a spring attached to the bottom of the hoop. For simplicity, we assume that the equilibrium length of the spring is zero, so that the force resulting from the spring is $-kr$, where r is the instantaneous length of the spring, as shown in Figure 2.15. The bead is released at the top of the hoop with negligible speed. How fast is the bead moving at the bottom of the hoop?

REFERENCES

Braginski, V. B. and Panov, V. I. Verification of the equivalence of initial and gravitational mass, *Soviet Physics-JETP*, 34, 463, 1971.

Roll, P., Krotkov, R., and Dicke, R. The equivalence of initial and passive gravitational mass, *Annals Phys.*, 26, 442, 1967.

3 Integration of Newton's Equation of Motion

3.1 INTRODUCTION

In Chapter 2, we saw that Newton's second law can be written as

$$\frac{d\vec{P}}{dt} = \vec{F},$$

where $\vec{P}$ is the momentum of the particle, and $\vec{F}$ is the external applied force. When the mass m is constant, we can write it as a second-order differential equation:

$$m\frac{d^2\vec{r}}{dt^2} = \vec{F}$$

and it is called the equation of motion of the particle. When the force $\vec{F}$ is known, the equation of motion becomes a second-order differential equation for the unknown function $\boldsymbol{r}(t)$, the path or orbit of the particle. The objective of this chapter is to seek solutions to this equation of motion for various given force functions. Solving problems by integrating the equation of motion includes the following operations:

(a) Write the differential equations. For this, we start by choosing a frame of reference (e.g., a coordinate system). Very often, the solution can be greatly simplified by a suitable choice of coordinate system. For Cartesian coordinates, the system's origin is usually chosen to coincide with the initial position of the particle.
(b) Integrate the equations of motion. The integration will depend upon the form of force function on the right-hand side of the equations of motion.
(c) Determine the constants of integration. The integration of the equations of motion will lead to the introduction of constants of integration; the values of the integration constants are determined from the initial conditions. If the differential equation of motion is an equation with separable variables, rather than introducing integration constants, we can evaluate the definite integrals on both sides of the equation over the appropriate range.
(d) Determine the required quantities and analyze the obtained results. A particular problem is solved by integrating the differential equations of motion and evaluating the integration constants. In order to be able to analyze the solution, it should be carried out in the most general form, inserting the numerical data only in the final results.

In general, the force function may be a function of position, velocity, or time. A problem where the applied force is a function of all three variables simultaneously is difficult to solve. The problems that we shall consider, therefore, generally fall into one of the following four categories:

1. The applied force is constant.
2. The applied force is time dependent.

3. The applied force is velocity dependent.
4. The applied force is position dependent.

3.2 MOTION UNDER CONSTANT FORCE

We consider the motion of a projectile under the action of the gravitational force near the Earth's surface. Although the gravitational acceleration varies with locality, in a small local region, it is constant to a good approximation both in magnitude and direction.

Example 3.1: Motion of a Projectile

A particle of mass m is projected with an initial speed $\boldsymbol{v}_0$ at an angle α with the horizontal. Find the following:

(a) The position vector of the particle at any time
(b) The time to reach the highest point
(c) The maximum height reached
(d) The time of flight
(e) The range
(f) The equation of the trajectory
(g) The parabola of safety (region of space in which the projectile can reach)

Solution:

We assume the air resistance is negligible; then the force acting on the projectile after its firing is the gravitational force mg. Let the constant force have a y-axis direction. If the axes are oriented so that the initial velocity $\boldsymbol{v}_0$ of the particle in the xy plane (Figure 3.1), then the subsequent motion will be entirely in this plane.

(a) The position vector $\boldsymbol{r}(t)$ at time t.
The equation of motion of the particle is

$$m\frac{d^2\vec{r}}{dt^2} = -mg\hat{j} \tag{3.1}$$

where

$$\vec{r} = x\hat{i} + y\hat{j}.$$

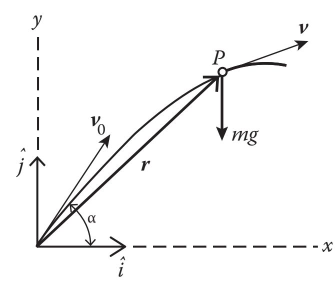


FIGURE 3.1 Path of a projectile.

Integrating once,

$$\vec{v} = \mathrm{d}\vec{r}/\mathrm{d}t = -gt\hat{j} + \vec{C}_1 \tag{3.2}$$

where $\vec{C}_1$ is a constant vector of integration, which can be determined from the initial condition $\vec{v} = \vec{v}_0$ at $t = 0$:

$$\vec{C}_1 = \vec{v}_0 = v_0 \cos\alpha\hat{i} + v_0 \sin\alpha\hat{j}.$$

Then Equation 3.2 becomes

$$\vec{v} = \mathrm{d}\vec{r}/\mathrm{d}t = v_0 \cos\alpha\hat{i} + (v_0 \sin\alpha - gt)\hat{j}. \tag{3.3}$$

Integrating Equation 3.3 once,

$$\vec{r} = v_0 t \cos\alpha\hat{i} + \left(v_0 t \sin\alpha - \frac{1}{2}gt^2\right)\hat{j} \tag{3.4}$$

where we have set the constant of integration $\vec{C}_2 = 0$ because $\boldsymbol{r} = 0$ at $t = 0$. In component form, we have

$$x = v_0 \cos\alpha t, \quad y = v_0 \sin\alpha t - \frac{1}{2}gt^2. \tag{3.5}$$

(b) The time t_h to reach the highest point.
At the highest point of the path, the y component of the velocity vanishes. Thus, we have, from Equation 3.3,

$$v_0 \sin\alpha - gt_h = 0, \qquad \text{or} \qquad t_h = v_0 \sin\alpha/g. \tag{3.6}$$

(c) The maximum height.
Eliminating t between Equations 3.5 and 3.6, we get the maximum height h reached by the projectile:

$$h = y_{\max} = \frac{v_0^2 \sin^2\alpha}{2g}. \tag{3.7}$$

(d) The time of flight.
At the end of flight, $y = 0$. Setting $y = 0$ in Equation 3.5 gives the time of flight:

$$t = 2v_0 \sin\alpha/g. \tag{3.8}$$

It is twice the time to reach the highest point.

(e) The range.
Eliminating t between Equations 3.5 and 3.8, we find the horizontal range R to be

$$R = x = \frac{v_0^2 \sin 2\alpha}{g}. \tag{3.9}$$

The maximum range v_0^2/g occurs at $\alpha = 45°$. Any range less than this can be attained from two angles of projection, one of which is the complement of the other.

(f) The equation of path.
Eliminating t from the two equations in Equation 3.5, we obtain the path of the projectile.

$$y = x\tan\alpha - \frac{g}{2v_0^2\cos^2\alpha}x^2. \tag{3.10}$$

This is a parabola with its vertex at the highest point of the path. It can be put in a simpler form

$$(x - R/2)^2 = (R/2)^2\ (h - y)/h. \tag{3.11}$$

(g) The parabola of safety.
If we want to know the farthest region of space that the projectile can reach, it is necessary to compute the limits of the various paths. Mathematically, we seek the envelope of the one-parameter family of curves, which is the solid, large, parabolic curve of Figure 3.2. We see that the equation of the path (Equation 3.10) is a one-parameter (the projection angle α) family of curves, and we write it as

$$f(x, y, \alpha) = 0. \tag{3.11}$$

Then

$$\mathrm{d}f = \frac{\partial f}{\partial x}\mathrm{d}x + \frac{\partial f}{\partial y}\mathrm{d}y + \frac{\partial f}{\partial \alpha}\mathrm{d}\alpha = 0$$

or

$$\nabla f \cdot \mathrm{d}\vec{s} + \frac{\partial f}{\partial \alpha}\mathrm{d}\alpha = 0$$

where
$\mathrm{d}\vec{s} = \mathrm{d}x\hat{i} + \mathrm{d}y\hat{j}$ is the tangent to the curve at the point $P(x,y)$, and $\nabla f = \hat{i}\partial f/\partial x + \hat{j}\partial f/\partial y$ is perpendicular to the curve at $P(x,y)$.
Hence,

$$\nabla f \cdot \mathrm{d}\vec{s} = 0,$$

and accordingly, $\partial f/\partial\alpha = 0$.
This is the condition from which the equation of the envelope is to be computed. Now, we rewrite Equation 3.10 to give the explicit form of f:

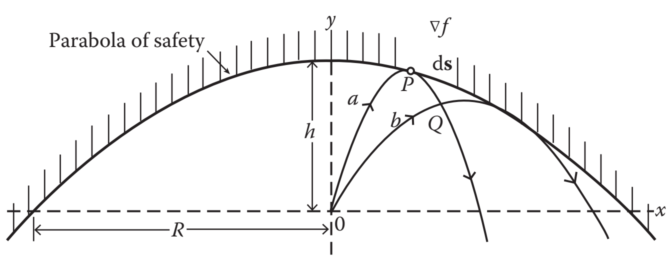


FIGURE 3.2 Parabola of safety and typical trajectories.

$$f = y - \tan\alpha x + \frac{g}{2v_0^2\cos^2\alpha}x^2 = 0.$$

Differentiating once with respect to α, we get

$$v_0^2 = g\tan\alpha x.$$

Substituting this into Equation 3.10 to eliminate α, we obtain

$$x^2 = -\frac{2v_0^2}{g}\left(y - \frac{v_0^2}{2g}\right) \tag{3.12}$$

which is the equation of the envelope, a vertical parabola with the vertex on the y-axis at a distance $v_0^2/2g$ above the origin. No points lying outside the envelope can be reached from O by a projectile having the initial velocity $\vec{v}_0$. A point on the envelope is touched by only a single trajectory, whereas two paths corresponding to two different angles of projection may reach points in the interior. This suggests to us a simple alternative derivation of the equation of the envelope, and we now outline it below.

First, let us rewrite Equation 3.10 in a different form. To this end, we put $w = \tan\alpha$; then Equation 3.10 becomes

$$y = wx - \frac{gx^2}{2v_0^2}(1 + w^2)$$

or

$$w^2 - \frac{2v_0^2}{gx}w + \frac{2v_0^2}{gx^2}y + 1 = 0. \tag{3.13}$$

The two roots of this quadratic in w (= tan α) specify the angles of projection of two paths, such as a and b in Figure 3.2 passing through the point Q. As the two paths approach each other, Q approaches the envelope of the paths. Consequently, the equation of the envelope is the relationship between x and y that makes the roots of Equation 3.13 equal, that is, the relationship obtained by equating to zero the discriminant of the quadratic. This equation is identical to Equation 3.12.

We now calculate the time t_P when the projectile meets the parabola of safety. To this end, we substitute Equation 3.5 into Equation 3.12 and solve for t, obtaining

$$t_P = \frac{v_0}{g\sin\alpha}.$$

We also note that

$$t_P - t_h = \frac{v_0\cos^2\alpha}{g\sin\alpha} \geq 0, \quad 0 \leq \alpha \leq \pi.$$

Thus, the trajectory touches the parabola of safety after it has reached the highest point.

Obviously, some changes must be made in the preceding analysis when applying it to the case where a projectile is launched to a higher or lower level than its starting point. For example, the launching angle for maximum range is no longer 45°. We can show this easily. The coordinates of point P (Figure 3.3) are found to be the solution of the equation of trajectory:

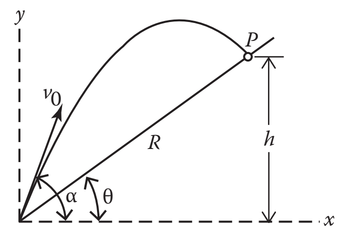


FIGURE 3.3 Trajectory of a projectile.

$$y = x\tan\alpha - \frac{g}{2v_0^2\cos^2\alpha}x^2$$

and

$$y = x \tan \theta$$

where θ is the slope of the inclined plane with respect to the horizontal. Eliminating y, we obtain

$$\begin{aligned} R &= (x)_P \sec\theta \\ &= \left(v_0^2 \sin 2\alpha/g - 2v_0^2\cos^2\alpha\tan\theta/g\right)\sec\theta. \end{aligned}$$

The maximum range along the incline, for a given θ, occurs at

$$(\alpha)_{max} = \theta/2 + \pi/4 \tag{3.14}$$

and the maximum range itself is

$$R_{max} = v_0^2 g^{-1}(1 - \sin\theta)\sec^2\theta.$$

Figure 3.3 illustrates the case where a projectile is launched to a higher level; θ and h are positive. For the case where a projectile is launched to a lower level, θ and h are negative. Alternatively, we can find the maximum range of a projectile along an inclined plane in a simpler way without recourse to calculus. We resolve the motion into two components: x' (parallel to the slope) and y' (perpendicular to the slope). Then, we have

$$x' = v_0\cos\phi t - \frac{1}{2}g\sin\theta t^2$$

$$y' = v_0\sin\phi t - \frac{1}{2}g\cos\theta t^2$$

where ϕ is the angle between $\boldsymbol{v}_0$ and the inclined plane, that is, $\phi = \alpha - \theta$. Solving t when $y' = 0$ and substituting into the equation for x', we obtain the range R along the inclined plane:

$$R = \frac{2v_0^2\sin\phi\cos(\phi+\theta)}{g\cos^2\theta}.$$

Using the identity

$$2\cos A\sin B = \sin(A + B) - \sin(A - B),$$

the range R becomes

$$R = \frac{v_0^2}{g\cos^2\theta}\left[\sin(2\phi + \theta) - \sin\theta\right],$$

and the maximum range occurs at

$$\phi = \pi/4 - \theta/2$$

which is identical to the previous result (Equation 3.14).

3.3 FORCE IS A FUNCTION OF TIME

If the force acting on a particle is a function of time only, we write the equation of motion in the form

$$m\frac{d\vec{v}}{dt} = \vec{F}(t).$$

Integrating once, we obtain

$$m(\vec{v} - \vec{v}_0) = \vec{P} - \vec{P}_0 = \int_{t_0}^{t} \vec{F}(t')\,dt' \tag{3.15}$$

in which $\vec{v}$ is the velocity of the particle at time t, and $\boldsymbol{v}_0$ is the initial velocity at time t_0. Integrating once, we obtain the particle path $\boldsymbol{r}(t)$:

$$\vec{r}(t) - \vec{r}_0 = \vec{v}_0(t - t_0) + \frac{1}{m}\int_{t_0}^{t} dt' \int_{t_0}^{t''} \vec{F}(t'')\,dt'' \tag{3.16}$$

where we have used the initial condition: $\boldsymbol{r} = \boldsymbol{r}_0$ at $t = t_0$. If the necessary integration can be performed explicitly, then we have an explicit solution for the position of the particle as a function of time. Otherwise, we can perform a numerical integration and obtain a numerical solution to the problem.

Example 3.2: Reflection of Radio Waves from the Ionosphere

The interaction of radio waves with charged particles in the Earth's upper atmosphere is an interesting and important problem. These charged particles are electrons and positive ions, which are formed when the ultraviolet light from the sun is absorbed by the atoms and molecules of the upper atmosphere. The charged particles tend to be trapped by the magnetic field of the Earth and stay in the upper regions, forming the ionosphere. Because the electrons are much lighter than the positive ions, the electric fields of radio waves passing through the ionosphere accelerate them more; thus, these electrons are more effective in modifying the propagation of the radio waves. The ionosphere may be regarded as a tenuous free electron gas.

A free electron of charge $-e$ in the electric field of an electromagnetic plane wave has the equation of motion

$$m\frac{d^2x}{dt^2} = -eE_0\sin(\omega t) \tag{3.17}$$

where $E = E_0 \sin(\omega t)$ is the electric field of the electromagnetic plane wave. Integrating once, we obtain the velocity as a function of time:

$$v = v_0 - \frac{eE_0}{m\omega}(1 - \cos\omega t). \tag{3.18}$$

Integrating once again, we obtain the position as a function of time:

$$x = x_0 + \left(v_0 - \frac{eE_0}{m\omega}\right)t + \frac{eE_0}{m\omega^2}\sin\omega t \tag{3.19}$$

where we assume that at $t = 0$, the electron has the position x_0 and the velocity v_0. The first two terms indicate that the electron is drifting with a uniform velocity. Superimposed on this drifting motion is an oscillating motion represented by the last term; the oscillating frequency ω of the electron is the same as the frequency of the incident electromagnetic waves. This oscillating part of the displacement x gives rise to an electric dipole moment

$$\vec{p} = -\frac{e^2}{m\omega^2}\vec{E}. \tag{3.20}$$

Each of the electrons of the electron gas experiences the externally applied field $\vec{E}$ and an internal field caused by the induced dipole moments of the other electrons. But the electron density N is very low in the ionosphere; the second contribution can be neglected; and the macroscopic polarization, $\vec{P}$, is given by

$$\vec{P} = N\vec{p} = -\frac{Ne^2}{m\omega^2}\vec{E}. \tag{3.21}$$

The refractive index n of the electron gas will tell us what happens to radio waves traveling through the ionosphere. The refractive index n of a medium is given by $n = c/u$, where c and u are the velocities of light in a vacuum and in a medium, respectively, and they are given by

$$c = \frac{1}{\sqrt{\varepsilon_0\mu_0}}$$

and

$$u = \frac{1}{\sqrt{\varepsilon\mu}}$$

where ε_0 and ε are the electric permittivities, and μ_0 and μ are the magnetic permeabilities, with $\mu_0/\mu \cong 1$. Thus,

$$n = c/u = \sqrt{\varepsilon/\varepsilon_0} = \sqrt{\chi}$$

and χ is referred to as the relative permittivity, which is related to $\vec{E}$, $\vec{P}$, and the electric displacement $\vec{D}$ by the relationship

$$\vec{D} = \varepsilon_0\vec{E} + \vec{P} = \varepsilon\vec{E} = \chi\varepsilon_0\vec{E}$$

from which we have

$$\chi = 1 + \frac{1}{\varepsilon_0}\frac{P}{E} = 1 - \frac{Ne^2}{\varepsilon_0 m\omega^2}.$$

By putting

$$\nu_p = \frac{1}{2\pi}\frac{Ne^2}{\varepsilon_0 m\omega}, \quad \omega = 2\pi\nu$$

we can write

$$\chi = 1 - (\nu_p/\nu) \quad \text{and} \quad n = \sqrt{1-(\nu_p/\nu)^2}.$$

It is easy to see that when v is equal to v_p, both χ and n become zero. When v is less than v_p, χ becomes negative and n becomes imaginary. We are now ready to investigate what happens to radio waves traveling through the ionosphere:

1. n is real but less than unity ($v > v_p$). According to Snell's law ($n_1 \sin\theta_1 = n_2 \sin\theta_2$), a radio wave would be refracted away from the normal when it strikes the ionosphere. The angle of refraction θ_r becomes 90° when $\sin\theta_i = n$. For angles of incidence θ_i greater than this, the wave is totally reflected. In fact, the edge of the ionosphere is not sharp, but reflection must occur for all angles of incidence given by $\sin\theta_i \geq n_{\min}$; $n_{\min}$ is the minimum refractive index of the ionosphere, occurring at the height of the maximum electron density. These effects are illustrated in Figure 3.4. Figure 3.4a shows that for the highest-frequency waves, the refractive index n of the ionosphere is approximately unity, and the waves are slightly refracted away from the normal. In Figure 3.4b, we see that, at lower frequencies, n is smaller, and the waves are refracted back toward the Earth. Figure 3.4c shows that, at still lower frequencies, n is less than $\sin\theta_i$ at the bottom of the ionosphere, and the wave is totally reflected.
2. n is imaginary ($v \leq v_p$). It has been shown that there is no net flow of energy and no energy absorption by the ionosphere. Because of these phenomena, we can conclude that the wave is completely reflected by the ionosphere for any angle of incidence.

We now return to Equation 3.15. The time integral of the force on the right-hand side is the impulse of the force delivered to the particle, and Equation 3.15 is referred to as the impulse–momentum relationship. Though perfectly general, it is particularly useful in analyzing the result of applying

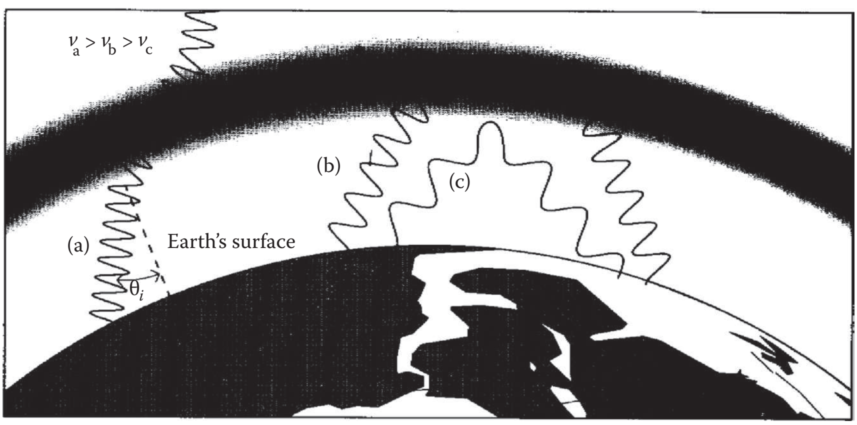


FIGURE 3.4 Refraction and reflection of radio waves. (a) For higher-frequency waves, $n \sim 1$, (b) for lower-frequency waves, $n <\sim 1$ (c) for still lower-frequency waves, total reflection occurs.

a large force to the particle for a very short duration. Utilizing the concept of impulse, George Green (1793–1841) developed a powerful solution technique for general force functions. We now review Green's method.

3.3.1 Impulsive Force and Green's Function Method

Green's method is based upon representing the force $\vec{F}(t)$ as a sum of impulsive forces $\vec{F}_n(t)$, each acting during a very short time interval Δt and delivering an impulse $\vec{F}_n(t)\Delta t$ (Figure 3.5):

$$\vec{F}(t) = \sum_{n=-\infty}^{\infty} \vec{F}_n(t) \tag{3.22}$$

where

$$\vec{F}_n(t) = \begin{cases} 0, & t < t_n \\ \vec{F}(t_n), & t_n \le t \le t_{n+1} \\ 0, & t > t_{n+1} \end{cases} \tag{3.23}$$

where $t_n = n\Delta t$. As $\Delta t \to 0$, the sum of all the impulsive forces $\vec{F}_n(t)$ will approach $\vec{F}(t)$.

To illustrate this method explicitly, consider a free particle of mass m initially at rest. What is the motion of the particle when it is subjected to a force $\vec{F}(t)$ at $t = 0$ and acting continuously thereafter? Instead of using the second law of motion to investigate the motion of the particle, we follow George Green and break the force $\vec{F}(t)$ up into a series of impulsive forces occurring at time t'. If acting alone, each of these impulsive forces would produce a velocity and a displacement of the particle at time t given by, respectively,

$$v(t,t') = \frac{\bar{F}(t')\Delta t}{m}$$

$$x(t,t') = \frac{\bar{F}(t')\Delta t}{m}(t - t'), \quad t > t'$$

where $\bar{F}$ is the average force over the short time interval Δt. By the principle of super-position (it is valid here because of the linearity of the system), the actual displacement is caused by all the impulsive forces that $F(t)$ is thought to be composed of, and in the limit, we have

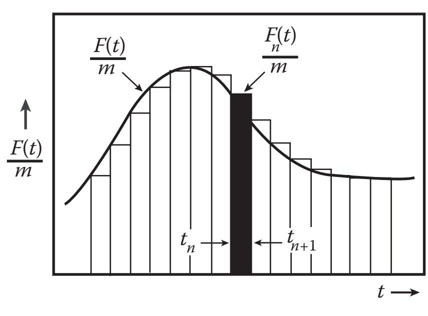


FIGURE 3.5 Representation of a force as a sum of impulses.

$$x(t)=\int_0^t \frac{t-t'}{m}F(t')\,\mathrm{d}t'.$$

We can rewrite this as

$$x(t)=\int_0^t G(t,t')F(t')\,\mathrm{d}t', \tag{3.24}$$

where we have defined

$$G(t,t') = (t - t')/m. \tag{3.25}$$

$G(t,t')$ is called Green's function for this particular system and is here a function only of the difference $(t - t')$. The equation for $x(t)$ is an example of an integral equation, $G(t,t')$ being a so-called integral kernel, which, here, converts the applied force function $F(t)$ into the displacement function $x(t)$.

To complete the above example, we assume that $F(t) = F_0$, a constant, for $t \geq 0$. Then

$$x(t)=\int_0^t \frac{t-t'}{m}F(t')\,\mathrm{d}t' = \frac{F_0}{m}\int_0^t (t-t')\,\mathrm{d}t'$$

$$= \frac{F_0}{m}(tt' - t'^2/2)\Big|_0^t = \frac{1}{2}\frac{F_0}{m}t^2$$

which is the expected result for a constant applied force with the assumed initial condition.

The Green's function method is generally very useful for the solution of linear, inhomogeneous differential equations. Green's function $G(t, t')$ is the solution of the differential equation for an infinitesimal element of the inhomogeneous part. The Green's function appropriate to a given problem depends on the nature of the linear system considered as well as on the particular set of initial conditions. As in the preceding example, the Green's function $G(t, t')$ represents the response of the system to an impulsive force. By imagining the actual disturbance made by such impulsive forces and using Green's function to propagate the effect of each impulse up to a given time, we get the system's response to each of the impulses at that time. Finally, we add all of the responses together to get the response of the system at that time to the actual disturbances. The Green's function method is particularly well adapted to the numerical solution of the problem when $F(t)$ is known numerically.

In engineering books, $G(t, t')$ is usually referred to as the impulse response function.

3.4 FORCE IS A FUNCTION OF VELOCITY

If the force acting on the particle is known explicitly as a function of velocity, the equation of motion may not have solutions for the general case. However, solutions do exist for one-dimensional motion; the equation of motion now takes the form

$$m\mathrm{d}v/\mathrm{d}t = F(v) \tag{3.26}$$

or

$$mv\mathrm{d}v/\mathrm{d}x = F(v). \tag{3.27}$$

Equations 3.26 and 3.27 are both first-order differential equations and, therefore, always have solutions. From Equation 3.26, we obtain time as a function of the velocity:

$$t - t_0 = \int_{v_0}^{v} \frac{m\mathrm{d}v}{F(v)} \tag{3.28}$$

where v_0 is the initial velocity at time t_0. Equation 3.27 gives the position as a function of the velocity:

$$x - x_0 = \int_{v_0}^{v} \frac{mv\mathrm{d}v}{F(v)}\,. \tag{3.29}$$

Equations 3.28 and 3.29 together represent a parametric solution to the equation of motion. We can eliminate v between these two equations to directly obtain the position as a function of time.

Example 3.3: Motion under Gravity with Resistance

We consider, as an example, the sport of skydiving as motion under the action of a velocity-dependent force: a jumper falling in the atmosphere from a height smaller than the Earth's radius so that both the jumper's weight and the air density can be considered constant. Assuming the parachute opens at the beginning of the fall, there are two forces acting on the parachute: the downward force of gravity and the upward force of air resistance. If we choose a coordinate system with $y = 0$ at the Earth's surface and increasing upward, then the equation of motion of a falling object becomes

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = -mg + kv^n \tag{3.30}$$

where k is a positive constant, containing dependent factors other than velocity. In general, the air resistance (or the retarding forces) is very complicated, but the power-law approximation is useful in many instances in which the velocity does not vary appreciably. Experiments show that for a relatively small object moving in air, $n = 1$ for velocities less than 24 m/s; at higher but still subsonic velocities up to approximately 300 m/s, the resistance is approximately proportional to the square of the velocity ($n = 2$). The term Stokes' law of resistance is often applied for the case of $n = 1$, and with Newton's law of resistance for the case of $n = 2$. For the skydiving case, we approximate the air resistance on the parachute as square law, and so the equation of motion takes the form

$$m\frac{\mathrm{d}v}{\mathrm{d}t} = -mg + kv^2$$

which is separable as

$$\frac{\mathrm{d}v}{v^2 - v_t^2} = \frac{k}{m}\mathrm{d}t \tag{3.31}$$

where $v_t^2 = mg/k$. Now,

$$\frac{1}{v^2 - v_t^2} = \frac{1}{(v+v_t)(v-v_t)} = \frac{1}{2v_t}\left(\frac{1}{v-v_t} - \frac{1}{v+v_t}\right).$$

Equation 3.31 becomes

$$\frac{1}{2v_t}\left(\frac{1}{v-v_t} - \frac{1}{v+v_t}\right)\mathrm{d}v = \frac{k}{m}\mathrm{d}t.$$

Integrating yields

$$\frac{1}{2v_t}\ln\left(\frac{v-v_t}{v+v_t}\right)=\frac{k}{m}t+C$$

where C is an integration constant. Solving for v, we finally obtain

$$v(t)=v_t\,\frac{1+B\exp(-2gt/v_t)}{1-B\exp(-2gt/v_t)} \tag{3.32}$$

where $B = \exp(2v_tC)$. For the constants of integration, we need to know the initial conditions and the value of k, which is approximately 30 km/s for the Earth's atmosphere and a standard-size parachute.

It is easy to see that as $t \gg v_t/2g$, $\exp(-2gt/v_t) \to 0$, and so $v \to v_t$; that is, if the diver falls from a sufficient height, the diver will eventually fall at a constant velocity given by v_t, the terminal velocity. To determine the constants of integration, we need to know the value of k, which is approximately 30 km/s for the Earth's atmosphere and a standard parachute.

Example 3.4: Projectile Motion in a Viscous Medium

In example 3.1, we have neglected the resistance of the air. If air exerts a resistance of $-mk\mathbf{v}$ on the projectile (this means that the velocity of the projectile is low), the equation of motion then takes the form

$$\frac{d\vec{v}}{dt}=-g\hat{j}-k\vec{v} \tag{3.33}$$

where k is a constant. Multiplying by the integrating factor e^{kt}, we obtain

$$\left(\frac{d\vec{v}}{dt}+k\vec{v}\right)e^{kt}=-ge^{kt}\hat{j}$$

or

$$\frac{d}{dt}\left(\vec{v}e^{kt}\right)=-ge^{kt}\hat{j}.$$

Integration yields

$$\vec{v}e^{kt}=-\frac{g}{k}e^{kt}\hat{j}+\vec{C}_1. \tag{3.34}$$

Using the initial velocity at $t = 0$,

$$\vec{v}_0=v_0\cos\alpha\hat{i}+v_0\sin\alpha\hat{j}$$

we obtain

$$\vec{C}_1=v_0\cos\alpha\hat{i}+(v_0\sin\alpha+g/k)\hat{j}.$$

Substituting this into Equation 3.34, we obtain

$$\vec{v} = \frac{d\vec{r}}{dt} = v_0(\cos\alpha\hat{i} + \sin\alpha\hat{j})e^{-kt} - \frac{g}{k}(1-e^{-kt})\hat{j}. \tag{3.35}$$

Thus, after a long time has elapsed, the velocity of the projectile approaches the terminal value $-g/k$.

Integrating Equation 3.35 once and using the initial condition $\boldsymbol{r} = 0$ at $t = 0$, we find

$$\vec{r} = \frac{v_0}{k}(\cos\alpha\hat{i} + \sin\alpha\,\hat{j})(1-e^{-kt}) - \frac{g}{k}\left[t + \frac{1}{k}(e^{-kt}-1)\right]\hat{j} \tag{3.36}$$

or

$$x = \frac{v_0\cos\alpha}{k}(1-e^{-kt}) \tag{3.37}$$

$$y = -\frac{g}{k}t + \frac{1}{k}(v_0\sin\alpha + g/k)(1-e^{-kt}). \tag{3.38}$$

From Equation 3.37, we see that the limiting value of x for the large value of t is

$$\lim_{t\to\infty} x = (v_0/k)\cos\alpha.$$

Thus, the trajectory approaches asymptotically the vertical line $(v_0/g)\cos\alpha$ with a terminal speed of $-g/k$.

Eliminating t between Equations 3.37 and 3.38, we obtain the approximate equation of the path of the trajectory (Figure 3.6):

$$y = \frac{v_0\sin\alpha + g}{v_0\cos\alpha}x + \frac{g}{k^2}\ln\left(1 - \frac{k}{v_0\cos\alpha}x\right). \tag{3.39}$$

When a problem is generalized by the inclusion of an effect previously ignored, it is a good idea to check that the original result is still obtained when the magnitude of the extra effect tends to zero in the revised result. Let us now consider the limit of Equation 3.39 when the air resistance is negligible. When the value of the quantity $kx/(v_0\cos\alpha)$ in the logarithm term is much smaller than unity, we can use the logarithm series

$$\ln(1-z) = \ln[1+(-z)] = -(z + z^2/2 + z^3/3 + \ldots),\ z^2 < 1.$$

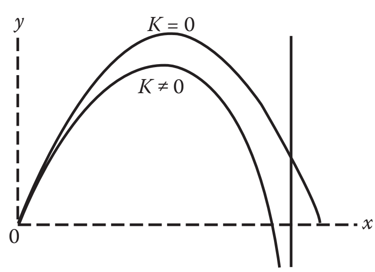


FIGURE 3.6 Resisted projectile.

Now, let $z = kx/(v_0 \cos\alpha)$ in the logarithm term, and using the logarithm series, we find that Equation 3.39 reduces to

$$y = \tan\alpha x - \frac{1}{2}\frac{g}{v_0^2\cos^2\alpha}x^2 - \frac{1}{3}\frac{kg}{v_0^3\cos^3\alpha}x^3 - \frac{1}{4}\frac{k^2 g}{v_0^4\cos^3\alpha}x^4 + \ldots$$

which, when $k = 0$, reduces to the equation of the path of the trajectory in air without resistance (Equation 3.10).

Example 3.5: Motion of Charged Particles in Magnetic Fields

The motion of a charged particle of mass m and velocity $\vec{v}$ in prescribed electric and magnetic fields $\vec{E}$ and $\vec{B}$ is governed by the Lorentz equation of motion:

$$m\mathrm{d}\vec{v}/\mathrm{d}t = q(\vec{E} + \vec{v}\times\vec{B}). \tag{3.40}$$

The right-hand side is the Lorentz force $\vec{f} = q(\vec{E} + \vec{v}\times\vec{B})$. Both vectors $\vec{E}$ and $\vec{B}$ can be a function of position and time. In many applications, the particles used can interact with each other and with other particles, primarily through collisions. It is beyond the scope of this book to include the effect of collisions in Equation 3.40, so we shall not try to do so.

Knowledge of the trajectories followed by charged particles in various electric and magnetic field configurations is of considerable interest in plasma physics, astrophysics, and high-energy physics (in the design of apparatus for producing particles with large kinetic energies). We consider only some simple but interesting cases, beginning with the motion of a charged particle in a uniform magnetic field, then generalizing to a nearly uniform magnetic field with field lines slowly converging in space.

3.4.1 Motion in a Uniform Magnetic Field

We assume $\vec{B} \neq 0$ and is independent of time. Then, with $\vec{E} = 0$, Equation 3.40 reduces to

$$m\mathrm{d}\vec{v}/\mathrm{d}t = q\vec{v}\times\vec{B} \tag{3.41}$$

and the Lorentz force is $\vec{f} = q\vec{v}\times\vec{B}$. Because the Lorentz force $\vec{f}$ is perpendicular to $\vec{v}$, the magnetic force does no work on the particle, and so its kinetic energy remains constant. The latter point can be demonstrated easily. Taking the dot product of Equation 3.41 with $\boldsymbol{v}$ and noting that $\vec{v}(\vec{v}\times\vec{B}) = 0$, we obtain

$$m\vec{v}\cdot\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}mv^2\right) = 0.$$

That is, the kinetic energy of the particle is conserved. This result is quite general and is valid for any arbitrary magnetic field.

A convenient way to study Equation 3.40 is to separate the velocity $\boldsymbol{v}$ into two principal directions, one parallel to and the other perpendicular to $\vec{B}$, $\vec{v} = \vec{v}_\perp + \vec{v}_\parallel$. Equation 3.40 is then split into two equations, one describing the parallel and the other the perpendicular motion of the particles:

$$\mathrm{d}\vec{v}_\parallel/\mathrm{d}t = 0, \quad \mathrm{d}v_\perp/\mathrm{d}t = (q/m)(v_\perp\times\vec{B}).$$

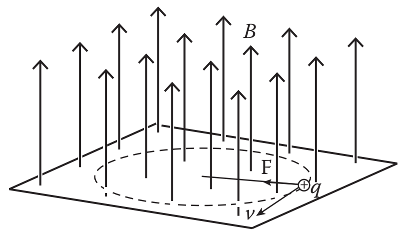


FIGURE 3.7 Path of a charged particle moving perpendicular to a uniform magnetic field.

We see that $\vec{v}_{\parallel}$ = constant; the particle, therefore, moves with a uniform velocity along the direction of $\vec{B}$. Also, $\mathrm{d}\vec{v}_{\perp}/\mathrm{d}t$ is always perpendicular to both $\vec{v}_{\perp}$ and $\vec{B}$, so the perpendicular component makes the particle travel in a circular path (Figure 3.7).

To find the radius R of the circle, we recognize that the Lorentz force $qv_{\perp}B$ provides the particle with the centripetal force $mv_{\perp}^2/R$ that keeps it moving in a circle. Equating the Lorentz and centripetal forces, we have

$$qv_{\perp}B = mv_{\perp}^2/R.$$

Solving for R,

$$R = \frac{mv_{\perp}}{qB}. \tag{3.42}$$

The radius R is often called the Larmor radius of the particle. We can obtain the period of revolution T and the angular frequency ω_c by noting that in one period, the particle travels a total distance $2\pi R$ so that $T = 2\pi R/v_{\perp}$, and ω_c is given by $\omega_c = 2\pi/T$. Combining these expressions with Equation 3.42, we obtain

$$T = 2\pi m/qB \text{ and } \omega_c = qB/m,$$

where ω_c is known as the cyclotron frequency. The complete motion of the charged particle is described as a gyration of the particle in an orbit (known as the Larmor orbit) superimposed on the uniform motion of the orbit center, or guiding center, along a magnetic field line. The resulting helical motion is shown in Figure 3.8.

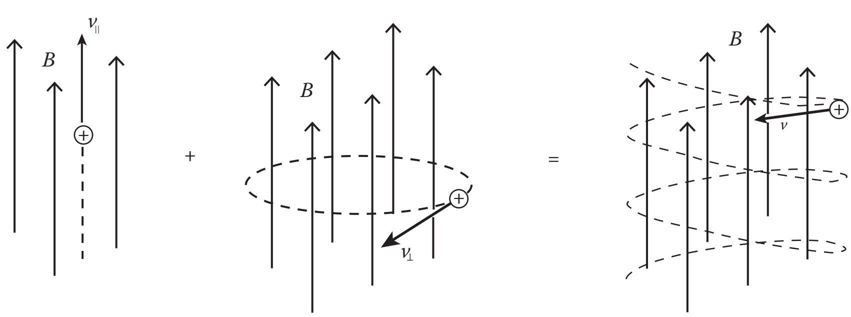


FIGURE 3.8 Particle motion in a uniform magnetic field.

An interesting quantity, which we will use later, is the magnetic moment of the grating particle. By definition, the magnetic moment of a current loop, $\vec{\mu}$, is given by

$$\vec{\mu} = \text{current} \times \text{area} = \frac{qv_\perp}{2\pi R} \times (\pi R^2) = \frac{K_\perp}{B}. \tag{3.43}$$

There, $K_\perp$ is the kinetic energy associated with the motion perpendicular to $\vec{B}$.

3.4.2 Motion in Nearly Uniform Magnetic Field

Each case in which $\vec{B}$ is not uniform generally has to be considered separately. But there is one case of considerable interest that we can discuss approximately. This is the one in which $\vec{B}$ has axial symmetry, and the field lines are slowly converging in space. Thus, if we take the z-axis of cylindrical coordinates to be the axis of symmetry, we can write $\vec{B}(z, \rho)$ and assume that the variation of $\vec{B}$ with both z and ρ is not great. This would correspond to lines of $\vec{B}$ like those in Figure 3.9. Then, we can expect that the motion of the particle will still be approximately helical as shown in Figure 3.9.

Taking the z-component of Equation 3.41, we obtain

$$m\frac{\mathrm{d}v_z}{\mathrm{d}t} = -qv_\phi B_\rho. \tag{3.44}$$

We can find the approximate value of B_ρ by using $\mathrm{div}\vec{B} = \nabla\vec{B} = 0$, which now becomes

$$\frac{1}{\rho}\frac{\partial(\rho B_\rho)}{\partial\rho} = -\frac{\partial B_z}{\partial z}. \tag{3.45}$$

Because the field lines are converging slowly, we can set $\partial B_z/\partial z \cong$ constant over the orbit cross section; then, the integration of Equation 3.45 gives

$$B_\rho = -\frac{1}{2}\rho\frac{\partial B_z}{\partial z}. \tag{3.46}$$

In a slowly varying magnetic field $\boldsymbol{B}$, the radius of the helix will not change very much for each turn, so we can take $v_\rho \ll v_\phi$ and then $v_\phi \cong -v_\perp$ because $v_\perp$ and $\hat{e}_\phi$ are directed oppositely for $+q$. Making this substitution, along with Equation 3.46, and using Equation 3.42 to express ρ as the radius of the orbit, we find that

$$m\frac{\mathrm{d}v_z}{\mathrm{d}t} = -\frac{1}{2}qv_\perp\rho\frac{\partial B_z}{\partial z} = -\frac{1}{2}\frac{mv_\perp^2}{B}\frac{\partial B_z}{\partial z} = -\vec{\mu}\frac{\partial B_z}{\partial z} \tag{3.47}$$

which indicates that the direction of the z-component of the force is such that it accelerates particles toward the weaker part of the field. As the particle is gyrating toward the region of the stronger magnetic field, $v_\parallel$ is decreased; on the other hand, conservation of energy requires that the orbital motion $v_\perp$ be simultaneously sped up. Thus, gyrating particles that are approaching regions of stronger magnetic field are slowed down. If the magnetic field is sufficiently convergent, the particle will gyrate in an ever-tighter helical spiral until it is reflected back into the weaker field. This is known as the magnetic mirror effect.

Magnetic mirrors are found both in the laboratory and in nature. In a nuclear fusion laboratory, a pair of them are used as a magnetic bottle (Figure 3.10) to contain hot plasma. If a solid container

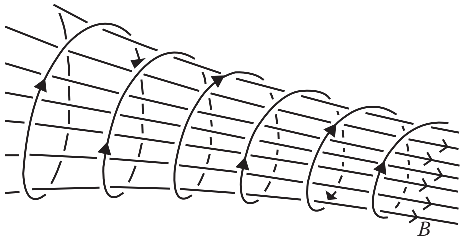


FIGURE 3.9 Particle motion in a nonuniform magnetic field.

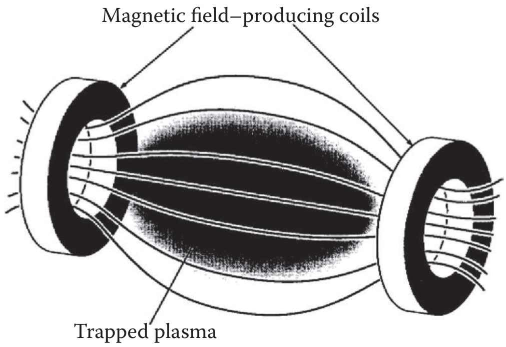


FIGURE 3.10 "Magnetic bottle."

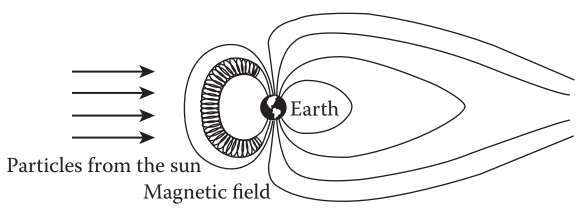


FIGURE 3.11 Trapping of electrons and protons from the sun in geomagnetic field.

was used, contact with its walls would cool the plasma and the ions would not have enough energy for fusion to take place.

Electrons and protons from the sun strike the Earth's magnetic field, which is approximately configured as shown in Figure 3.11. The particles become trapped in an oscillation back and forth between mirror points in the high latitudes, forming the Van Allen belt. Those particles that enter the mirror system with high values of $2mv_{\parallel}^2$, such that the magnetic field cannot reflect them, are dumped into the upper portion of the atmosphere at these latitudes. Here, they collide with atmospheric atoms and molecules, exciting them and thus producing the aurora displays.

3.5 FORCE IS A FUNCTION OF POSITION

Forces that are functions of position occur frequently in nature. One example is the gravitational force. The equation of motion can be written as

$$m\ddot{\vec{r}} = \vec{F}(r). \tag{3.48}$$

The left-hand side of this equation is a time derivative, and the right-hand side is a function of position; hence, it is desirable to integrate the left-hand side with respect to the time and the right-hand side with respect to r. This is accomplished by making use of the identity

$$\frac{d\vec{r}}{dt}dt = d\vec{r}.$$

Taking the dot product of Equation 3.48 with this identity, we obtain

$$m\frac{d^2\vec{r}}{dt^2}\cdot\frac{d\vec{r}}{dt}dt = \vec{F}(r)\cdot d\vec{r}.$$

Integrating from the initial time t_0 to the time t

$$\frac{1}{2}m\frac{d\vec{r}}{dt}\cdot\frac{d\vec{r}}{dt}\Bigg|_{t_0}^{t} = \int_{r_0}^{r}\vec{F}(r)\cdot d\vec{r}$$

where $\boldsymbol{r}_0$ is the initial position at t_0. If $\boldsymbol{v}_0$ is the initial velocity vectors at t_0, then

$$\frac{1}{2}m\left(v^2 - v_0^2\right) = \int_{r_0}^{r}\vec{F}(r)\cdot d\vec{r}. \tag{3.49}$$

If the force $\vec{F}$ is a conservative one, the work integral can always be evaluated, either explicitly or numerically.

$$\int_{r_0}^{r}\vec{F}\cdot d\vec{r} = -\int_{r_0}^{r}\nabla V(r)\cdot d\vec{r} = V(r_0) - V(r).$$

Combining this with Equation 3.49, we have the law of conservation of energy:

$$mv^2/2 + V(r) = mv_0^2/2 + V(r_0). \tag{3.50}$$

3.5.1 Bounded and Unbounded Motion

In many mechanical problems, a qualitative or quantitative analysis of a particle's motion, influenced by a force that depends on the position of the particle, may be carried out by using the law of conservation of energy and an energy diagram without integrating the equation of motion. For example, for one-dimensional motion, we have, from the energy conservation law (Equation 3.50),

$$v = \dot{x} = \sqrt{\frac{2}{m}[E - V(x)]} \tag{3.51}$$

where E is the total energy of the particle. Thus, when $V(x) < E$, v is real. When $V(x) = E$, the velocity v goes to zero, which means that the particle will come to rest at these points and then reverse its motion, returning to its original position. These points are called the turning points of the motion. For example, if the potential energy function $V(x)$ has the form shown in Figure 3.12, where the horizontal line is the total energy E, then the regions between points x_1 and x_3 and points x_5 and x_7—that is, regions II and IV—are inaccessible to the particle. In these two regions, the kinetic energy of the

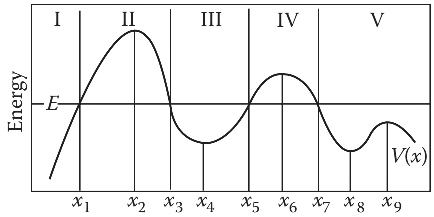


FIGURE 3.12 Energy diagram.

particle is negative, and so its velocity is not real. All the other regions in Figure 3.12 are physically accessible while points x_1, x_3, x_5, and x_7 are the turning points of the motion. Region III has two turning points, a bounded region. A particle moving in such a region is said to perform bound motion. For one-dimensional bound motion, it is necessarily periodic. Both regions I and V have only one turning point, extending from the other end to infinity. So the motion of the particle in these regions is unbounded. Region I includes the origin, and the particle can pass through it. However, classically, we cannot always solve a problem in which the particle passes through the origin or comes very close to it. So we shall exclude such motion from meaningful classical considerations. Regions II and IV are bounded regions that are not accessible to particles and are called potential barriers, and the accessible bounded region III is called the potential well. These concepts are very important in modern physics. For example, the nuclear shell model requires an understanding of potential wells, and the quantum tunnel effect relates to potential barriers.

3.5.2 Stable and Unstable Equilibrium

As shown in Figure 3.12, the slope of the potential energy function vanishes at points x_4, x_6, x_8, and x_9:

$$\frac{\mathrm{d}V(x)}{\mathrm{d}x} = 0. \tag{3.52}$$

Thus, the force is zero at these points, and when a particle is placed there with zero velocity, it remains at rest. The particle is said to be in equilibrium, so the point is called an equilibrium point. It is easy to see that points x_4 and x_8, at which the potential energy has a minimum value, are stable, and points x_6 and x_9, at which the potential energy has a maximum value, are unstable. To show this distinction, if the displacement of the particle from x_e, the equilibrium point, is small, then we can approximate the potential by a series expansion about $x = x_e$:

$$V(x) = V(x_e) + (x - x_e)\left(\frac{\mathrm{d}V}{\mathrm{d}x}\right)_{x=x_e} + \frac{1}{2}(x - x_e)^2\left(\frac{\mathrm{d}^2V}{\mathrm{d}x^2}\right)_{x=x_e} + \ldots \tag{3.53}$$

in which the subscript $x = x_e$ means that the quantity is to be evaluated at point x_e. Now, $(\mathrm{d}V/\mathrm{d}x)_{x=x_e} = 0$. And for the configuration near x_e, the x^2 term, although itself small, is still very large compared with the terms in x^3, x^4,… that follow it. We may neglect these latter terms altogether. And so Equation 3.53 reduces to

$$V(x) - V(x_e) = \frac{1}{2}(x - x_e)^2\left(\frac{\mathrm{d}^2V}{\mathrm{d}x^2}\right)_{x=x_e}$$

where the value of $(\mathrm{d}^2V/\mathrm{d}x^2)_{x=x_e}$ may be either positive or negative. If it is positive, then $V(x) - V(x_e)$ is positive, whatever the value of x, so that $V(x)$ is minimum at x. If a particle is situated at x_e and experiences a slight disturbance, it will return to x_e when the disturbance is removed. Equilibrium of this kind is called stable equilibrium. On the other hand, if $(\mathrm{d}^2V/\mathrm{d}x^2)_{x=x_e}$ is negative, then $V(x) < V(x_e)$. If the particle is slightly displaced from x_e, it will move away from its original position of equilibrium when the disturbance is removed. Equilibrium of this kind is, therefore, unstable. Accordingly, points x_4 and x_5 in Figure 3.12 are stable equilibrium points, and point x_9 is an unstable equilibrium point.

Example 3.6

A uniform rod *AB* of weight *W* can turn freely around a horizontal axis through one end *A*. A fine cord is attached to point *C* vertically above *A,* passes through an eyelet fixed on the rod at end *B*, and has a hanging weight W_1 suspended from its other end. Prove that, in the absence of friction, the red will be in equilibrium when *BC* is equal to $W_1 \cdot AC/(W/2 + W_1)$. Also, prove that this position of equilibrium is unstable.

Solution:

As shown in Figure 3.13, *a* is the vertical distance between the points *A* and *C*; the length of the fine cord is *l*. At equilibrium, the rod makes an angle θ with the vertical axis. If the length of the rod *AB* is 2*b*, then we have, by law of cosine,

$$x^2 = a^2 + (2b)^2 - 2a(2b)\cos\theta$$

or

$$\cos\theta = \frac{a^2 + 4b^2 - x^2}{4ab}.$$

The potential energy of the system is

$$\begin{aligned} V &= Wb\cos\theta + W_1[2b\cos\theta - (l - x)] \\ &= \frac{W(a^2 + 4b^2 - x^2)}{4a} + W_1\left[\frac{a^2 + 4b^2 - x^2}{2a} - l + x\right]. \end{aligned}$$

Equilibrium points occur where d*V*/d*x* = 0:

$$\frac{\mathrm{d}V}{\mathrm{d}x} = -\frac{Wx}{2a} - \frac{W_1 x}{a} + W_1 = 0\cdot$$

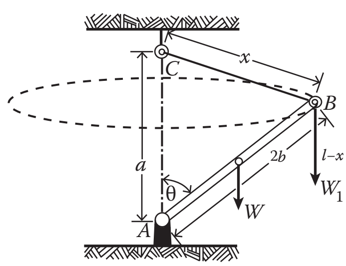


FIGURE 3.13 Freely turning rod.

Thus, there is only one equilibrium point, at

$$x = \frac{aW_1}{W_1 + W/2}.$$

Because $d^2V/dx^2 = -(W/2a + W_1/a) < 0$, the equilibrium is unstable.

3.5.3 Critical and Neutral Equilibrium

If $(d^2V/dx^2)_{x=x_e} = 0$ at a position of equilibrium, the equilibrium is called critical. To investigate the behavior of a particle when it is slightly displaced from the critical equilibrium position, we need to include higher-order terms in the $V(x)$ series expansion:

$$V(x) - V(x_e) = \frac{(x-x_e)^3}{3!}\left(\frac{d^3V}{dx^3}\right)_{x=x_e} + \frac{(x-x_e)^4}{4!}\left(\frac{d^4V}{dx^4}\right)_{x=x_e} + \ldots$$

in which the most important term in the value of $V(x) - V(x_e)$ that is $(x - x_e)^3$. So we shall take it as our next approximation to $V(x) - V(x_e)$. If $(d^3V/dx^3)_{x=x_e} > 0$, $V(x) - V(x_e)$ is positive or negative according to whether $(x - x_e)$ is positive or negative. If $(d^3V/dx^3)_{x=x_e} < 0$, $V(x) - V(x_e)$ is positive or negative according to whether $(x - x_e)$ is negative or positive. Thus, $V(x) - V(x_e)$ changes sign on passing the point $x = x_e$, and $V(x)$ has a point of inflection at $x = x_e$. That is, the equilibrium is unstable (Figure 3.14).

If $(d^3V/dx^3)_{x=x_e}$ is zero, it is necessary to investigate the next higher derivative. In this case, we can write

$$V(x) - V(x_e) = \frac{(x-x_e)^4}{4!}\left(\frac{d^4V}{dx^4}\right)_{x=x_e}.$$

This case can be treated the same as the second derivative, and we find that the equilibrium is stable or unstable according to whether $(d^2V/dx^2)_{x=x_e}$ is positive or negative.

The preceding considerations can be carried to any number of higher-order derivatives, and we easily obtain the following general rules:

1. If the first nonvanishing derivative of $V(x)$ is of odd order, the equilibrium is unstable, regardless of the sign of the derivative.
2. If the first nonvanishing derivative of $V(x)$ is of even order, the equilibrium is stable or unstable according to whether this derivative, evaluated at the equilibrium point, is positive or negative.

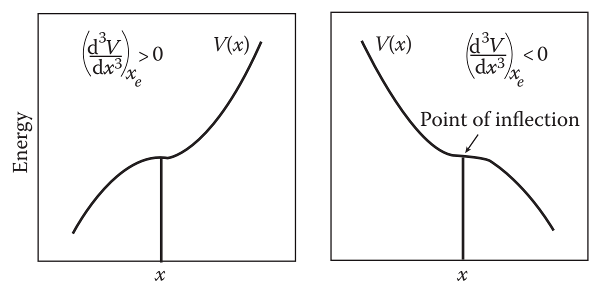


FIGURE 3.14 Unstable equilibrium.

If $V(x)$ is a constant throughout the whole of the range surrounding the configuration under consideration, then the body can be displaced, and there will be no force tending to move it away from its new configuration. Thus, every configuration is one of equilibrium. This kind of equilibrium is called neutral equilibrium.

It should be pointed out that, although, in the case of statically stable systems, the equilibrium configuration is generally associated with potential minima, dynamic stability is not necessarily identified with potential minima. Actually, dynamic stability may occur at potential maxima at fixed potentials other than extrema with continuously varying potential or even without any potential dependence. The treatment of dynamic stability is beyond the syllabus of this text.

We have considered some examples in which the force acting on a particle is a function of one of the following: time, velocity, and position. Problems in which the force depends on two or even all three of the variables just mentioned are more complicated and difficult to solve. No general simplification is possible, and the method of solution of the differential equation of motion will vary from one problem to another. For rectilinear motion, however, some simple methods may be applied:

1. $\vec{F} = \vec{F}(\dot{x}, t)$, that is, when the force function is independent of position.
   In this case, we can try to put

   $$v = \dot{x}$$

   then write the equation of motion as a first-order differential equation:

   $$m\ \mathrm{d}v/\mathrm{d}t = F(\dot{x}, t)$$

   which may be possible to solve analytically.
2. $\vec{F} = \vec{F}(x, \dot{x})$, that is, when the force function is independent of time.
   In this case, we may also try

   $$v = \dot{x}$$

   then write acceleration as

   $$\ddot{x} = v\frac{\mathrm{d}v}{\mathrm{d}x},$$

   and the equation of motion as a first-order differential equation:

   $$mv\frac{\mathrm{d}v}{\mathrm{d}x} = F(x, v)$$

   which may be solved analytically.

3.6 TIME-VARYING MASS SYSTEM (ROCKET SYSTEM)

As a final example of the integration of Newton's equation of motion, we shall study the motion of a variable-mass system. Thus far, we have assumed the particle mass to remain constant. This assumption was based on the fact that in most problems in classical mechanics, speeds are but a small fraction of the speed of light, so the relativistic effect can be ignored. But a particle may change its mass with time, such as a raindrop losing mass by evaporation or gaining mass by condensation, or a rocket in the short time interval during which its fuel is being consumed.

Because of its special interest, we choose to study the motion of the rocket. During the time of propellant, the aerodynamic drag is negligible compared with the force resulting from the rocket motor. We consider here the rocket as a particle losing mass. A more realistic approach is to treat the rocket as a rigid body and pay attention to its angular orientation because the propellant force acts along the axis of the projectile. We will revisit the motion of the rocket in Chapter 12 on rigid bodies.

Consider the rocket at time t. Let M be the mass of the rocket and its fuel at this time, and its velocity relative to a fixed, inertial coordinate system is $\vec{v}$. Between t and $t + \Delta t$, a small amount of fuel Δm is burned and ejected as gas with velocity $\vec{u}$ relative to the rocket, and the resultant change in linear momentum provides the thrust to drive the rocket forward. The velocity $\vec{u}$ is determined by the nature of the propellants, the throttling of the engine, etc., but it is independent of the velocity of the rocket. At time $t + \Delta t$, the velocity of the rocket is $\vec{v} + \Delta\vec{v}$; the velocity of the exhaust relative to the fixed coordinate system is $\vec{v} + \Delta\vec{v} - \vec{u}$. The initial momentum is $\vec{P}(t) = M\vec{v}$, and the final momentum is

$$\vec{P}(t + \Delta t) = (M - \Delta m)(\vec{v} + \Delta\vec{v}) + \Delta m(\vec{v} + \Delta\vec{v} - \vec{u}).$$

Thus, the change in momentum is

$$\begin{aligned}\Delta\vec{P} = \vec{P}(t + \Delta t) - \vec{P}(t) &= (M - \Delta m)(\vec{v} + \Delta\vec{v}) \\ &+ \Delta m(\vec{v} + \Delta\vec{v} - \vec{u}) - M\vec{v} = M\vec{v} - \Delta m\vec{u}.\end{aligned}$$

From this, we have

$$\frac{\mathrm{d}\vec{P}}{\mathrm{d}t} = \lim_{\Delta t \to 0} \frac{\Delta\vec{P}}{\Delta t} = M\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} - \vec{u}\frac{\mathrm{d}m}{\mathrm{d}t}.$$

Equating this rate of change of momentum with the external force $\vec{F}$ on the system, we obtain the equation of motion

$$\vec{F} = M\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} - \vec{u}\frac{\mathrm{d}m}{\mathrm{d}t}$$

or

$$M\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \vec{F} + \vec{u}\frac{\mathrm{d}m}{\mathrm{d}t}.$$

It is inconvenient to have both m and M in the equation of motion. We can eliminate one of them. Note that $\mathrm{d}m/\mathrm{d}t$ is the rate of increase in the exhaust mass, and because this mass comes from the rocket,

$$\mathrm{d}m/\mathrm{d}t = -\mathrm{d}M/\mathrm{d}t.$$

Substituting this into the equation of motion, we obtain

$$M\frac{\mathrm{d}\vec{v}}{\mathrm{d}t} = \vec{F} - \vec{u}\frac{\mathrm{d}M}{\mathrm{d}t}. \tag{3.54}$$

We now solve this equation for two different cases: the rocket in free space and the rocket in a gravitational field.

Case a: The rocket in free space. In free space, there is no external force on a rocket, and its equation of motion is given by

$$M\frac{d\vec{v}}{dt} = -\vec{u}\frac{dM}{dt}$$

or

$$\frac{d\vec{v}}{dt} = -\frac{\vec{u}}{M}\frac{dM}{dt}.$$

Generally, the exhaust velocity is constant; it depends on the type of fuel that is burned. Then, integration of the equation of motion with respect to t gives

$$\int_{t_0}^{t_f}\frac{d\vec{v}}{dt}dt = -\vec{u}\int_{t_0}^{t_f}\frac{1}{M}\frac{dM}{dt}dt = -\vec{u}\int_{M_0}^{M_f}\frac{dM}{M}$$

or

$$\vec{v}_f - \vec{v}_0 = -\vec{u}\ln\frac{M_f}{M_0}$$

where subscripts i and f represent initial and final values, respectively. The exhaust velocity and the ratio of initial and final mass determine the final velocity of the rocket. It is independent of how the mass is released; the fuel can be expended rapidly or slowly without affecting $\vec{v}_f$.

Case b: The rocket in a gravitational field.

For a rocket fired vertically upward from the surface of Earth (Equation 3.54), the equation of motion becomes

$$M\frac{dv}{dt} = -Mg - u\frac{dM}{dt}$$

or

$$\frac{dv}{dt} = -g - \frac{u}{M}\frac{dM}{dt}$$

where we have select up is positive, and both u and g are directed down and are assumed to be constant. Integrating with respect to time, we obtain

$$\int_{t_0}^{t_f}\frac{dv}{dt}dt = -g(t_f - t_0) - u\int_{t_0}^{t_f}\frac{1}{M}\frac{dM}{dt}dt$$

or

$$v_f = u\ln\frac{M_0}{M_f} - gt_f$$

where we have applied the initial condition $v_0 = 0$ at $t_0 = 0$. It is easy to notice that the shorter the burn time, the greater the velocity. Thus, a rocket burns its fuel as quickly as possible during its takeoff.

PROBLEMS

1. A particle of mass m, which at time $t = 0$ has position x_0 and velocity v_0, is being acted upon by a sinusoidal force

$$F = F_0 \sin(\omega t - \phi)$$

   where F_0, ω, and ϕ are constants. Find the equations for position and velocity for all positive time t.
2. A particle of mass m is whirled on the end of a string of length R. The motion is in a vertical plane in the Earth's gravitational field. The instantaneous speed is v when the string makes angle θ with the horizontal. Find the tension T in the string and the tangential acceleration at this instant.
3. A uniform rope of length L is pulled by gravity from a smooth table. Assuming that a length L initially hangs over the table and that the rope starts from rest, find x as a function of t for $L < x < L$, where x represents the distance from the lower end of the rope to the tabletop. The height of the table is H.
4. A mass M hangs from a string of length L that is attached to a rod rotating at constant angular frequency ω as shown in Figure 3.15. The mass moves with a steady speed in a circular path of constant radius. Find α, the angle the string makes with the vertical axis.
5. A block A of mass M_A on a frictionless table is connected to block B of mass M_B hanging beneath the table by a string of negligible mass that passes through a hole in the center of the table. Initially, B is held stationary, and A rotates at constant radius r_0 with steady angular velocity ω_0. If B is released at $t = 0$, what is its acceleration immediately afterward?
6. A particle that is made to move along the x-axis is subject to a restoring force $-kx$ and a constant force F. Discuss its motion, and find the frequency and position of the point of equilibrium.
7. A body of mass m released with velocity v_0 in a viscous fluid is retarded by a force Cv. Find the motion, assuming no other forces are acting on it.

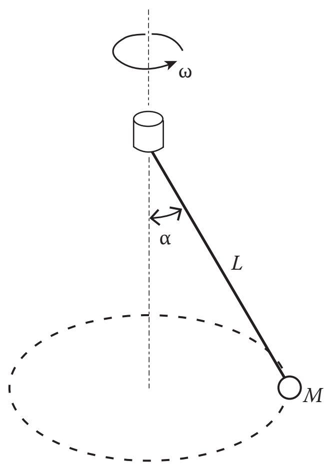


FIGURE 3.15 Conical pendulum.

8. A shot-putter at the top of a cliff of height h above sea level throws a shot at speed v and angle of elevation α to the horizontal axis. Find the range R (at ground level) and α in terms of v. Air resistance is negligible.
9. Two identical blocks A and B, both of mass M, are connected by a spring of length L and spring constant k. Initially, they are at rest on a frictionless straight track. At $t = 0$, block A is hit sharply, giving it an instantaneous velocity v_0 to the right. Find the velocities for subsequent times.
10. According to Yakawa's theory of nuclear forces, the attractive force between two nucleons has a potential of the form

$$V(r) = \frac{Ke^{-\alpha r}}{r}, \quad K < 0, \quad \alpha > 0$$

    (a) Find the force.
    (b) Discuss the types of motion that are possible for mass m under such a force.
    (c) Find the angular momentum L and total energy E for motion on a circle of radius a.
    (d) Find the period of circular motion and the period of small radial oscillations.
11. Using Green's function method, solve the differential equation of the motion of a particle of mass M moving in a resistive medium of coefficient R under the influence of an external force $F(t)$.
12. A force $F = a - 2bx$ acts upon a particle of mass M, where a and b are constants.
    (a) Find the potential energy $V(x)$.
    (b) Make plots of $F(x)$ and $V(x)$.
    (c) Discuss the motion of the particle for different values of energy.
13. A particle of mass m and charge q is moving with velocity V in a constant uniform magnetic field $\boldsymbol{B}$. If we choose the magnetic field $\boldsymbol{B}$ parallel to the OZ-axis, and the initial velocity of the particle is given by $\boldsymbol{V}o = (0, U, V)$, then show that
    (a) $V_x = U \sin \omega t, \quad V_y = U \cos \omega t, \quad V_z = V$
    (b) $x = x_0 + (U/\omega)(1 - \cos \omega t)$, $y = y_0 + (U/\omega)\sin \omega t$, $z = z_0 + Vt$
    where x_0, y_0, and z_0 are the initial positions of the particle along the x-, y-, and z-axes, respectively.
    (c) Also show that the projection of the path on the xy-plane is an equation of circle given by

$$(x - x_0 - U/\omega)^2 + (y - y_0)^2 = U^2/\omega^2.$$

    The particle, thus, moves on the surface of the right circular cylinder whose cross section is the circle represented by the preceding equation.
14. A teeter toy consists of two identical weights that hang from a peg on drooping arms as shown in Figure 3.16. The arrangement is unexpectedly stable—the toy can be spun or

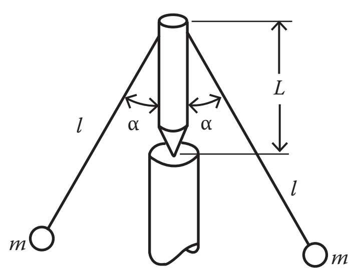


FIGURE 3.16 Teeter toy.

rocked with little danger of toppling over. Analyze why this is so by looking at its potential energy.

15. A uniform solid cube with sides $2a$ rests in the equilibrium position on the top of a cylindrical log of radius r. The plane of one side of the cube is normal to the axis of the log. Determine whether or not the equilibrium is stable. Assume perfectly rough contact.
16. A force causes a particle of mass m to make a rectilinear, simple harmonic motion toward 0. The force is of magnitude $m\omega^2 x$; x is the particle's displacement from 0. When passing through 0, the particle's velocity is V, and when its velocity has become $V/2$ in the same direction, an impulse I is applied to the particle in the direction of motion. Assuming the same force law, find the time and total distance traveled from 0 to the first position of instantaneous rest.
17. A spherical raindrop falling through fog accumulates mass at a rate proportional to its velocity and cross-sectional area. Find the acceleration of the raindrop in terms of its radius and velocity. Assume that the raindrop starts from rest and is infinitely small.
18. A particle is slightly displaced from rest in its position of equilibrium at the highest point of a smooth, fixed sphere of radius a. Determine the point where the particle leaves the sphere.
19. A light string is wrapped round a rough circular cylinder and is in a plane at right angles to the cylinder's generators. Show that when the string is at the point of losing its tension, the two ends are in the ratio $\exp(f_i\theta)$, where θ is the angle between the normals at the two ends of the string that are in contact with the cylinder.
20. A loaded spring gun, initially at rest on a horizontal frictionless surface, fires a marble at an angle of elevation e. The mass of the gun is M, the mass of the marble is m, and the muzzle velocity of the marble is v_0. What is the final motion of the gun? *Hint:* This problem can be solved using Newton's laws directly. But by using conservation of momentum, you can find the final motion of the system in a few steps.
21. Sand falls from a stationary hopper onto a freight car that is moving with uniform velocity v. The sand falls at the rate $\mathrm{d}m/\mathrm{d}t$. How much force is needed to keep the freight car moving at the speed v?
22. Two atoms of masses m_1 and m_2 are bound together in a molecule with energy so low that their separation is always close to the equilibrium value r_0. With the parabola approximation, the effective spring constant is

$$k = \left.\frac{\mathrm{d}^2 U}{\mathrm{d}r^2}\right|_{r_0}.$$

Find the vibration frequency of the molecule.

4 Lagrangian Formulation of Mechanics

Descriptions of Motion in Configuration Space

We have gained considerable experience in setting up Newton's equations of motion in a variety of problems. If the system is not subject to external constraints, the equations of motion are usually easy to set up in Cartesian coordinates. If either the system is subject to external constraints or Cartesian coordinates are not used, then the equations of motion may be difficult to solve or even to formulate. Lagrange found a way to circumvent this problem by the use of generalized coordinates q_i (to be defined soon). In terms of these generalized coordinates, we could write the equations of motion in a form that is equally suitable for all coordinates. Furthermore, the introduction of generalized coordinates can take advantage of constraints on a dynamic system. It is generally a much more pressing matter to take care of the constraints imposed on the motion of the dynamic system. The existence of constraints gives rise to two difficulties. First, the coordinates of the dynamic system are connected by the equations of constraints, so they are not all independent. As a result, the equations of motion are also not all independent. Second, the forces of the constraints are usually very complex or unknown. Thus, we may find ourselves unable to write the equations of motion. In order to circumvent these difficulties, alternative formulations to Newtonian theory have been developed. Each is based on the idea of energy and is so constructed that Newtonian theory may be recovered from it. Moreover, each is expressed in terms of generalized coordinates. This chapter presents a discussion on Lagrangian dynamics, which were developed by Joseph-Louis Lagrange (1736–1813) and others.

4.1 GENERALIZED COORDINATES AND CONSTRAINTS

4.1.1 Generalized Coordinates

Any convenient set of parameters or quantities that can be used to specify the configuration or state of the system can be used as generalized coordinates. That is, the generalized coordinates can be any quantities that can be observed to change with the motion of the system, and they need not be geometrical quantities (lengths or angles). In suitable circumstances, they can be electric currents (see Example 5.8). We shall write the generalized coordinates as q_i, $i = 1,2,3,\ldots, n$. It should be emphasized that we have generalized coordinates but not a generalized coordinate system.

Just as with the nomenclature for rectangular coordinates, we call $\dot{q}_i$ the time derivative of q_i, the generalized velocities corresponding to the generalized coordinate q_i.

4.1.2 Degrees of Freedom

An important characteristic of a given mechanical system is its number of degrees of freedom. The number of degrees of freedom is the smallest number of coordinates required to specify completely the configuration or state of the system. Thus, for a free particle, it is 3. For a system of N particles that is free from constraints, a total of $3N$ coordinates q_i ($i = 1, 2,\ldots, 3N$) is required to describe its configuration completely, that is, the system has $3N$ degrees of freedom. But if the N particle system is subject to k constraints, the number of independent coordinates for the system is reduced to $S = 3N - k$, and the system is said to have S degrees of freedom. A set of n independent generalized coordinates whose number equals the number of degrees of freedom and that are not restricted by the constraints is called a proper set of generalized coordinates.

The idea of generalized coordinates may be summarized below:

(1) They are general in the sense that they need not be lengths or angles in particular.
(2) They are in number just equal to the number of degrees of freedom of the system.
(3) They are independent of one another.

It is obvious that various sets of coordinates could be used to specify the configuration or state of a given system. These sets do not necessarily have the same number of coordinates nor do they have the same number of constraints. But different sets of coordinates will be functionally related, and we can obtain one set of coordinates from the other by a coordinate transformation. If $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N$ and $q_1, q_2, \ldots, q_n$ are two sets of coordinates for a system of N particles, we will assume that the transformation equations are of the form

$$\vec{r}_j = \vec{r}_j = \left(q_1, q_2, \ldots q_n, t\right), \quad j = 1, 2, \ldots, N. \tag{4.1}$$

The time t is included here expressly for the case of moving coordinates. If the Jacobian determinant of the transformation is not zero, we can solve Equation 4.1 for the q's as functions of the $\vec{r}$'s and t.

4.1.3 Configuration Space

We have seen that the configuration of a system can be specified by the values of the n independent generalized coordinates $q_1, q_2, \ldots, q_n$. It is convenient to think of these n numbers as the coordinates of a single point in an n-dimensional space where the q's form the n coordinate axes. This n-dimensional space is known as configuration space. We can consider a vector q to be drawn from the origin to the given configuration point. The evolution of the system in time is given by a curve (the path of motion) in this configuration space. Each point on the curve represents the entire system configuration at some given instant of time. As the generalized coordinates are not necessarily position coordinates, configuration space is not necessarily connected to the physical three-dimensional space, and the path of motion in configuration space does not necessarily resemble the path in space of any actual particle.

4.1.4 Constraints

When the motion of a dynamic system is not permitted to extend freely in three dimensions, the system is said to be subject to constraints. The usual constraints can be classified as the following four types, but we will concentrate on holonomic constraints.

4.1.4.1 Holonomic and Nonholonomic Constraints

Suppose the configuration of a dynamic system is specified by the n generalized coordinates $q_1, q_2, \ldots, q_n$. If the conditions of the constraint imposed on the motion of the system can be expressed as equations connecting the coordinates of the dynamic system and the time in the following form:

$$f_j(q_1, q_2, \ldots, q_n, t) = 0, \quad j = 1, 2, \ldots, k \tag{4.2}$$

then the constraints are said to be holonomic. For example, a bead constrained to slide on a circular wire of radius a in the xy-plane must have its x- and y-coordinates satisfy the following condition:

$$x^2 + y^2 = a^2$$

which is the same form as Equation 4.2. Holonomic constraints are also known as integrable constraints. The term integrable is employed here because Equation 4.2 is equivalent to a differential equation:

$$\sum_{k=1}^{n} \frac{\partial f_j}{\partial q_k} \mathrm{d}q_k = 0 \tag{4.3}$$

which is readily integrated into Equation 4.2. For the sliding bead, the differential relationship $x\mathrm{d}x + y\mathrm{d}y = 0$ shows that the coordinates x and y cannot be independently varied. This expression integrates into $x^2 + y^2 = a^2$ easily.

A dynamic system whose constraint equations are all of the holonomic form given in Equation 4.2 is known as a holonomic system.

Constraints not expressible in the form of Equation 4.2 are said to be nonholonomic. In general, a system of m constraints, which are written as nonintegrable differential expressions of the form

$$\sum_{k=1}^{n} a_{jk}\mathrm{d}q_k + a_{jt}\mathrm{d}t = 0,\; j = 1,2,\ldots,m \tag{4.4}$$

are nonholonomic constraints. Here, the a's are, in general, functions of the q's and t. We cannot integrate these differential equations to obtain functions of the form given in Equation 4.2, nor is it possible to find a set of independent generalized coordinates. Hence, nonholonomic dynamic systems are always described by more coordinates than there are degrees of freedom. For example, consider the rolling of a vertical disk of radius a (Figure 4.1). The contact point P of the disk moves along a curve C in a horizontal plane defined by OX and OY. Let θ be the angle that the tangent at $P(x, y)$ to C makes with OX, and let ϕ be the angular rotation of the disk when P has traveled a distance s along C; then the four coordinates (x, y, θ, ϕ) specify the position of the disk at any instant. Now the position of the disk is varied from the state defined by (x, y, θ, ϕ) to $(x + \mathrm{d}x, y + \mathrm{d}y, \theta + \mathrm{d}\theta, \phi + \mathrm{d}\phi)$. Using $s = a\phi$ so that $\mathrm{d}s = a\mathrm{d}\phi$, we obtain

$$\mathrm{d}x = (a\mathrm{d}\phi)\cos\theta, \text{ and } \mathrm{d}y = (a\mathrm{d}\phi)\sin\theta. \tag{4.4a}$$

These relationships show clearly that the variation $\mathrm{d}x$, $\mathrm{d}y$ in x and y depend upon the variation $\mathrm{d}\phi$ in ϕ. Thus, the system (x, y, θ, ϕ) is nonholonomic. Equation 4.4a is not integrable because ϕ is

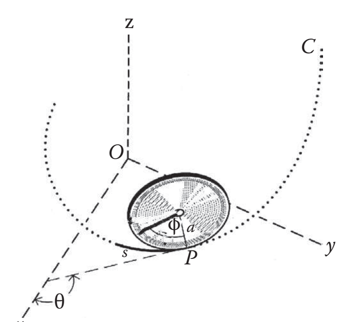


FIGURE 4.1 Vertical disk rolling on a horizontal plane.

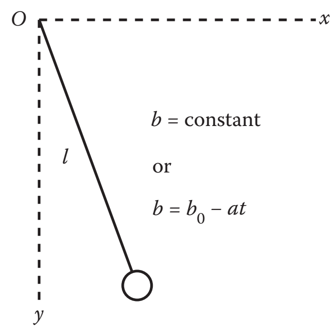


FIGURE 4.2 Simple pendulum.

arbitrary. Because nonholonomic constraints may also appear in the form of inequalities, such as a particle moving on the surface of a sphere of radius a, the condition of the constraints is an inequality: $x^2 + y^2 + z^2 \geq a^2$.

A dynamic system that is subject to one or more nonholonomic constraints is called a nonholonomic dynamic system.

4.1.4.2 Scleronomic and Rheonomic Constraints

A scleronomic constraint is one that is independent of time, whereas a rheonomic constraint contains time explicitly. For example, a pendulum with an inextensible cord is scleronomic, and the condition of the constraints is $x^2 + y^2 = b^2$, where x and y are the coordinates of the pendulum bob, and b is the length of the cord. On the other hand, a pendulum with an extensible cord is rheonomic; the condition of constraint in this case is $x^2 + y^2 = (b_0 - at)^2$ where b_0 is the length of the cord at the time $t = 0$, and a is the rate of increase of the cord (Figure 4.2).

For holonomic constraints, the difficulty of the non-independent coordinates mentioned earlier can be handled by using the equations of constraint to eliminate variables from the equations of motion. But this procedure often involves algebraic difficulties and is therefore rarely used. Instead, we bypass the difficulty of non-independent coordinates by introducing independent generalized coordinates. For nonholonomic constraints, the equations of constraint cannot be used to eliminate the extra dependent coordinates. Because there is no general way to treat nonholonomic systems, we must consider each problem separately.

The second difficulty, which is caused by the forces of constraint not given at the start, is still present even though we are using generalized coordinates. To overcome the difficulty, we shall formulate dynamics in the Lagrangian form, in which the unknown forces of constraint do not appear but can be obtained from the solution we are looking for.

Having discussed the four constraints, we shall concentrate our major effort on holonomic systems because virtually all the aspects of dynamics that are important in the study of atomic and molecular properties involve only holonomic systems.

4.2 KINETIC ENERGY IN GENERALIZED COORDINATES

The kinetic energy T is considered to be a function of the generalized coordinates and velocities $\left(q_i, \dot{q}_i\right)$ and possibly time t; we can show this explicitly. In fixed, rectangular coordinates, the kinetic energy T is a homogeneous, quadratic function of the velocities $\dot{\vec{r}}_i$:

$$T = \frac{1}{2}\sum_{i}^{N} m_i \dot{r}_i^2.$$

From Equation 4.1, we have

$$\dot{\vec{r}}_i = \sum_j^n \frac{\partial \vec{r}_i}{\partial q_j}\dot{q}_j + \frac{\partial \vec{r}_i}{\partial t}.$$

Note that $\dot{\vec{r}}_i$ is linear in the $\dot{q}$'s and that $\partial r_i/\partial q_j$ and $\partial r_i/\partial t$ are functions of the q's and t.

Evaluating the square of $\vec{r}_i$, we obtain

$$\dot{r}_i^2 = \dot{\vec{r}}_i \cdot \dot{\vec{r}}_i = \left(\sum_j^n \frac{\partial \vec{r}_i}{\partial q_j}\dot{q}_j + \frac{\partial \vec{r}_i}{\partial t}\right)\cdot\left(\sum_k^n \frac{\partial \vec{r}_i}{\partial q_k}\dot{q}_k + \frac{\partial \vec{r}_i}{\partial t}\right)$$

$$= \sum_{j,k}^n \frac{\partial \vec{r}_i}{\partial q_j}\cdot\frac{\partial \vec{r}_i}{\partial q_k}\dot{q}_j\dot{q}_k + 2\sum_j^n \frac{\partial \vec{r}_i}{\partial q_j}\cdot\frac{\partial \vec{r}_i}{\partial t} + \frac{\partial \vec{r}_i}{\partial t}\cdot\frac{\partial \vec{r}_i}{\partial t}.$$

The kinetic energy T becomes

$$T = \sum_{j,k}^n a_{jk}\dot{q}_j\dot{q}_k + \sum_j^n b_j\dot{q}_j + c \tag{4.5}$$

where a_{jk}, b_j, and c are functions of the q's and t:

$$a_{jk} = \sum_{i=1}^n \frac{1}{2}m_i \frac{\partial \vec{r}_i}{\partial q_j}\cdot\frac{\partial \vec{r}_i}{\partial q_k}, \quad b_j = \sum_{i=1}^n m_i \frac{\partial \vec{r}_i}{\partial q_j}\cdot\frac{\partial \vec{r}_i}{\partial t}, \quad c = \sum_{i=1}^n \frac{1}{2}m_i \frac{\partial \vec{r}_i}{\partial t}\cdot\frac{\partial \vec{r}_i}{\partial t}.$$

For a fixed or scleronomic system, the time t does not appear explicitly in the transformation equation 4.1, so $b_j = c = 0$, and Equation 4.5 reduces to

$$T = \sum_{j,k}^n a_{jk}\dot{q}_j\dot{q}_k. \tag{4.5a}$$

That is, the kinetic energy T is, as in rectangular coordinates, a homogeneous, quadratic function of the generalized velocities.

Differentiating T with respect to $\dot{q}_i$, we obtain

$$\frac{\partial T}{\partial \dot{q}_i} = \sum_{j,k}^n \left(a_{jk}\dot{q}_j \frac{\partial \dot{q}_k}{\partial \dot{q}_i} + a_{jk}\frac{\partial \dot{q}_j}{\partial \dot{q}_i}\dot{q}_k\right)$$

$$= \sum_{j,k}^n \left(a_{jk}\dot{q}_j\delta_{ki} + a_{jk}\delta_{ji}\dot{q}_k\right) = \sum_{j=1}^n a_{ji}\dot{q}_j + \sum_{k=1}^n a_{ik}\dot{q}_k.$$

Multiplying this equation by $\dot{q}_i$ and summing over I, we obtain

$$\sum_{i=1}^n \dot{q}_i \frac{\partial T}{\partial \dot{q}_i} = \sum_{i,k=1}^n a_{ik}\dot{q}_i\dot{q}_k + \sum_{i,j=1}^n a_{ji}\dot{q}_j\dot{q}_i.$$

Now, all of the indices, i, j, and k are dummies. If we change i to j in the first term on the right-hand side and i to k in the second term, then both terms become identical, and we have

$$\sum_{i=1}^{n} \dot{q}_i \frac{\partial T}{\partial \dot{q}_i} = 2\sum_{j,k=1}^{n} a_{jk}\dot{q}_j\dot{q}_k = 2T. \tag{4.6}$$

This is a special case of Euler's theorem, which states that if $f(x_i)$ is a homogeneous function of the x_i of order n, that is,

$$f(\lambda x_i) = \lambda^n f(x_i)$$

then

$$\sum_i x_i \frac{\partial f}{\partial x_i} = nf.$$

As an illustration of Euler's theorem, consider the motion of a particle of mass m under the influence of a force derived from a potential dependent on position only. The kinetic energy T in Cartesian coordinates is

$$T = \frac{1}{2}m\sum_{i=1}^{3} \dot{x}_i^2.$$

Obviously, T is a homogeneous function of x_i and $n = 2$. According to Euler's theorem, we then have

$$\sum_{i=1}^{3} \dot{x}_i \frac{\partial T}{\partial \dot{x}_i} = 2T.$$

This can be verified by direct computation.

4.3 GENERALIZED MOMENTUM

Let us revisit the particle of the above illustrative example again, if this particle moves in a potential field V, we can introduce a new function, denoted by L, to characterize it; this new function L is called the Lagrange function or Lagrangian, and is defined as:

$$L = T - V = \frac{1}{2}m\sum_{i=1}^{3} \dot{x}_i^2 - V.$$

Taking the partial derivative with respect to $\dot{x}_i$ we obtain

$$\frac{\partial L}{\partial \dot{x}_i} = m\dot{x}_i = p_i$$

which is the x_i component of the linear momentum of the particle. This result suggests an extension of the concept of momentum. The generalized momentum p_i corresponding to q_i shall be defined as

$$p_i = \partial L/\partial \dot{q}. \tag{4.7}$$

It is, in general, a function of the q's, $\dot{q}$'s, and t. Note, however, that as the Lagrangian is, at most, quadratic in the $\dot{q}$'s, p_i is a linear function of the $\dot{q}$'s. It is evident that if q_i is not a Cartesian coordinate, p_i does not necessarily have the dimension of linear momentum. Also, if there is a velocity-dependent potential, then even with a Cartesian coordinate q_i and the corresponding generalized momentum p_i, as defined earlier, it will not be identical to the usual mechanical momentum.

4.4 LAGRANGIAN EQUATIONS OF MOTION

We now proceed to derive Lagrange's equations of motion, which are the bedrock of the Lagrangian dynamics. Lagrange's equations can be obtained either from a differential principle or from an integral principle. Both approaches involve only work and energy.

In the differential principle approach, we consider the instantaneous state of the system and a small virtual displacement about the instantaneous state. J. Bernoulli and D'Alembert originally introduced the concepts of virtual displacement, virtual work, and inertial forces. Lagrange cast these ideas in a form especially applicable in dynamic problems. We shall not follow this approach to derive Lagrange's equations of motion in the main text. For the interested readers, we listed the derivation in an appendix.

In the integral principle approach, we adopt a global view, that is, what can we say about the whole path of our system? The actual path of a dynamic system differs in a remarkable way from other paths taking the same time. There is a certain quantity that, if integrated with respect to time over the path, gives a result that has a stationary (minimum or maximum) value for the actual path. For conservative systems, the certain quantity is simply the Lagrangian of the system! The integral is known as the action, so what we are talking about is the principle of least action, commonly known as Hamilton's principle, introduced in 1833 by William R. Hamilton.

4.4.1 HAMILTON'S PRINCIPLE

To formulate Hamilton's principle, we start with the action or action integral

$$I = \int_{t_1}^{t_2} L(q_i,\ \dot{q}_i;\ t)\,\mathrm{d}t \tag{4.8}$$

where the Lagrangian L is given, and $q_i(t)$ takes on the prescribed values at t_1 and t_2 but may be arbitrarily varied for values of t between t_1 and t_2.

For each choice of the set of functions $q_i(t)$, Equation 4.8 defines a numerical value for L. Hamilton's principle states that for a conservative, holonomic dynamic system, the motion of the system from its position in configuration space at time t_1 to its position at time t_2 follows a path for which the action integral, Equation 4.8, has a stationary (maximum or minimum) value. In other words, of all possible paths along which a dynamic system might move from one point to another in configuration space within a specified time interval and consistent with the constraints, the path followed by the dynamic system is that where the value of the action is stationary. This means that if a set of functions $q_i(t)$ gives the integral I a minimum or maximum value, then any neighboring set, no matter how close to the set $q_i(t)$, must make I increase (or decrease). In terms of calculus of variations, this means

$$\delta I = \delta \int_{t_1}^{t_2} L(q_i,\ \dot{q}_i;\ t)\,\mathrm{d}t = 0 \tag{4.9}$$

where $q_i(t)$, and hence $\dot{q}_i(t)$, is to be varied subject to $\delta\, q_i(t_1) = \delta\, q_i(t_2) = 0$. The symbol δ refers to the variation in a quantity between two paths while d, as usual, refers to a variation along a path.

4.4.2 Lagrange's Equations of Motion from Hamilton's Principle

The symbol δ in Equation 4.9 refers to the variation of the integrand L between two paths. Two such paths, out of an infinite number of possibilities, are shown in Figure 4.3. We define the variation in q, labeled by δq, to be $q_2(t) - q_1(t)$, that is, to be the change in q from path to path at the same value of t. The slope of the paths in q also is different for the same t. The variation in slope is $\delta\dot{q} = \dot{q}_2(t) - \dot{q}_1(t)$.

It is evident from Figure 4.3 that as t changes from t_1 to t_2, δq will change and so is a function of t, which we will assume to be continuous at all points between $t = t_1$ to $t = t_2$. If we differentiate it with respect to t, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}(\delta q) = \dot{q}_2(t) - \dot{q}_1(t) = \delta\dot{q}.$$

We shall have constant occasion to use this fundamental relationship permitting interchange of the operations of differentiation and variation. In a similar way, it may be shown that the operations of integration and variation commute.

We now describe the difference between paths by introducing a new function $\eta_i(t)$ and a scaling parameter α. The function $\eta_i(t)$ is arbitrary except that $\eta_i(t_1) = \eta_i(t_2) = 0$, and $\eta_i(t)$ must be differentiable. Each possible path can therefore be written as

$$q_i(t,\alpha) = q_i(t,0) + \alpha\eta_i(t) \tag{4.10}$$

where $q_i(t, 0)$ is the actual dynamical path followed by the system (as yet unknown). In terms of the variation symbol δ, we can write

$$\delta q_i = (\partial q_i/\partial\alpha)\mathrm{d}\alpha = \eta_i \mathrm{d}\alpha \tag{4.11}$$

which corresponds to a virtual displacement in q_i from the actual dynamical path to a neighboring varied path.

The action integral I is now a function of α only for any given $\eta_i(t)$:

$$I(\alpha) = \int_{t_1}^{t_2} L\{q_i(t,\alpha),\ \dot{q}_i(t,\alpha);\ t\}\,\mathrm{d}t. \tag{4.12}$$

Hence, the stationary values of $I(\alpha)$ occur when $\partial I/\partial\alpha = 0$. But by our choice of $q_i(t, 0)$, we know that this happens when $\alpha = 0$, so the necessary condition that the action integral has a stationary

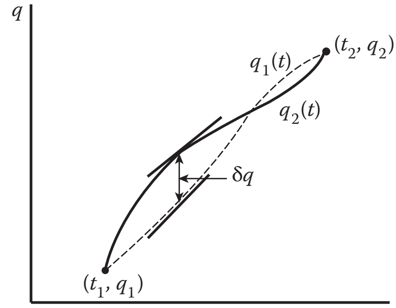


FIGURE 4.3 Variation of a curve between fixed end points.

value is $\partial I/\partial\alpha = 0$ when $\alpha = 0$. In terms of the variation symbol δ, this necessary condition can be written as

$$\delta I = \left(\frac{\partial I}{\partial \alpha}\right)_{\alpha=0} \mathrm{d}\alpha = 0 \tag{4.13}$$

for arbitrary $\eta_i(t)$ and nonzero α, and the subscript $\alpha = 0$ means that we evaluate the derivative $\partial I/\partial\alpha$ at $\alpha = 0$.

We now expand the integrand L in Equation 4.12 in a Taylor's series:

$$I(\alpha) = \int_{t_1}^{t_2} \left(L(q_i(t,0),\ q_i(t,0);\ t) + \alpha\eta_i \frac{\partial L}{\partial q_i} + \alpha\dot{\eta}_i \frac{\partial L}{\partial \dot{q}_i} + O(\alpha^2) \right) \mathrm{d}t.$$

Because the integration limits t_1 and t_2 are not dependent on α, we can differentiate under the integral sign with respect to α and obtain

$$\frac{\partial}{\partial \alpha} I(\alpha) = \int_{t_1}^{t_2} \left(\eta_i \frac{\partial L}{\partial q_i} + \dot{\eta}_i \frac{\partial L}{\partial \dot{q}_i} + O(\alpha^2) \right) \mathrm{d}t. \tag{4.14}$$

Dropping terms in α^2, α^3, and so forth, and integrating by parts, the second term yields

$$\int_{t_1}^{t_2} \frac{\partial L}{\partial \dot{q}_i} \dot{\eta}_i \,\mathrm{d}t = \left[\frac{\partial L}{\partial \dot{q}_i} \eta_i(t) \right]_{t_1}^{t_2} - \int_{t_1}^{t_2} \eta_i(t) \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} \mathrm{d}t. \tag{4.15}$$

The first term on the right-hand side is 0 because $\eta_i(t_1) = \eta_i(t_2) = 0$. Substituting the resulting Equation 4.15 into Equation 4.14, we obtain

$$\left(\frac{\partial I}{\partial \alpha}\right)_{\alpha=0} = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} \right]_{\alpha=0} \eta_i(t)\mathrm{d}t. \tag{4.16}$$

To obtain the stationary condition in terms of the variation symbol δ, we multiply Equation 4.16 through by $\mathrm{d}\alpha$, resulting in

$$\left(\frac{\partial I}{\partial \alpha}\right)_{\alpha=0} \mathrm{d}\alpha = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} \right] \mathrm{d}\alpha\eta_i(t)\mathrm{d}t,$$

or

$$\delta I = \left(\frac{\partial I}{\partial \alpha}\right)_{\alpha=0} \mathrm{d}\alpha = \int_{t_1}^{t_2} \left[\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_i} \right] \delta q_i(t)\mathrm{d}t, \tag{4.17}$$

where δq_i corresponds to the previously defined virtual displacement. Because δq_i is arbitrary, except that $\delta q_i(t_1) = \delta q_i(t_2) = 0$, a necessary condition for $\delta I = 0$ is that the square bracket vanishes, yielding Lagrange's equations of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0, \quad j = 1, 2, \ldots, n. \tag{4.18}$$

This is also a sufficient condition for a stationary value of the action integral I. The sufficient condition is apparent from the fact that Equation 4.18 implies that the integral in Equation 4.17 vanishes, resulting in the variation δI being zero.

If the forces on the system are not conservative, then we have to use the generalized version of Hamilton's principle:

$$\int_{t_1}^{t_2}\left(\delta T+\sum_{j=1}^{n}Q_j\delta q_j)\right)\mathrm{d}t=0. \tag{4.19}$$

The variation of the first integral on the right-hand side can be written as

$$\int_{t_1}^{t_2}\delta T(q_i,\ \dot{q}_i;\ t)\,\mathrm{d}t=\int_{t_1}^{t_2}\sum_i\left[\frac{\partial T}{\partial q_i}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_i}\right]\delta q_i\mathrm{d}t.$$

Substituting this into Equation 4.19, the generalized version of Hamilton's principle, we get

$$\int_{t_1}^{t_2}\sum_i\left[\frac{\partial T}{\partial q_i}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_i}+Q_i\right]\delta q_i\mathrm{d}t=0.$$

Because the constraints are holonomic, δq_i are arbitrary and all independent, and we get

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_i}-\frac{\partial T}{\partial q_i}=Q_i,\ I=1,2,\ldots,n. \tag{4.20}$$

If some of the forces acting on the system are conservative, Equation 4.20 will take the following form:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j}-\frac{\partial L}{\partial q_j}=Q'_j,\ j=1,2,\ldots,n \tag{4.21}$$

where L contains the potential of the conservative force, and Q'_j represents the force not arising from a potential.

Lagrange's equations of motion are a set of n second-order differential equations for n unknown functions $q_i(t)$, and the general solution contains $2n$ integration constants. To determine these $2n$ constants and thereby define uniquely the motion of the system, it is necessary to know the state of the system's initial conditions at some given instant, for example, the initial values of all the q's and $\dot{q}$'s.

Example 4.1: Simple Pendulum

Describe the motion of a pendulum bob with the Lagrange method (Figure 4.4).

Solution:

We choose the angle θ made by the string of the pendulum with the vertical axis as the generalized coordinate. The kinetic energy of the bob is then given by

$$T=\frac{1}{2}m\left(\dot{r}^2+r^2\dot{\theta}^2\right)=\frac{1}{2}ml^2\dot{\theta}^2\ (r=l)$$

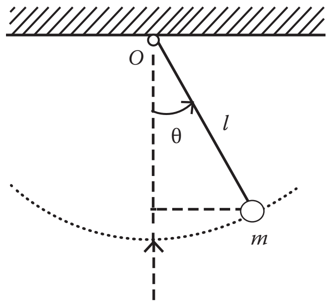


FIGURE 4.4 Simple pendulum.

and the potential energy is

$$V = mg(l - l\cos\theta) = mgl(1 - \cos\theta).$$

Then the Lagrangian L is

$$L = T - V = \frac{1}{2}ml^2\dot{\theta}^2 - mgl(1 - \cos\theta)$$

and the Lagrange's equation is

$$\frac{d}{dt}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = \ddot{\theta} + (g/l)\sin\theta = 0.$$

For a small angle of oscillation, $\sin\theta \cong \theta$, the Lagrange's equation of motion reduces to the familiar harmonic oscillator type of equation

$$\ddot{\theta} + (g/l)\theta = 0.$$

Example 4.2

A bead slides on a smooth wire that is rotating uniformly about a point O in a horizontal plane (the xy-plane). Describe the motion of the bead (Figure 4.5).

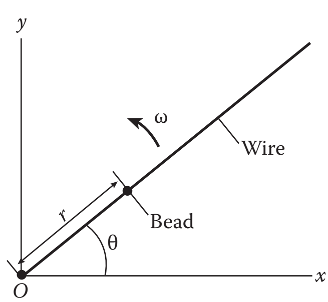


FIGURE 4.5 A bead slides on a smooth wire.

Solution:

The use of Newton's equation of motion (the second law) would be complicated by the constraint of the wire; we therefore use Lagrange's approach. The kinetic energy is

$$T = (m/2)\left(\dot{x}^2 + \dot{y}^2\right) = (m/2)\left(\dot{r}^2 + r^2\dot{\theta}^2\right),$$

and the potential energy V is zero. The constraint is that the angular speed is a constant, $\dot{\theta} = \omega$ (a constant), and so the Lagrangian L is

$$L = T - V = T = \frac{1}{2}m\left(\dot{r}^2 + r^2\omega^2\right).$$

The Lagrange's equation gives

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = m\ddot{r} - mr\omega^2 = 0,$$

or

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

This equation can be integrated by multiplying both sides by the integrating factor $\dot{r}$ to give

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

from which we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\dot{r}^2}{2}\right) = \omega^2\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{r^2}{2}\right)$$

and hence

$$\dot{r}^2 = \sqrt{\omega^2 r^2 + C_1},$$

where C_1 is an integration constant. The last equation can be directly integrated to give

$$r = \sqrt{C_1/\omega^2}\,\sinh(\omega t + C_2).$$

The two integration constants, C_1 and C_2, can be determined by the initial conditions.

Example 4.3: The S.H.O. under Sinusoidal Force

Solution:

There is one generalized coordinate, namely, x; the kinetic and potential energies are given by, respectively,

$$T = m\dot{x}^2/2 \quad \text{and} \quad V = kx^2/2$$

where k is the spring constant of the spring (Figure 4.6). The Lagrangian L of the system is

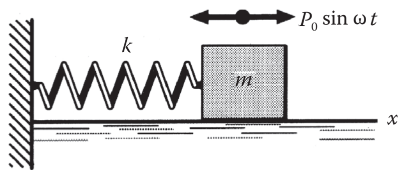


FIGURE 4.6 Force-driven oscillator.

$$L = T - V = m\dot{x}^2/2 - kx^2/2.$$

The applied force is not conservative: $Q' = P_0 \sin \omega t$. Substituting these into Lagrange's Equation 4.21,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = Q'_j$$

we find

$$m\ddot{x} + kx = P_0 \sin \omega t.$$

Example 4.4

A bead of mass m slides freely on a frictionless circular wire of radius b that rotates in a horizontal plane about a point on the circular wire with a constant angular velocity ω. Find Lagrange's equation of motion of the bead. Show that the bead oscillates as a pendulum of length $l = g/\omega^2$.

Solution:

The circular wire rotates in the xy-plane about the point O as shown in Figure 4.7. The rotation is in the counterclockwise direction; C is the center of the circular wire, and the angles θ and φ are as indicated. As the wire rotates counterclockwise with an angular velocity ω, we have $\varphi = \omega t$. Now, the coordinates x and y of the bead are easily seen to be

$$x = b\cos \omega t + b\cos(\theta + \omega t),\ y = b\sin \omega t + b\sin(\theta + \omega t).$$

Thus, the problem is one involving only a single degree of freedom, and the generalized coordinate is θ.

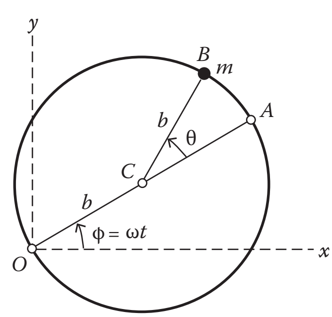


FIGURE 4.7 A bead slides on a circular wire that rotates in a horizontal plane about a point on the circular wire.

The potential energy of the bead can be taken to be zero, and its kinetic energy is

$$T = \frac{1}{2}m\left(\dot{x}^2 + \dot{y}^2\right) = \frac{1}{2}mb^2\left[\omega^2 + \left(\dot{\theta} + \omega\right)^2 + 2\omega\left(\dot{\theta} + \omega\right)\cos\theta\right]$$

which is also the Lagrangian L of the system (the bead). Inserting the L into Lagrange's equation of motion

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = 0,$$

we obtain

$$mb^2\left(\ddot{\theta} - \omega\dot{\theta}\sin\theta\right) + mb^2\omega\left(\dot{\theta} + \omega\right)\sin\theta = 0, \text{ or } \ddot{\theta} + \omega^2\sin\theta = 0.$$

Thus, we see that the bead oscillates about the line OA like a pendulum of length $l = g/\omega^2$.

Example 4.5

A double pendulum consists of two particles suspended by rods of negligible mass. Assuming all motion is in a vertical plane, find Lagrange's equations of motion, and then linearize these equations for small oscillations (Figure 4.8).

Solution:

The coordinates of the two particles are given by

$$x_1 = l_1 \cos\theta_1,\ y_1 = l_1 \sin\theta_1,\ x_2 = l_1 \cos\theta_1 + l_2 \cos\theta_2, \text{ and } y_2 = l_1 \sin\theta_1 + l_2 \sin\theta_2.$$

The kinetic energy of the system is

$$\begin{aligned} T &= \frac{1}{2}m_1\left(\dot{x}_1^2 + \dot{y}_1^2\right) + \frac{1}{2}m_2\left(\dot{x}_2^2 + \dot{y}_2^2\right) \\ &= \frac{1}{2}m_1 l_1^2\dot{\theta}_1^2 + \frac{1}{2}m_2\left[l_1^2\dot{\theta}_1^2 + l_2^2\dot{\theta}_2^2 + 2l_1 l_2\dot{\theta}_1\dot{\theta}_2\cos(\theta_1 - \theta_2)\right]. \end{aligned}$$

Choosing the reference level for potential energy at a distance $l_1 + l_2$ below the point of suspension, we obtain

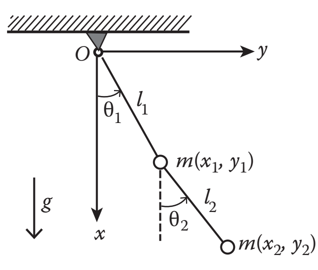


FIGURE 4.8 Double pendulum.

$$V = m_1 g(l_1 + l_2 - l_1 \cos \theta_1) + m_2 g[l_1 + l_2 - (l_1 \cos \theta_1 + l_2 \cos \theta_2)].$$

The Lagrangian L of the system is then given by

$$L = \frac{1}{2} m_1 l_1^2 \dot{\theta}_1^2 + \frac{1}{2} m_2 \left[l_1^2 \dot{\theta}_1^2 + l_2^2 \dot{\theta}_2^2 + 2 l_1 l_2 \dot{\theta}_1 \dot{\theta}_2 \cos(\theta_1 - \theta_2) \right]$$
$$- m_1 g(l_1 + l_2 - l_1 \cos \theta_1) - m_2 g[l_1 + l_2 - (l_1 \cos \theta_1 + l_2 \cos \theta_2)].$$

The problem has two degrees of freedom; θ_1 and θ_2 are the two generalized coordinates. From the two Lagrange's equations

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \theta_i} - \frac{\partial L}{\partial \theta_i} = 0, \quad i = 1, 2$$

we obtain

$$(m_1 + m_2) l_1^2 \ddot{\theta}_1 + m_2 l_1 l_2 \cos(\theta_1 - \theta_2) \ddot{\theta}_2 + m_2 l_1 l_2 \sin(\theta_1 - \theta_2) \dot{\theta}_2^2 = -(m_1 + m_2) g l_1 \sin \theta_1$$

and

$$m_2 l_2^2 \ddot{\theta}_2 + m_2 l_1 l_2 \ddot{\theta}_1 \cos(\theta_1 - \theta_2) - m_2 l_1 l_2 \cos(\theta_1 - \theta_2) \dot{\theta}_1^2 = -m_2 g l_2 \sin \theta_2.$$

When $m_1 = m_2$ and $l_1 = l_2$, the preceding two equations reduce:

$$2 l \ddot{\theta}_1 + l \cos(\theta_1 - \theta_2) \ddot{\theta}_2 + l \sin(\theta_1 - \theta_2) \dot{\theta}_2^2 = -2 g \sin \theta_1$$

and

$$l \cos(\theta_1 - \theta_2) \ddot{\theta}_1 + l \ddot{\theta}_2 - l \sin(\theta_1 - \theta_2) \dot{\theta}_1^2 = -g \sin \theta_2.$$

We now linearize the last two equations for the case of small oscillations. Assuming θ_1, θ_2, and their time derivatives are much smaller than one, then

$$\cos(\theta_2 - \theta_1) \cong 1 \text{ and } \sin(\theta_2 - \theta_1) \cong \theta_2 - \theta_1$$

and the last two equations reduce to

$$2 l \ddot{\theta}_1 + l \ddot{\theta}_2 = -2 g \theta_1 \text{ and}$$
$$l \ddot{\theta}_1 + l \ddot{\theta}_2 = -g \theta_2.$$

In the next example of the use of Lagrange's equations, the introduction of an appropriate coordinate system is dictated by the symmetry of the problem. A constant of the motion can be found easily in this chosen coordinate system.

Example 4.6: The Spherical Pendulum

Consider a particle of mass m attached to a fixed point by a rod of length l with negligible mass. The particle is free to swing in any direction under the action of gravity. Because the particle is constrained to move on the inner surface of a sphere, this system is called a "spherical pendulum." Find its differential equations of motion.

Solution:

Using spherical coordinates as indicated in Figure 4.9, we see that the kinetic energy of the particle is

$$T = \frac{1}{2}m\left(l^2\dot{\theta}^2 + l^2\dot{\phi}^2\sin^2\theta\right).$$

If we take the support point as the reference level for the potential energy, then we have

$$V = mgl\cos\theta$$

and the Lagrangian of the particle is

$$L = \frac{1}{2}m\left(l^2\dot{\theta}^2 + l^2\dot{\phi}^2\sin^2\theta\right) - mgl\cos\theta.$$

There are two degrees of freedom; θ and φ are the two independent generalized coordinates. From Lagrange's equation for coordinate θ, we obtain

$$ml^2\ddot{\theta} - ml^2\dot{\phi}^2\sin\theta\cos\theta - mgl\sin\theta = 0. \tag{4.22}$$

Similarly, we have the equation of motion for coordinate φ:

$$ml^2\ddot{\phi}\sin^2\theta + 2ml^2\dot{\theta}\dot{\phi}\sin\theta\cos\theta = 0. \tag{4.23}$$

Note that these equations are nonlinear, although $\ddot{\theta}$ and $\ddot{\phi}$ appear linearly. The φ equation of motion is immediately integrable because $\partial L/\partial\varphi = 0$, and therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\phi}} = 0.$$

Hence,

$$\partial L/\partial\dot{\phi} = ml^2\dot{\phi}\sin^2\theta = p_\phi$$

where p_φ, the generalized momentum corresponding to φ, is a constant of the motion. It is the angular momentum about a vertical axis through the support point O.

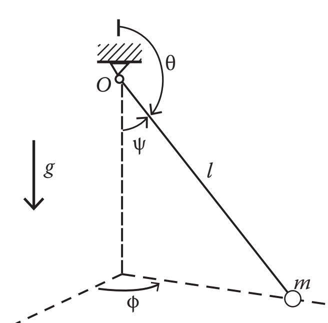


FIGURE 4.9 Spherical pendulum.

The general motion of a spherical pendulum can be determined in terms of elliptic functions. However, there are two particular motions that can be discussed quite simply.

The first is a motion in which the particle performs small oscillations near the lowest point of the spherical surface, and the second is motion in a horizontal circle. To investigate these two motions, let us first introduce the constant h, defined by

$$h = \sin^2\theta\,\dot{\phi} = p_\phi/ml^2. \tag{4.24}$$

With this, Equation 4.22 becomes

$$\ddot{\theta} - (g/l)\sin\theta - h^2\cos\theta\sin^{-3}\theta = 0, \tag{4.25}$$

which becomes, in terms of ψ $(= \pi - \theta)$,

$$\ddot{\psi} + (g/l)\sin\psi - h^2\cos\psi\sin^{-3}\psi = 0. \tag{4.26}$$

If the angle φ is constant (so $h = 0$), Equation 4.26 reduces to

$$\ddot{\psi} + (g/l)\sin\psi = 0. \tag{4.27}$$

Equation 4.27 is just the equation of motion of the simple pendulum. The motion takes place in the plane φ, and φ_0 = constant.

On the other hand, if $\psi = \psi_0$ = constant, the particle describes a horizontal circle (the conical pendulum), and Equation 4.26 reduces to

$$(g/l)\sin\psi_0 - h^2\cos\psi_0\sin^{-3}\psi_0 = 0$$

or

$$h^2 = (g/l)\sin^4\psi_0\sec\psi_0. \tag{4.28}$$

From Equations 4.24 and 4.28, we find the condition for conical motion of the pendulum:

$$\dot{\phi}_0^2 = (g/l)\sec\psi_0. \tag{4.29}$$

Example 4.7: Electric Oscillations

As an illustration of the generality of Lagrangian dynamics, we consider its application to an *LC* circuit (inductive–capacitive circuit) as shown in Figure 4.10. At some instant of time, the charge on the capacitor C is $Q(t)$, and the current flowing through the inductor L is $I(t) = \mathrm{d}Q/\mathrm{d}t = \dot{Q}(t)$. The voltage drop around the circuit is, according to Kirchhoff's law,

$$L\frac{\mathrm{d}I}{\mathrm{d}t} + \frac{1}{C}\int I(t)\,\mathrm{d}t = 0$$

or, in terms of charge Q,

$$L\ddot{Q} + Q/C = 0.$$

This equation is exactly the same form as that for a simple mechanical oscillator:

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

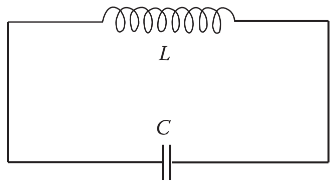


FIGURE 4.10 LC circuit.

If the electric circuit also contains a resistor R, Kirchhoff's law will give

$$L\ddot{Q} + R\dot{Q} + Q/C = 0.$$

This equation is of exactly the same form as that for a damped oscillator:

$$m\ddot{x} + b\dot{x} + kx = 0$$

where b is the damping coefficient. Linear oscillations, which include simple harmonic oscillators, damped oscillators, and forced oscillators without and with damping, will be treated in Chapter 7.

By comparing the corresponding terms in these equations, an analogy between mechanical and electric quantities can be established:

x	displacement	Q	Charge
$\dot{x}$	velocity	$\dot{Q}$	current I
m	mass	L	inductance
$1/k$	1/(spring constant)	C	capacitance
b	damping constant	R	electric resistance
$\frac{1}{2} m\dot{x}^2$	kinetic energy	$\frac{1}{2} L\dot{Q}^2$	energy stored in inductance
$\frac{1}{2} kx^2$	potential energy	$Q^2/2C$	energy stored in capacitance

If we recognize in the beginning that the charge Q in a loop plays the role of a generalized coordinate, and for this loop, $T = \frac{1}{2} LQ^2$ and $V = \frac{1}{2} Q^2/C$, then Lagrangian L is

$$L = T - V = \frac{1}{2}L\dot{Q}^2 - \frac{1}{2}\frac{Q^2}{C}$$

and Lagrange's equation gives

$$L\ddot{Q} + Q/C = 0,$$

which is exactly the one given by Kirchhoff's law.

4.5 NONUNIQUENESS OF THE LAGRANGIAN

The Lagrangian of a system is defined only to within an additive total time derivative of any function of coordinates and time. For example, if $L(q,\dot{q},t)$ and $L'(q,\dot{q},t)$ are two different choices of Lagrangian for a system, they differ by the total derivative with respect to the time of some function $f(q, t)$ of coordinates and time:

$$L'(q,\dot{q},t) = L(q,\dot{q},t) + \frac{\mathrm{d}f(q,t)}{\mathrm{d}t}. \tag{4.30}$$

The actions from these two Lagrangians then differ by a quantity, which gives zero on variation:

$$I' = \int_{t_1}^{t_2} L'(q,\dot{q},t)\,\mathrm{d}t = \int_{t_1}^{t_2} L(q,\dot{q},t)\,\mathrm{d}t + \int_{t_1}^{t_2} \frac{\mathrm{d}f}{\mathrm{d}t}\,\mathrm{d}t = I + f\{q(t_2),t\} - f\{q(t_1),t\}$$

where $\delta[f\{q(t_2), t\} - f\{q(t_1), t\}]$ vanishes because of $\delta q(t_1) = \delta q(t_2) = 0$. Thus, for any choice of a differentiable function $f(q,t)$, the conditions $\delta I' = 0$ and $\delta I = 0$ are equivalent, and the form of the equations of motion, the Lagrange's equations, is unchanged. The Lagrangians related by a transformation of the form given by Equation 4.30 are dynamically equivalent. However, the transformation does carry with it a change in the generalized momentum $p_i = \partial L/\partial \dot{q}_i$ conjugated to each degree of freedom q_i. Because

$$\frac{\mathrm{d}f}{\mathrm{d}t} = \sum_i \frac{\partial f}{\partial q_i}\dot{q}_i + \frac{\partial f}{\partial t}$$

therefore

$$p_i' = \frac{\partial L'}{\partial \dot{q}_i} = \frac{\partial}{\partial \dot{q}_i}\left(L + \frac{\mathrm{d}f}{\mathrm{d}t}\right) = p_i + \frac{\partial f}{\partial q_i}.$$

Equation 4.30 is often called a "gauge transformation" because it does regauge the dependence of the Lagrangian on the independent variables.

It is a real problem to determine how to select a correct Lagrangian that describes the system. Because of the lack of an established criterion, the choice at present is largely a matter of taste and art. This usually does not pose a serious problem for classical mechanics because the central theme of classical mechanics is the study of the equations of motion that are directly associated with observed motion. However, if we are interested in a quantum mechanical treatment of the system, the choice of a physical Lagrangian that describes the system correctly is essential because quantum mechanics is based on the Lagrangian or the Hamiltonian. We shall see later that the Hamiltonian of a system is defined in terms of the Lagrangian, generalized velocities, and their conjugate generalized momenta. Two different Lagrangians that produce the correct classical equations of motion usually lead to contradictory physical results after quantization.

Example 4.8

To see how different choices of Lagrangians can lead to the same equations of motion, consider this extremely simple system: a projectile of mass m projected under gravity in vacuum. Let x denote the horizontal component of the particle's position and y the vertical component. The equations of motion are well known:

$$\ddot{x} = 0, \quad \ddot{y} = -g.$$

The standard Lagrangian for the particle is

$$L_1 = T - V = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) - mgy.$$

But note that the following two functions

$$L_2 = m\dot{x}\dot{y} - mgx \quad \text{and} \quad L_3 = mgy - \frac{1}{2}mgx\dot{y}^2 - m\dot{x}^3\dot{y} - \frac{1}{2}m\dot{y}^2$$

also lead to the same equations of motion, so they are also Lagrangians for this system. But because the action integral of the Lagrangians L_2 and L_3 is not the classical action, we can use this as a guide to reject them and select a correct Lagrangian for the system.

4.6 INTEGRALS OF MOTION AND CONSERVATION LAWS

Lagrange's equations of motion for a system of n degrees of freedom are a set of n second-order differential equations. The solution of each equation has two constants of integration, which usually can be determined from the initial values of the generalized coordinate q and the generalized velocity $\dot{q}$. Sometimes Lagrange's equations can be solved in terms of known functions but not always. In general, most problems either cannot be solved completely or are too tedious to solve. Fortunately, very often, a great deal of information about the system is contained in a number of so-called first integrals of the motion, which are often of greater interest and importance than a complete knowledge of all the q's as a function of time t. The first integrals of motion are functions of the generalized coordinates q_j and the generalized velocities $\dot{q}_j$ of the form $f(q\text{'s}, \dot{q}\text{'s}, t) = \alpha_i$ (constant) whose values remain constant during the motion of the system and are dependent only on the initial conditions of the system. The conservation laws of energy, momentum, and angular momentum that we deduced in Newtonian formalism are of this exact type. These conservation laws can be deduced easily in Lagrangian formalism in a very general and elegant fashion. In the process, they make quite clear the relationship between conservation laws and the symmetry properties of the system. The association goes beyond these conservation laws, beyond classical systems; it finds wide application in modern physics, especially in quantum field theories and particle physics. Hence, the study of symmetry and its uses in learning about the laws of nature is very important.

4.6.1 Cyclic Coordinates and Conservation Theorems

We begin by examining the first integrals of the motion associated with the so-called cyclic coordinates. Coordinates that do not appear explicitly in the Lagrangian of a system are said to be cyclic or ignorable. Be aware that this definition is not universal. Other authors may use them differently. If q_i is a cyclic coordinate, then the Lagrangian L will take the form

$$L = L(q_1, \ldots, q_{i-1}, q_{i+1}, \ldots, q_n, \dot{q}_1, \ldots, \dot{q}_i, \ldots, \dot{q}_n, t) \tag{4.31}$$

and so

$$\partial L/\partial q_i = 0.$$

Lagrange's equations of motion reduce to

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i} = 0, i = 1, 2, \ldots, n$$

or

$$\mathrm{d}p_i/\mathrm{d}t = 0,$$

From this, it follows that

$$P_i = \alpha_i,\ i = 1, 2, \ldots, n \tag{4.32}$$

where the α_i are constants evaluated from the initial conditions. They are called the first integrals of motion. We now find a general conservation theorem: The generalized momentum conjugate to a cyclic coordinate is conserved during the motion. This conservation theorem for generalized momentum is more general than the conservation theorems for linear momentum and angular momentum. In fact, the latter two conservation theorems are contained in the general conservation theorem. For example, if q_1 does not appear in L and if a change δq_1 in q_1 corresponds to a translation of the system through a distance along a certain direction, say x, then

$$\frac{\partial L}{\partial \dot{q}_1} = \frac{\partial}{\partial \dot{q}_1}\left(\frac{1}{2}\sum_i m_i(\dot{x}_i^2 + \dot{y}_i^2 + \dot{z}_i^2\right) = \sum_i m_i\left(\dot{x}_i \frac{\partial \dot{x}_i}{\partial \dot{q}_1} + \dot{y}_i \frac{\partial \dot{y}_i}{\partial \dot{q}_1} + \dot{z}_i \frac{\partial \dot{z}_i}{\partial \dot{q}_1}\right)$$

$$= \sum_i m_i\left(\dot{x}_i \frac{\partial x_i}{\partial q_1} + \dot{y}_i \frac{\partial y_i}{\partial q_1} + \dot{z}_i \frac{\partial z_i}{\partial q_1}\right) = \sum_i m_i\left(\dot{x}_i \frac{\partial l}{\partial q_1} + 0 + 0\right) = \sum_i m_i \dot{x}_i = \alpha_1$$

or

$$p_i = \sum_i m_i \dot{x}_i = \alpha_i \quad \text{(constant)}. \tag{4.33}$$

This is the law of momentum conservation along the x_i axis.

In a similar fashion, it can be shown that if cyclic coordinate q_j is such that δq_j corresponds to a rotation of the system around some axis, then the conservation of its conjugate momentum corresponds to the conservation of an angular momentum. We do not plan to give a general proof here because of its length and tediousness and refer interested students to the book by Goldstein (1980). The two-dimensional harmonic oscillator described in plane polar coordinates gives a simple illustrative example. The Lagrange L is

$$L = T - V = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - \frac{1}{2}kr^2$$

where θ is a cyclic coordinate. Hence,

$$\partial L/\partial \dot{\theta} = mr^2\dot{\theta} = \beta = \text{constant}.$$

$mr^2\dot{\theta}$ is seen to be the angular momentum about the origin.

Summarizing these results, we see that if a translation coordinate is cyclic, the system is invariant under translation along a given direction, and the corresponding linear momentum is conserved. Similarly, when a rotation coordinate is cyclic, the system is invariant under rotation about the given axis, and the conjugate angular momentum is conserved. That is, when arbitrary changes δq_i of a coordinate q_i make no difference to the description of the motion by the Lagrangian, q_i need not appear in the Lagrangian and the description possesses symmetry expressible as an invariance of the system to q_i. The consequent conservation property testifies to a close association between conservation laws and invariances or symmetries. In the following, we shall explore the purely geometric types of symmetry that reflect the general properties of the homogeneity of space and time and the isotropy of space in an inertial reference frame.

4.6.2 Symmetries and Conservation Laws

The homogeneity of space and time means that there are no fixed reference points in space and that there is no preferred instant in time. In other words, the displacement in space of the system as a whole or a shift in time will not change the mechanical properties of a closed system (i.e., one that does not interact with other systems). The isotropy of space means that all directions in space are equivalent; hence, rotation in space does not change the properties of a closed system.

In Lagrangian formalism, the laws of motion of a system are given by Lagrange's equations of motion and so are uniquely determined by the Lagrange of the system. In other words, the effects of a symmetry operation on a system's equations of motion can be determined from its effect on the Lagrangian L. That is what we shall proceed to do, and we begin with the law of energy conservation.

4.6.2.1 Homogeneity of Time and Conservation of Energy

Assume that a system of particles is in unchanging external conditions; this occurs if the system is closed or in a stationary force field (a time-independent constant external force field). In this case, the time, because of its homogeneity, cannot enter the Lagrangian explicitly, and so we have $\partial L/\partial t = 0$. Then the total derivative of the Lagrangian becomes

$$\frac{\mathrm{d}L}{\mathrm{d}t} = \sum_i \left(\frac{\partial L}{\partial q_i}\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i \right) = \sum_i \left(\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i} \right)\dot{q}_i + \frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i \right) = \sum_i \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\dot{q}_i \right)$$

or

$$\sum_i \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L \right) = 0.$$

Thus, we see that, for a closed system or one in a stationary external force field, the quantity in the parentheses, a function of the generalized coordinates and the generalized velocities, remains constant during the motion:

$$\sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L = \alpha \quad \text{(a constant).}$$

Functions of the quantities q_i and $\dot{q}_i$ that remain, during the motion, a constant value determined by the initial conditions are called integrals of motion. Accordingly, α is an integral of constant.

The quantity $\sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L$ is denoted by the symbol H, called the Hamiltonian of the system:

$$H = \sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L = \sum_i p_i\dot{q}_i - L. \tag{4.34}$$

If the potential energy V is velocity-independent, and if the equations of transformation (Equation 4.1) do not depend on time explicitly, then H is equal to the total energy of the system. It is easy to show this explicitly. By the first condition, we have $V = V(x_{i\alpha})$, where $i = 1, 2, \ldots, n$ (number of particles) and $\alpha = 1, 2, 3$ (coordinate axes). By making use of the second condition, namely, the equations of transformation connecting the rectangular and generalized coordinates do not depend on time explicitly [$x_{i,\alpha} = x_{i,\alpha}(q_j)$ or $q_j = q_j(x_{i\alpha})$], we can express V in terms of q_j as $V = V(q_j)$, and so $\partial V(q_j)/\partial \dot{q}_j = 0$. Furthermore, under the second condition, the kinetic energy T is a homogeneous, quadratic function of the q's, and Euler's theorem gives

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

Substituting these into the expression for the Hamiltonian H, we obtain

$$H = \sum_i \frac{\partial L}{\partial \dot{q}_i}\dot{q}_i - L = \sum_i \frac{\partial (T-V)}{\partial \dot{q}_i}\dot{q}_i - (T-V)$$

$$= \sum_i \frac{\partial T}{\partial \dot{q}_i}\dot{q}_i - (T-V) = 2T - (T-V) = T + V = E.$$

Hence, the homogeneity of time leads to the following law: The energy of the system of a closed conservative system of particles (or a system in a stationary external force field) remains constant.

When the equations of transformation connecting the rectangular and generalized coordinates depend on time explicitly, the Hamiltonian H is no longer the total energy of the system, but it is still conserved.

4.6.2.2 Spatial Homogeneity and Momentum Conservation

Consider a closed system. Because of the homogeneity of space, the displacement of the system by a small amount δr must not change the mechanical properties of the system, and so the Lagrangian must retain its previous value. This would not be true for an unclosed system because such a displacement would cause a change in the arrangement of the particles relative to the bodies interacting with them, and so the mechanical properties of the system would be affected. As the displacement δr is very small, we can write

$$\delta L = \sum_j \frac{\partial L}{\partial \vec{r}_j}\cdot \delta \vec{r}_j = \delta \vec{r}\cdot \sum_j \frac{\partial L}{\partial \vec{r}_j} = 0 \tag{4.35}$$

where j is the number of particles, and we have made use of the fact that each particle in the system is displaced by the same amount, and so $\delta \vec{r}_j = \delta \vec{r}$. Before we continue further, we ought to digress for a moment to explain a mathematical notation here: the derivative of a scalar function with respect to a vector quantity. By the derivative of the scalar ϕ with respect to the vector λ, it is understood as a vector having the components $\partial\phi/\partial\lambda_x$, $\partial\phi/\partial\lambda_y$, and $\partial\phi/\partial\lambda_z$. Consequently, the symbol $\partial\phi/\partial r$ stands for a vector with the components $\partial\phi/\partial x$, $\partial\phi/\partial y$, and $\partial\phi/\partial z$, and

$$\frac{\partial \varphi}{\partial \vec{r}}\cdot d\vec{r} = \frac{\partial \varphi}{\partial x}dx + \frac{\partial \varphi}{\partial y}dy + \frac{\partial \varphi}{\partial z}dz.$$

We now return to Equation 4.35. Because $\delta r \neq 0$, we have

$$\sum_j \frac{\partial L}{\partial \vec{r}_j} = 0. \tag{4.36}$$

Lagrange's equations allow us to write

$$\frac{\partial L}{\partial x_j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{x}_j}, \quad \frac{\partial L}{\partial y_j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{y}_j}, \quad \frac{\partial L}{\partial z_j} = \frac{d}{dt}\frac{\partial L}{\partial \dot{z}_j}.$$

Multiplying the first, second, and third of these equations by the unit vectors $\hat{e}_x, \hat{e}_y, \hat{e}_z$, respectively, and summating them, we obtain the expression

$$\frac{\partial L}{\partial \vec{r}_j} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\vec{r}}_j}.$$

Equation 4.36 can thus be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_j \frac{\partial L}{\partial \dot{\vec{r}}_j} = 0. \tag{4.37}$$

The quantity $\partial L/\partial \dot{\vec{r}}_j$ is a vector with the components $\partial L/\partial \dot{x}_j$, $\partial L/\partial \dot{y}_j$, and $\partial L/\partial \dot{z}_j$. These are the projections of the conventional (not generalized) momentum p_j of the jth particle onto the coordinate axes. Hence,

$$\partial L/\partial \dot{\vec{r}}_j = \vec{p}_j.$$

Accordingly, Equation 4.37 can be written as

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_j \vec{p}_j = 0.$$

Hence, it follows that

$$\vec{P} = \sum_j \vec{p}_j = \text{constant}. \tag{4.38}$$

Thus, the homogeneity of space leads to the momentum conservation: The total momentum of a closed system of particles remains constant, that is, it is also an integral of motion.

4.6.2.3 Isotropy of Space and Angular Momentum Conservation

Because of the isotropy of space, the Lagrangian of a closed system should not be affected by an infinitesimal rotation of the system as a whole in space. Accordingly, the Lagrangian should be unchanged, $\delta L = 0$. We are now interested in the increment of the Lagrangian δL in an arbitrary very small rotation of a system through an angle $\delta\theta$. All the vectors characterizing the system will rotate together with it. As a result, they will receive certain increments that will be of the same order as $\delta\theta$. According to Equation 4.36, we have

$$\delta\vec{r}_\alpha = \delta\vec{\theta} \times \vec{r}_\alpha, \quad \text{and} \quad \delta\dot{\vec{r}}_\alpha = \delta\vec{v}_\alpha = \delta\vec{\theta} \times \vec{v}_\alpha. \tag{4.39}$$

Because of the smallness of the quantities $\delta\vec{r}_\alpha$ and $\delta\vec{v}_\alpha$, we have

$$\delta L(\vec{r}_\alpha, \vec{v}_\alpha) = \sum_\alpha \frac{\partial L}{\partial \vec{r}_\alpha}\cdot\delta\vec{r}_\alpha + \sum_\alpha \frac{\partial L}{\partial \vec{v}_\alpha}\cdot\delta\vec{v}_\alpha \tag{4.40}$$

which becomes, in view of Equation 4.39,

$$\delta L(\vec{r}_\alpha, \vec{v}_\alpha) = \sum_\alpha \frac{\partial L}{\partial \vec{r}_\alpha} \cdot (\delta\vec{\theta} \times \vec{r}_\alpha) + \sum_\alpha \frac{\partial L}{\partial \vec{v}_\alpha} \cdot (\delta\vec{\theta} \times \vec{v}_\alpha). \tag{4.41}$$

Now, a cyclic transposition of the multipliers may be performed in a scalar triple product, $\vec{A} \cdot (\vec{B} \times \vec{C}) = \vec{B} \cdot (\vec{C} \times \vec{A}) = \vec{C} \cdot (\vec{A} \times \vec{B})$. Such a transposition in Equation 4.41 yields

$$\delta L = \sum_\alpha \delta\vec{\theta} \cdot \left(\vec{r}_\alpha \times \frac{\partial L}{\partial \vec{r}_\alpha} \right) + \sum_\alpha \delta\vec{\theta} \cdot \left(\vec{v}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right) = \delta\vec{\theta} \cdot \sum_\alpha \left(\vec{r}_\alpha \times \frac{\partial L}{\partial \vec{r}_\alpha} + \vec{v}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right).$$

δL can be further simplified with the help of Lagrange's equation:

$$\frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \vec{v}_\alpha} - \frac{\partial L}{\partial \vec{r}_\alpha} = 0 \quad \text{or} \quad \frac{\partial L}{\partial \vec{r}_\alpha} = \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \vec{v}_\alpha}$$

$$\begin{aligned} \delta L &= \delta\vec{\theta} \cdot \sum_\alpha \left(\vec{r}_\alpha \times \frac{\partial L}{\partial \vec{r}_\alpha} + \vec{v}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right) \\ &= \delta\vec{\theta} \cdot \sum_\alpha \left(\vec{r}_\alpha \times \frac{\mathrm{d}}{\mathrm{d}t} \frac{\partial L}{\partial \vec{v}_\alpha} + \vec{v}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right) = \delta\vec{\theta} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \sum_\alpha \left[\vec{r}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right]. \end{aligned} \tag{4.42}$$

Because $\delta\vec{\theta} \neq 0$, the condition $\delta L = 0$ is equivalent to the condition

$$\frac{\mathrm{d}}{\mathrm{d}t} \sum_\alpha \left[\vec{r}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right] = 0 \tag{4.43}$$

from which it follows that

$$\vec{L} = \sum_\alpha \left[\vec{r}_\alpha \times \frac{\partial L}{\partial \vec{v}_\alpha} \right] = \sum_\alpha \left[\vec{r}_\alpha \times \vec{p}_\alpha \right] = \text{constant}. \tag{4.44}$$

The symbol $\vec{L}$ for angular momentum should not be confused with the Lagrangian L.

Thus, the isotropy of space leads to the angular momentum conservation law: The resultant angular momentum of a closed system of particles remains constant. The angular momentum of a closed system, like its energy and momentum, is also an integral of motion.

Although the conservation law for angular momentum is valid only for a closed system, the conservation law may hold in a more restricted form for a system in an external force field that possesses an axis of symmetry. In such a field, the Lagrangian of the system is invariant about the symmetry axis; hence, the angular momentum of the system about the axis of symmetry is constant in time, that is, it is conserved. The most important such case is that of a central force field that will be examined in Chapter 6. We consider here a simple illustrative example: the motion of a particle on the inner surface of a cone.

Example 4.9

A particle of mass m is constrained to move under the influence of gravity on the smooth inner surface of the paraboloid of revolution $x^2 + y^2 = az^2$, where a is a constant. Show that the angular momentum of the particle about the axis of symmetry of the system is conserved.

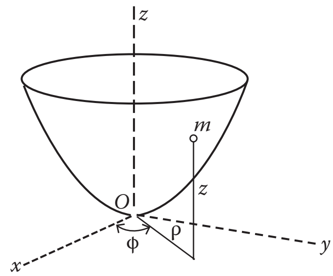


FIGURE 4.11 Particle constrained to move under gravity on the smooth inner surface of the paraboloid.

Solution:

The problem possesses cylindrical symmetry, so we choose ρ, ϕ, and z as the generalized coordinates, and we let the axis of the paraboloid correspond to the z-axis and the vertex of the paraboloid be located at the origin (Figure 4.11). The Lagrangian of the system is

$$L = T - V = \frac{1}{2}m(\dot{\rho}^2 + \rho^2\dot{\varphi}^2 + \dot{z}^2) - mgz$$

where the reference level for the potential energy is set at the vertex of the paraboloid.

As ϕ is a cyclic coordinate, $\partial L/\partial\phi = 0$. Then, Lagrange's equation for coordinate ϕ reduces to

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial\dot{\varphi}} = \frac{\mathrm{d}}{\mathrm{d}t}(m\rho^2\dot{\varphi}) = 0$$

from which we obtain

$$m\rho^2\dot{\varphi} = \text{constant}. \tag{4.45}$$

Note that $m\rho^2\dot{\varphi} = m\rho^2\omega$ is just the angular momentum of the particle about the system's axis of symmetry, the z-axis. Thus, Equation 4.45 simply expresses the fact that the angular momentum of the particle about the axis of symmetry is conserved.

We have learned that the laws of conservation of energy, linear momentum, and angular momentum are an immediate consequence of the general symmetry properties of space and time. We should also note that these laws also explain, from their derivations, why the following pairs of variables are associated with each other:

$$(\vec{r}, \vec{p}), (\vec{\theta}, \vec{L}), \text{ and } (\mathrm{t}, E).$$

4.6.2.4 Noether's Theorem

By now, we have all learned that symmetries of the Lagrangian gave rise to constants of the motion. But the constants of the motion do not always come from the obvious symmetries of the Lagrangian, nor do they always have a simple form. Mathematician Emmy Noether took a general approach to this problem in 1919 and found a theorem that states essentially that if corresponding to a variable α in the Lagrangian, and the Lagrangian remains unchanged for a change of α to α + ε where ε is infinitesimal, we will have a conservation principle. The conservation laws of energy, momentum, and angular momentum are just an example of Noether's theorem. A straightforward and simple exposition of Noether's work is given in Appendix 3.

4.7 SCALE INVARIANCE

It is evident that the multiplication of the Lagrangian of a system by any constant does not affect Lagrange's equations of motion. This is often called scale invariance. Very often, this property of the Lagrangian enables us to obtain useful results without integrating the equations of motion.

Consider a system whose potential energy V is a homogeneous function of the coordinates q_1, $q_2,\ldots,q_n$. If the coordinates undergo a scale transformation

$$q_i \to q_i' = \alpha q_i \text{ and } i = 1,2,\ldots,n \tag{4.46}$$

then

$$V(q_1,\ldots,q_n) \to V(\alpha q_1,\ldots,\alpha q_n) = \alpha^k V(q_1,\ldots,q_n) \tag{4.47}$$

where α is a dimensionless constant parameter, and k is the degree of homogeneity of the function V. Let there also be a simultaneous scaling of the time coordinate:

$$t \to t' = \beta t \tag{4.48}$$

with β being a dimensionless constant parameter. Under these transformations on the space–time coordinates, all the velocities are changed by a factor of α/β:

$$\dot{q}_i \to \dot{q}_i' = (\alpha/\beta)\dot{q}_i. \tag{4.49}$$

The kinetic energy T is quadratic in the velocities for a scleronomic (time-independent) system. Thus, under these scale transformations, T changes by a factor $(\alpha/\beta)^2$, and V changes by a factor α^k. If

$$\alpha^k = (\alpha/\beta)^2$$

or

$$\alpha^{2-k} = \beta^2 \tag{4.50}$$

then both T and V, and, hence, the Lagrangian L of the system, get multiplied by a same constant parameter, so under these scale transformations, the equations of motion remain unchanged. Thus, we conclude that for a homogeneous potential of degree k, under the transformations

$$\begin{aligned} q_i &\to q_i' = \alpha q_i, \qquad i = 1,2,\ldots,n, \\ t &\to t' = \alpha^{(2-k)/2} t, \end{aligned} \tag{4.51}$$

the equations of motion remain invariant. However, the equations of motion now permit a series of geometrically similar paths differing only in size, and the time of the motion between two corresponding points is in the ratio

$$\frac{t'}{t} = \beta = \alpha^{(2-k)/2} = \left(\frac{\lambda'}{\lambda}\right)^{(2-k)/2} \tag{4.52}$$

where λ'/λ is the ratio of the linear dimensions of the two paths. With the aid of this simple relationship, we can make some useful inferences about the properties of the motion without actually integrating the equations of motion. For example, we saw that in small oscillations the potential energy is quadratic in the coordinates, so $k = 2$. We find then, from Equation 4.52, that the period of small oscillations is independent of their amplitudes. As another example, in the gravitational interaction of two masses or the Coulomb interaction of two charges, the potential energy is inversely proportional to the distance between them, so it is a homogeneous function of degree $k = -1$. Then, from Equation 4.52, we find that the square of the period of revolution in the orbit is proportional to the cube of the size of the orbit. We shall see in Chapter 6 on central force motion that this is Kepler's third law of planetary motion.

4.8 NONCONSERVATIVE SYSTEMS AND GENERALIZED POTENTIAL

Thus far, we have concentrated on conservative systems. If the forces acting on the system are not conservative, then Lagrange's equations of motion will take the basic form of Equation 4.30:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} = Q_j,\ j = 1, 2, \ldots, n$$

where Q_j stands for a nonconservative force that cannot be obtained from a scalar potential function. However, when Q_j can be obtained from a velocity-dependent potential function $U(q, \dot{q}, t)$ in accordance with the equation

$$Q_j = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial \dot{q}_j} - \frac{\partial U}{\partial q_j} \tag{4.53}$$

then Equation 4.21 can be put in the standard form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0,\ j = 1, 2, \ldots, n$$

with the Lagrangian given by

$$L = T - U. \tag{4.54}$$

The velocity-dependent potential U is often called the generalized potential.

4.9 CHARGED PARTICLE IN ELECTROMAGNETIC FIELD

One example is the important Lorentz force that can be obtained from such a velocity-dependent potential. In rationalized units, the Lorentz force is $\vec{F} = e\left(\vec{E} + \vec{v} \times \vec{B}\right)$, where the electric field intensity $\vec{E}$ and the magnetic induction vector $\vec{B}$ can be obtained from a scalar potential φ and a vector potential $\vec{A}$, and in terms of φ and $\vec{A}$, the Lorentz force is

$$\vec{F} = q\left\{-\nabla\phi - \partial\vec{A}/\partial t + \vec{v} \times \nabla \times \vec{A}\right\} \tag{4.55}$$

where

$$\vec{E} = -\nabla\phi - \frac{\partial\vec{A}}{\partial t}, \quad \vec{B} = \nabla \times \vec{A} \tag{4.56}$$

and φ and $\vec{A}$ are, in general, functions of position and time. Consider the x-component:

$$(\nabla\phi)_x = \partial\phi/\partial x,$$

and

$$\left(\vec{v}\times\nabla\times\vec{A}\right)_x = v_y(\partial A_y/\partial x - \partial A_x/\partial y) - v_z(\partial A_x/\partial z - \partial A_z/\partial x)$$
$$= v_y\partial A_y/\partial x + v_z\partial A_z/\partial x + v_x\partial A_x/\partial x - v_y\partial A_x/\partial y - v_z A_x/\partial z - v_x\partial A_x/\partial x$$

$$\mathrm{d}A_x/\mathrm{d}t = \partial A_x/\partial t + (v_x\partial A_x/\partial x + v_y\partial A_x/\partial y + v_x\partial A_z/\partial z).$$

And, finally, we have

$$\left(\vec{v}\times\nabla\times\vec{A}\right)_x = \frac{\partial}{\partial x}\left(\vec{v}\cdot\vec{A}\right) - \frac{\mathrm{d}A_x}{\mathrm{d}t} + \frac{\partial A_x}{\partial t}$$

and so the x-component of the Lorentz force is

$$F_x = e\left[-\frac{\partial\phi}{\partial x} + \frac{\partial}{\partial x}\left(\vec{v}\cdot\vec{A}\right) - \frac{\mathrm{d}A_x}{\mathrm{d}t}\right]$$
$$= e\left[-\frac{\partial\phi}{\partial x} + \frac{\partial}{\partial x}\left(\vec{v}\cdot\vec{A}\right) - \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial}{\partial\dot{x}}\left(\vec{v}\cdot\vec{A}\right)\right)\right].$$

Because the scalar potential φ is independent of velocity, this expression is equivalent to

$$F_x = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial U}{\partial\dot{x}} - \frac{\partial U}{\partial x}$$

with the velocity-dependent potential U given by

$$U = e\left(\phi - \vec{v}\cdot\vec{A}\right). \tag{4.57}$$

The Lagrangian L for a charged particle in an electromagnetic field can now be written as

$$L = T - U = \frac{1}{2}m\dot{x}^2 - e\left(\phi - \vec{v}\cdot\vec{A}\right). \tag{4.58}$$

The generalized momentum of the charged particle is

$$P_x = \frac{\partial L}{\partial\dot{x}} = m\dot{x} + eA_x$$

or

$$\vec{P} = m\vec{v} + e\vec{A} \tag{4.59}$$

which shows clearly that a part of the momentum is associated with the electromagnetic field. The total energy of the charged particle is

$$E_{\mathrm{tot}} = T + e\varphi \tag{4.60}$$

where $e\varphi$ is the potential energy associated with the electric field. This is as expected because the magnetic force, $e\vec{v}\times\vec{B}$, does not do work on the particle; hence, the vector potential does not enter into the energy expression (Equation 4.60).

4.10 FORCES OF CONSTRAINT AND LAGRANGE'S MULTIPLIERS

The Lagrangian method concentrates solely on active forces and ignores internal force effects resulting from joints, connectors, and contact with constraints. If forces of constraint happen to be the object of interest, what can we do then? Obviously, Lagrange's equation of motion would not give us any help. Fortunately, we can utilize Lagrange's multiplier method. What is Lagrange's multiplier method? It is the subject of this section.

We have discussed two methods that can be used in the analysis of systems with holonomic constraints, namely, the elimination of variables using the constraint equations and the use of independent generalized coordinates. Lagrange's multiplier method is the third method, and it can also be applied to certain types of nonholonomic constraints. We, therefore, first introduce Lagrange's multiplier method for nonholonomic systems and then show how it can also be applied to holonomic systems.

The derivation of Lagrange's equations for a holonomic system requires that the generalized coordinates be independent. With nonholonomic constraints, there are more generalized coordinates than number of degrees of freedom. Thus, the variations of q_i are no longer independent of each other for a virtual displacement consistent with the constraints. However, we are still able to treat nonholonomic systems, provided the equations of the constraints are of the following type:

$$\sum_k a_{jk}\mathrm{d}q_k + a_{jt}\mathrm{d}t = 0,\ j = 1,2,\ldots,m \tag{4.61}$$

where the coefficients a_{jk} and a_{jt} may be functions of the q's and time t. The virtual displacement δqs occurring in the variation and consistent with the constraints must meet the following conditions:

$$\sum_k a_{jk}\delta q_k = 0,\ j = 1,2,\ldots,m. \tag{4.62}$$

We can now use this to reduce the number of virtual displacements to independent ones by using the so-called Lagrange's multiplier method. If Equation 4.62 holds, then it is also true that

$$\lambda_j \sum_k a_{jk}\delta q_k = 0,\ j = 1,2,\ldots,m \tag{4.63}$$

where λ_j are some undetermined constants known as Lagrange's multipliers. They are, in general, functions of the coordinates and of the time t. Now, summing Equation 4.63 over j and then integrating the result with respect to the time from point 1 (time t_1) to point 2 (time t_2),

$$\int_1^2 \sum_{j,k} \lambda_j a_{jk}\delta q_k \mathrm{d}t = 0. \tag{4.64}$$

Hamilton's principle is assumed to hold for the nonholonomic systems:

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

Combining this with Equation 4.64, we obtain

$$\int_1^2 \mathrm{d}t \sum_{k=1}^{n} \left(\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} + \sum_j \lambda_j a_{jk} \right) \delta q_k = 0. \tag{4.65}$$

The δq's are not necessarily independent; they are connected by the m relationships (Equation 4.62). We rewrite Equation 4.65 as

$$\int_1^2 \mathrm{d}t \sum_{k=1}^{n-m} \left(\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} + \sum_j \lambda_j a_{jk} \right) \delta q_k$$

$$+ \int_1^2 \mathrm{d}t \sum_{k=n-m+1}^{n} \left(\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} + \sum_j \lambda_j a_{jk} \right) \delta q_k = 0.$$

The values of the λ_j's are at our disposal; we now choose them to be such that the second integral vanishes; that is,

$$\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} + \sum_j \lambda_j a_{jk} = 0,\ k = n-m+1,\ldots,n. \tag{4.66}$$

Having done this, we are free to choose $q_1, q_2, \ldots, q_{n-m}$ in the first integral arbitrarily. And, accordingly, we have

$$\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} + \sum_j \lambda_j a_{jk} = 0,\ k = 1,2,\ldots,n-m. \tag{4.67}$$

Equations 4.66 and 4.67 together give us the complete set of Lagrange's equations for nonholonomic systems:

$$\frac{\partial L}{\partial q_k} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} + \sum_j \lambda_j a_{jk} = 0,\ k = 1,2,\ldots,n. \tag{4.68}$$

Equation 4.68, together with the equations of constraints, now being written as first-order differential equations,

$$\sum_k a_{jk}\dot{q}_k + a_{jt} = 0 \tag{4.69}$$

constitute $n + m$ equations for $n + m$ unknowns, namely, the nq's and the $m\lambda_j$.

One question remains to be answered: what is the significance of λ_j? To answer this question, let us remove the constraints on the system in such a manner that we keep the motion of the system unchanged by applying external forces Q_k. Under the influence of these forces, the equations of motion that describe the motion of the system remain the same, and the applied forces Q_k must be equal to the forces of constraints:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = Q_k.$$

But these equations must be identical with Equation 4.68. Hence, we can identify $\sum_j \lambda_j a_{jk}$ with the forces of constraints Q_k. Thus, we see that, in Lagrange's multiplier method, the constraints enter the equations of motion in the form of constraint forces rather than in geometric terms, and λ's are linearly related to the forces of the constraints.

We now consider a holonomic system in which there are more generalized coordinates than degrees of freedom. Suppose there are m holonomic equations of constraint of the form

$$\phi_j(q_1, q_2, \ldots, q_n, t) = 0, j = 1, 2, \ldots, m. \tag{4.70}$$

Taking the total differential of φ_j, we obtain

$$\mathrm{d}\phi_j = \sum_k \frac{\partial \phi_j}{\partial q_k} \mathrm{d}q_k + \frac{\partial \phi_j}{\partial t} \mathrm{d}t = 0 \tag{4.71}$$

which is identical in form to Equation 4.61 and has the coefficients

$$a_{jk} = \frac{\partial \phi_j}{\partial q_k}, \quad a_{jt} = \frac{\partial \phi_j}{\partial t}. \tag{4.72}$$

Thus, we can use Lagrange's multiplier method for holonomic systems when (1) it is inconvenient to reduce all the q's to dependent coordinates or (2) we desire to solve for the forces of constraint.

The examples that follow will provide familiarity with and a feeling for the way the multiplier techniques are used.

Example 4.10

A particle of mass m is placed at the top of a smooth hemisphere of radius a (Figure 4.12). Find the reaction R of the hemisphere on the particle. If the particle is disturbed, at what height does it leave the hemisphere?

Solution:

The two generalized coordinates are r and θ. In terms of these two coordinates, the kinetic and potential energies of the particle are given by, respectively,

$$T = \frac{1}{2} m\left(\dot{r}^2 + r^2\dot{\theta}^2\right), \quad V = mgr\cos\theta$$

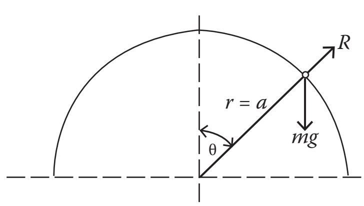


FIGURE 4.12 Particle at the top of a vertical hoop.

where the reference level for the potential energy is set at the bottom of the hemisphere. The Lagrangian L is

$$L = T - V = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - mgr\cos\theta.$$

The equation of constraint is $r = a$ or $\mathrm{d}r = 0$, so $a_r = l$ and $a_\theta = 0$. One Lagrangian multiplier λ is needed. The two Lagrange's equations are

$$ma\dot{\theta}^2 - mg\cos\theta + \lambda = 0 \quad (4.73)$$

$$ma^2\ddot{\theta} - mga\sin\theta = 0. \quad (4.74)$$

Now, let $p = \dot{\theta}$; then

$$\ddot{\theta} = \frac{\mathrm{d}\dot{\theta}}{\mathrm{d}t} = \frac{\mathrm{d}p}{\mathrm{d}t} = \frac{\mathrm{d}p}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = p\frac{\mathrm{d}p}{\mathrm{d}\theta}.$$

In terms of p, Equation 4.78 becomes

$$p\frac{\mathrm{d}p}{\mathrm{d}\theta} - \frac{g}{a}\sin\theta = 0$$

from which, after integration, we get

$$p^2 = \dot{\theta}^2 = -\frac{2g}{a}\cos\theta + \frac{2g}{a}$$

where the integration constant is $2g/a$ because $\dot{\theta} = 0$ at $\theta = 0$. Substituting this into Equation 4.73, we obtain the reaction force

$$\lambda = mg(3\cos\theta - 2).$$

If the particle is disturbed, it will leave the hoop when $\lambda = 0$, that is, at

$$\theta = \cos^{-1}(2/3).$$

Example 4.11

Consider Example 4.4 again. Find the reaction of the wire on the bead.

Solution:

The problem was solved in Example 4.6 as having one degree of freedom. We now consider it as having two coordinates, r and θ, but there is one constraint condition, namely, $r = b$. Thus, $\mathrm{d}r = 0$, and so $a_r = 1$ and $a_\theta = 0$.

The x and y coordinates of the bead are now given by

$$x = b\cos\omega t + r\cos(\theta + \omega t)$$

$$y = b\sin\omega t + r\sin(\theta + \omega t).$$

The Lagrangian L of the bead is

$$L = T = \frac{1}{2}m(\dot{x}^2+\dot{y}^2)$$
$$= \frac{m}{2}\left[b^2\omega^2+\dot{r}^2+r^2(\dot{\theta}+\omega)^2+2b\omega\dot{r}\sin\theta+2b\omega r(\dot{\theta}+\omega)\cos\theta\right].$$

Substituting this into the augmented Lagrange's equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{r}}-\frac{\partial L}{\partial r}=\lambda a_r$$

we obtain

$$m\left[\ddot{r}+b\omega\dot{\theta}\cos\theta-r(\dot{\theta}+\omega)^2-b\omega(\dot{\theta}+\omega)\cos\theta\right]=\lambda$$

which reduces to the following equation after inserting the constraint condition $r = b$:

$$\lambda=-mb\left[\omega^2\cos\theta+(\dot{\theta}+\omega)^2\right].$$

$\ddot{\theta}$ can be found in Example 4.6, from which we can obtain $\dot{\theta}$ by integrating once, thus eliminating it from the expression for λ.

Example 4.12

A ladder of length l and mass m is inclined against a frictionless wall and floor as shown in Figure 4.13. Find the equations of motion and reactions of the wall and the floor.

Solution:

We need three variables, x, y, and θ, to specify the position of the center of mass of the ladder and its orientation. The Lagrangian L is

$$L = T - V = \frac{1}{2}m(\dot{x}^2+\dot{y}^2)+\frac{1}{2}I\dot{\theta}^2-mgy$$

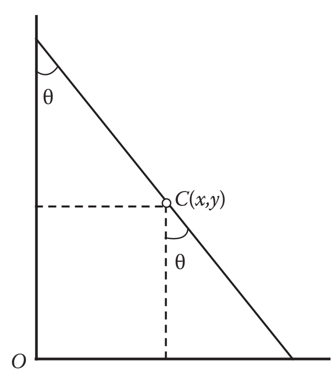


FIGURE 4.13 Ladder inclined against a wall and floor.

where the moment of inertia $I = ml^2/12$. The three variables, x, y, and θ, are related by two holonomic constraints:

$$x - \frac{l}{2}\sin\theta = 0 \quad \text{and} \quad y - \frac{l}{2}\cos\theta = 0$$

from which we have

$$dx - (l/2)\cos\theta \, d\theta = 0 \text{ and } dy + (l/2)\sin\theta \, d\theta = 0.$$

The augmented Lagrange's equations now take the form

$$m\ddot{x} = \lambda_x \tag{4.75}$$

$$m\ddot{y} + mg = \lambda_y. \tag{4.76}$$

From the constraint conditions, we find

$$\ddot{x} = -(l/2)\sin\theta\dot{\theta}^2 + (l/2)\cos\theta\ddot{\theta}$$

$$\ddot{y} = (l/2)\cos\theta\dot{\theta}^2 - (l/2)\sin\theta\ddot{\theta}.$$

From the first equation and Equation 4.75, we find

$$\lambda_x = m\left[(-l/2)\sin\theta\dot{\theta}^2 + (l/2)\cos\theta\ddot{\theta}\right].$$

Similarly, from the second equation and Equation 4.76, we find

$$\lambda_y = m\left[(l/2)\cos\theta\dot{\theta}^2 + (-l/2)\sin\theta\ddot{\theta} + g\right].$$

Because the three variables are connected by two constraints, therefore, there is only one degree of freedom and one generalized coordinate. If we choose θ to be the generalized coordinates and the constraint conditions to eliminate $\dot{x}$ and $\dot{y}$, then Equations 4.75 and 4.76 can be combined as one:

$$\frac{ml^2}{24}\ddot{\theta} = -\lambda_x l\cos\theta/2 + \lambda_y l\sin\theta/2.$$

Combining this with the two expressions for λ_x and λ_y, we obtain, after simplification

$$\ddot{\theta} = (12/5)\sin\theta\cos\theta\dot{\theta}^2 + (6g/5)\sin\theta.$$

This is the equation of motion of the ladder. After we solve it, in principle, we can then find λ_x and λ_y.

4.11 LAGRANGIAN VERSUS NEWTONIAN APPROACH TO CLASSICAL MECHANICS

The Newtonian and Lagrangian formulations of classical mechanics have been discussed; we should now compare the differences between them and weigh their relative advantages.

First, the Newtonian force-momentum formulation is vectorial in nature, whereas the Lagrangian work-energy approach involves only scalar functions. All the equations of motion come from a single scalar function: the Lagrangian. It is true that when dealing with simple systems the directional properties of the vectors involved assist our intuition in setting up the problem, but the Lagrangian method is simpler mathematically as the mechanical system becomes more complex.

Second, the Lagrangian method concentrates solely on active forces, completely ignoring all internal force effects resulting from joints, connectors, and contact with constraints, and the Newtonian approach must take the forces of constraint into consideration. So long as constraint forces per se are not a consideration, the Lagrangian method offers an advantage that becomes increasingly attractive as the mechanical system becomes more complex. If forces of constraint happen to be the object of interest, then Lagrangian multipliers must be utilized. In the Newtonian approach, these constraint effects appear naturally and directly in the fundamental formulation.

Third, the Lagrangian potential function is confined to conservative forces alone, putting the Lagrangian method at a serious disadvantage when dealing with velocity-dependent forces. Newton's law of motion, on the other hand, takes on all comers as grist for its mill and is thus universally applicable.

Fourth, in terms of generalized coordinates, we can write Lagrange's equations of motion in some form that is equally suitable for all coordinates. The Lagrangian approach can also be applied to a wide range of physical phenomena, particularly those involving fields, with which Newton's equations are not usually associated. Which method, then, is generally superior? The answer depends on the application. Each approach possesses certain inherent advantages in given situations and should be utilized accordingly.

PROBLEMS

1. What generalized coordinates can be used to completely specify the motion of each of the following?
   (a) A particle is constrained to move on a sphere.
   (b) A circular cylinder rolls down an inclined plane.
   (c) A disk rolls without slipping across a horizontal plane. The plane of the disk remains vertical, but it is free to rotate about a vertical axis.
2. Classify each of the following as to whether they are (1) scleronomic or rheonomic, (2) holonomic or nonholonomic, and (3) conservative or nonconservative.
   (a) A sphere rolls down from the top of a fixed sphere.
   (b) A cylinder rolls without slipping down a rough inclined plane.
   (c) A particle slides on a very long frictionless wire that rotates with constant angular velocity about a horizontal axis.
3. A simple pendulum of length b is held at an angle θ_0 to the vertical axis and is released with a velocity v toward the position of equilibrium. Employing Lagrange's method, find the amplitude and phase of the resulting motion.
4. A particle of mass m moves in a conservative force field. Find the Lagrangian function and the equation of motion in cylindrical coordinates (ρ, φ, z).
5. A bead slides without friction on a frictionless wire in the shape of a cycloid with equations

   $$x = a(\theta - \sin\theta)\ y = a(1 + \cos\theta)$$

   where $0 \le \theta \le 2\pi$. Find the Lagrangian function and the equation of motion.
6. Derive Lagrange's equation of motion for a uniformly thin disk that rolls without slipping on a horizontal plane under the influence of a horizontal force applied at its center. Determine the generalized force.
7. A particle, moving in the potential $V(x) = -Fx$, travels from the point $x = 0$ to the point $x = a$ in a time interval τ. Find the time dependence of the position of the particle, assuming it

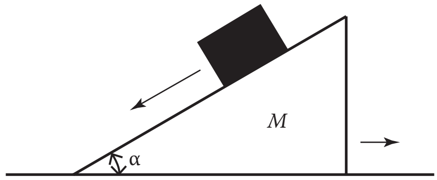


FIGURE 4.14 A wedge slides on a horizontal plane. A particle moves freely on the wegde.

to be of the form $x(t) = At^2 + Bt + C$ and determine the constants A, B, and C such that the action is a minimum.

8. A wedge of mass M and angle α (as shown in Figure 4.14) slides freely on a horizontal plane. A particle of mass m moves freely on the wedge. Determine the motion of the particle as well as that of the wedge.
9. Two small metal spheres of equal mass m and equal charge $+e$ are suspended by insulating threads of equal length b from two points at a distance a apart on the same horizontal line (Figure 4.15). If one sphere is pulled aside a small distance d_1 and the other a small distance d_2 (both d_1 and d_2 are small compared with a) in the plane formed by the equilibrium positions of both threads, determine the resulting motion in this plane.
10. A particle of mass m slides freely on a smooth straight wire that is constrained to rotate in a vertical plane with constant angular speed ω about a fixed point on the wire. Determine the Lagrangian function and Lagrange's equations of motion.
11. A hoop of mass M, radius a, and moment of inertia $\frac{1}{2}Ma^2$ about its axis has a simple pendulum of mass M and length $\frac{1}{2}a$ attached to its center. If the hoop rolls down a plane inclined at an angle α to the horizontal, determine the Lagrangian function and the equations of motion.
12. A particle of mass m slides on a straight, frictionless wire that is fixed at one end on a vertical axis. The wire rotates about this vertical axis with constant angular velocity ω.
    (a) Determine the Lagrangian function and Lagrange's equations of motion.
    (b) If the particle starts from rest at the upper end, how long will it take for it to reach the lower end, assuming that the length of the wire is b?
13. A bead of mass m slides freely on a smooth circular wire of radius b that rotates in a horizontal plane about a point on the circle with a constant angular velocity ω.
    (a) Determine the Lagrangian function and Lagrange's equations of motion.
    (b) Find the reaction of the wire.

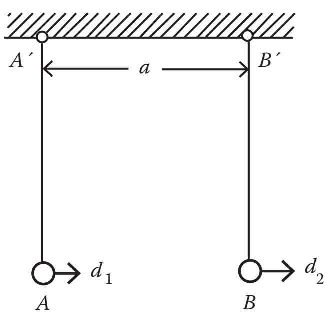


FIGURE 4.15 Two small charged metal spheres suspended by insulating threads.

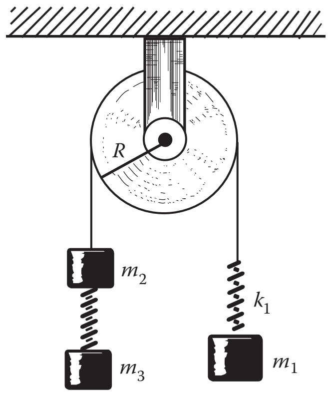


FIGURE 4.16 Pulley–spring system.

14. Use the method of Lagrangian undetermined multipliers to obtain the equations of motion for a particle in the field of gravity when it is constrained to move along a parabola $z = ax^2$ in a vertical plane.
15. Determine the Lagrangian function and the equations of motion for the pulley-spring system shown in Figure 4.16. Assume that the rope rolls off the disk, that the disk has a moment of inertia I, and that the unstretched length of each spring is b_0.
16. A particle of mass m moves under the influence of gravity on the inner surface of a smooth cone of half-angle α. The axis of the cone is vertical with its vertex downward. Determine the condition on the angular velocity ω such that the particle can describe a horizontal circle at a height h above the vertex. Gravity is acting vertically downward. Also show that if the particle is displaced slightly from this circular path, it will undergo oscillations about the path within a period $(2\pi/\cos\alpha)(h/3g)^{1/2}$.
17. A relativistic particle has mass

$$m = \frac{m_0}{\sqrt{1 - v^2/c^2}}$$

where m is the so-called rest mass, v is the velocity of the particle, and c is the velocity of light in a vacuum. Assuming that the motion is in one dimension, show that a Lagrangian function defined by

$$L = -m_0c^2\sqrt{1 - v^2/c^2} - V,$$

where V (the potential energy) is not velocity-dependent, provides the correct equations of motion. Also find the generalized momentum.
18. Derive Lagrange's equation of motion from Newton's laws.
19. Many nonconservative systems are caused by the presence of the dissipative forces, such as the friction by which energy is lost from the system. If the dissipative force is proportional to the velocity components of the particle in accordance with the simple law

$$F_f = -k_x v_x$$

where k_x is a constant, show that a dissipative force of this type can be incorporated into the analytical description by introducing a quantity R, the Rayleigh dissipative function

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

where the summation is over the particles of the system, and that Lagrange's equations take the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{q}_i}\right) - \frac{\partial L}{\partial q_i} + \frac{\partial R}{\partial \dot{q}_i} = 0$$

where $\partial R/\partial \dot{q}_i$ is the generalized force.

20. From the transformation Equation 4.1, show that

$$\frac{\partial \dot{\vec{r}}_i}{\partial \dot{q}_j} = \frac{\partial \vec{r}_i}{\partial q_j}\ .$$

REFERENCE

Goldstein, H. *Classical Mechanics*, 2nd ed., Addison Wesley, Readings, Massachusetts, 1980.

5 Hamiltonian Formulation of Mechanics

Descriptions of Motion in Phase Spaces

The Lagrangian dynamics have been shown to be elegant and straightforward. Half a century after Lagrange, William R. Hamilton introduced another way of writing the equations of motion of a system. Instead of a single differential equation of second order for each coordinate, Hamilton found a set of twice as many equations but only of the first order, that is, containing only first derivatives with respect to the time. How could Hamilton achieve this?

In the Lagrangian formulation, we can transform to a new set of coordinates as well; once the coordinates have been chosen, the corresponding velocities are also determined. Hamilton removed this subordinate feature of velocities by eliminating them in favor of the generalized momenta. One reason for this change is that the momenta are often conservative quantities, and symmetries is even more explicit in Hamilton's new formulation. Another reason is that the Lagrangian function has no useful physical meaning, but the Hamiltonian function, when conserved, is the energy of the system, a very important quantity. The Hamilton method is intimately connected with symmetry and conservation.

5.1 THE HAMILTONIAN OF A DYNAMIC SYSTEM

As we learned in Chapter 4, the Lagrangian L for a holonomic system of n degrees of freedom is defined in terms of q_i and $\dot{q}_i$ as

$$L = L(q_1,\ q_2,\ldots,\ q_n,\ \dot{q}_1,\ \dot{q}_2,\ldots,\ \dot{q}_n,\ t)$$

and the equations of motion are Lagrange's equations, a set of second-order differential equations,

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0, \quad j = 1, 2, \ldots n.$$

The generalized momentum conjugate to q_j is defined as

$$p_j = \frac{\partial L}{\partial \dot{q}_j}.$$

We are tempted, at this point, to search for a new way of describing the complete mechanical state of a system by giving q_j and p_j as functions of time, rather than q_i and $\dot{q}_i$. Hamilton first developed the way in which this can be done. He replaced the Lagrangian function with a quantity H, called the Hamiltonian or Hamilton's function of the system, by the defining relationship

$$H = \sum_j p_j \dot{q}_j - L \tag{5.1}$$

which we already saw in Chapter 4. In mathematics, Equation 5.1 is called a Legendre transformation, which is a procedure for the passage from one set of independent variables to another.

Although $\dot{q}_i$ explicitly appears in the defining expression (Equation 5.1), H is a function of the generalized coordinates q_j, the generalized momenta p_j, and the time t. This is because we can solve the defining expressions $p_j = \partial L/\partial \dot{q}_j$ explicitly for the $\dot{q}$'s in terms of p_j, q_j, and t. There are exception cases where the transformation from the Lagrangian to the Hamiltonian is hindered by the fact that the equation $p_j = \partial L/\partial \dot{q}_j$ cannot be solved for $\dot{q}_i$ as functions of p, q, and t. Such cases are called degenerate cases, and they were usually handled by special methods for each Lagrangian. A discussion of the general theory of this special topic goes beyond our syllabus.

5.1.1 Phase Space

The q's and p's are now treated on equal footing: $H = H(q_j, p_j, t)$. There are two quantities for each degree of freedom of the mechanical system: the generalized coordinate itself and the conjugate generalized momentum. Just as with the configuration spaces that are spanned by the n independent q's, we can imagine a space of $2n$ dimensions spanned by the $2n$ variables $q_1, q_2, \ldots, q_n, p_1, p_2, \ldots, p_n$. Thus, every point in this space represents both the positions and momenta of all particles in the system. Such space is called "phase space" and is particularly useful in both statistical mechanics and the study of nonlinear oscillations. The evolution of a representative point in this space is determined by a new set of equations of motion called the Hamilton's equations. We now proceed to find these new equations of motion.

5.2 HAMILTON'S EQUATIONS OF MOTION

5.2.1 Hamilton's Equations from Lagrange's Equations

We first take a total differentiation of the Hamiltonian $H = H(p, q, t)$, which gives

$$\mathrm{d}H = \sum_j \frac{\partial H}{\partial q_j}\mathrm{d}q_j + \sum_j \frac{\partial H}{\partial p_j}\mathrm{d}p_j + \frac{\partial H}{\partial t}\mathrm{d}t. \tag{5.2}$$

From Equation 5.1, we obtain

$$\mathrm{d}H = \sum_j p_j \mathrm{d}\dot{q}_j + \sum_j \dot{q}_j \mathrm{d}p_j - \sum_j \frac{\partial L}{\partial \dot{q}_j}\mathrm{d}\dot{q}_j - \sum_j \frac{\partial L}{\partial q_j}\mathrm{d}q_j - \frac{\partial L}{\partial t}\mathrm{d}t$$

which can be simplified. First, the first and third terms cancel each other; second, we rewrite the fourth term as

$$\frac{\partial L}{\partial q_j}\mathrm{d}q_j = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j}\mathrm{d}q_j = \frac{\mathrm{d}p_j}{\mathrm{d}t}\mathrm{d}q_j = \dot{p}_j \mathrm{d}q_j.$$

As a result, we obtain

$$\mathrm{d}H = \sum_j \dot{q}_j \mathrm{d}p_j - \sum_j \dot{p}_j \mathrm{d}q_j - (\partial L/\partial t)\mathrm{d}t. \tag{5.3}$$

We now compare Equations 5.2 and 5.3. Because q and p are independent variables, the variations $\mathrm{d}q_j$, $\mathrm{d}p_j$, and $\mathrm{d}t$ are mutually independent; their coefficients must be equal in Equations 5.2 and 5.3. Hence,

$$\dot{q}_j = \frac{\partial H}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial H}{\partial q_j}, \quad j = 1, 2, \ldots, n \tag{5.4a}$$

and

$$\frac{\partial H}{\partial t} = -\frac{\partial L}{\partial t}. \tag{5.4b}$$

The $2n$ first-order differential equations 5.4a and 5.4b are known as Hamilton's canonical equations of motion or, briefly, Hamilton's equations. They are also symmetrical. Only the algebraic sign distinguishes one set from the other. This symmetry opens up opportunities for more simplifications and other developments, especially in connection with complex systems, such as gases.

It is now useful to imagine the motion of a dynamic system being represented by the motion of a representative point in a $2n$-dimensional phase space. The equations representing the motion of the point in the phase space are Hamilton's canonical equations.

To apply Hamilton's equations to a dynamic system, the first step is to set up the Lagrangian L (q, q, t) of the system. The generalized momenta p_j are then obtained by using the definition $p_j = \partial L/\partial \dot{q}_j$, and the Hamiltonian H can be constructed in accordance with Equation 5.1. The equations of motion then follow by substituting H into Equation 5.4.

Both quantum mechanics and statistical mechanics are based on the Hamiltonian formalism rather than on the Lagrangian formalism, and Feynman's route to quantum mechanics and quantum field theory uses the Lagrangian approach. The Hamiltonian formalism is also especially suited for developing perturbation theory for systems in which we cannot obtain an exact solution of the equation of motion.

The properties of the Hamiltonian and, in particular, its physical significance when time is not an explicit variable is readily demonstrated. Equation 5.4a shows that time is very similar to a coordinate; if t is missing from the Lagrangian L, it is also missing from the Hamiltonian H as would be true for a coordinate. Now, consider the total time derivative of H. Because $H = H(q_j, p_j, t)$, we have for this the derivative

$$\frac{\mathrm{d}H}{\mathrm{d}t} = \sum_j \left[\frac{\partial H}{\partial q_j}\dot{q}_j + \frac{\partial H}{\partial p_j}\dot{p}_j \right] + \frac{\partial H}{\partial t}.$$

By Hamilton's equations, the contents of the square brackets become $-\dot{p}_j\dot{q}_j + \dot{q}_j\dot{p}_j$, which is obviously zero. Thus, we have

$$\dot{H} = \frac{\mathrm{d}H}{\mathrm{d}t} = \frac{\partial H}{\partial t}. \tag{5.5}$$

This says that H does not vary at all with respect to time unless it contains time explicitly. That is, if H does not explicitly depend on time t, H is a constant of the motion.

We saw in Chapter 4 that for a conservative force and if the holonomic constraints are time-independent, the Hamiltonian H of the system is equal to the total energy of the system and remains constant during the motion. This suggests a shortcut for constructing the Hamiltonian: We simply express the total energy in terms of the generalized coordinates and momenta. For a conservative system, but when the holonomic constraints are time-dependent, the Hamiltonian H is still a constant of the motion; however, it is no longer equal to the total energy of the system: that is, H = constant, but $H \neq E$.

When the Hamiltonian H depends on time t explicitly, it is no longer conserved ($H \neq$ constant), but it is still the total energy of the system ($H = E$), provided the potential energy does not depend on the velocities and the holonomic constraints are time independent.

The similarity between t and a coordinate was pointed out earlier. Students may also notice the similarity between H and a momentum because H assumes the role of a constant of the motion if its associated coordinate, that is, time, is absent from H (and so also from L). However, the association carries a minus sign: $-H$ is the momentum conjugate to t. In this way, a parallelism between Equation 5.5 and the second equation of Equation 5.4 is preserved.

Thus far, we have assumed that the dynamic system is a holonomic conservative system. Suppose we have a holonomic system, but some of the forces acting on the system are not conservative forces. In such a case, the Lagrange's equations take the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = Q_j \tag{5.6}$$

where L contains the conservative forces, and Q_j represents the forces not arising from a potential. In this case, the Hamilton's equations of motion take the form

$$\dot{q}_j = \frac{\partial H}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial H}{\partial q_j} + Q_j. \tag{5.7}$$

5.2.2 Hamilton's Equations from Hamilton's Principle

Hamilton's equations of motion can also be deduced from a variational principle, the Hamilton's principle. But as formulated in Chapter 4, Hamilton's principle refers to paths in configuration space. To extend the principle to phase space, we must modify it such that the integrand of the action integral is a function of both the generalized coordinates and momenta and their derivatives. The action integral I then could be evaluated over the trajectory of the system point in phase space, and the varied paths would be in the neighborhood of this phase space trajectory. Indeed, Hamilton's principle as stated in Chapter 4 would lead to this point if the Lagrangian L in the action integral I were expressed in terms of the Hamiltonian H by Equation 5.1, resulting in

$$\delta I = \delta \int_{t_1}^{t_2} \left(\sum_j p_j \dot{q}_j - H(p,q,t) \right) \mathrm{d}t = 0 \tag{5.8}$$

where $q_j(t)$ is a varied subject to $\delta q_j(t_1) = \delta q_j(t_2) = 0$, and p_j is varied without such endpoint restriction. In fact, there is no need for such a restriction on the variation of p_j. This point will become very clear in the process of deriving Hamilton's equations from this modified Hamilton's principle.

Carrying out the variations in the manner used previously in Chapter 4, we obtain

$$\int_{t_1}^{t_2} \sum_j \left(p_j \delta \dot{q}_j + \dot{q}_j \delta p_j - \frac{\partial H}{\partial q_j}\delta q_j - \frac{\partial H}{\partial p_j}\delta p_j \right) \mathrm{d}t = 0 \tag{5.9}$$

where the $\delta\dot{q}$'s are related to the δq's by the equation

$$\delta \dot{q}_j = \frac{\mathrm{d}}{\mathrm{d}t}\delta q_j. \tag{5.10}$$

Integrating the first term by parts, using Equation 5.10 and the end-point conditions on δq_j, we obtain

$$\int_{t_1}^{t_2}\sum_j p_j\delta\dot{q}_j dt = \int_{t_1}^{t_2}\sum_j p_j \frac{\mathrm{d}}{\mathrm{d}t}\delta q_j \mathrm{d}t$$
$$= \int_{t_1}^{t_2}\sum_j \frac{\mathrm{d}}{\mathrm{d}t}p_j\delta q_j \mathrm{d}t - \int_{t_1}^{t_2}\sum_j \dot{p}_j\delta q_j \mathrm{d}t = -\int_{t_1}^{t_2}\sum_j \dot{p}_j\delta q_j \mathrm{d}t. \tag{5.11}$$

Substituting this into Equation 5.9, we obtain

$$\int_{t_1}^{t_2}\sum_j\left[\left(\dot{q}_j - \frac{\partial H}{\partial p_j}\right)\delta p_j - \left(\dot{p}_j + \frac{\partial H}{\partial q_j}\right)\delta q_j\right]\mathrm{d}t = 0. \tag{5.12}$$

Because we view the modified Hamilton's principle as a variational principle in phase space, both δq's and δp's are arbitrary, so the coefficients of δq and δp in Equation 5.12 must vanish separately, which results in the $2n$ Hamilton's Equation 5.4:

$$\dot{q}_j = \frac{\partial H}{\partial p_j}, \quad \dot{p}_j = -\frac{\partial H}{\partial q_j}, \quad j = 1,2,\ldots n.$$

The first canonical equation usually reproduces the relationship $p_j = \partial L/\partial \dot{q}_j$, giving nothing new. It is the second or $\dot{p}_j$ equation that produces the equation of motion at once. The reader will see that Hamilton's canonical formalism is of no great assistance in getting an actual solution of the paths of the particles. But it is inestimably valuable in picturing the possible motions of a dynamic system in phase space, and as mentioned earlier, Hamilton's equations are so symmetrical, which opens up opportunities for further simplifications and other developments, especially in connection with complex systems.

Example 5.1

Solve the one-dimensional harmonic oscillator by Hamilton's method: find the Hamiltonian and the equations of motion. Identify any conserved quantities.

Solution:

The Lagrangian for a one-dimensional harmonic oscillator is

$$L = T - V = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2.$$

From the definition of p, we have

$$p = \partial L/\partial \dot{x} = m\dot{x}.$$

Then

$$T = m\dot{x}^2/2 = p^2/2m.$$

From the definition of the Hamiltonian, we have

$$H = (m\dot{x})\dot{x} - (m\dot{x}^2/2 - kx^2/) = m\dot{x}^2/2 + kx^2/2 = p^2/2m + kx^2/2$$

or, simply,

$$H = E = T + V = p^2/2m + kx^2/2.$$

Hamilton's equations give

$$\partial H/\partial p = p/m = \dot{x}, -\partial H/\partial x = kx = \dot{p} \text{ and } \partial H/\partial t = 0.$$

The first equation is the definition of momentum; the second equation is the equation of motion

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

and the last equation is the statement of conservation of energy.

Example 5.2

A particle of mass m is attracted to a fixed point O by an inverse square force $F_r = -GM/r^2 = -k/r^2$ (Figure 5.1). Solve this problem using Hamilton's equations.

Solution:

As shown in Figure 5.1, we use plane polar coordinates (r, θ). The velocity of the particle is given by

$$\vec{v} = \dot{r}\hat{e}_r + r\dot{\theta}\hat{e}_\theta.$$

Hence,

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

The Lagrangian L is

$$L = T - V = \frac{1}{2}m(\dot{r}^2 + r^2\dot{\theta}^2) - (-k/r)$$

and

$$p_r = \partial L/\partial \dot{r} = m\dot{r}, \quad p_\theta = \partial L/\partial \dot{\theta} = mr^2\dot{\theta}.$$

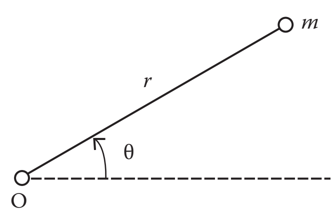


FIGURE 5.1 Particle attracted to a fixed point O.

Because L is not a function of the time, and V is not a function of the velocities, H is equal to the total energy of the particle, $H = T + V$. Thus, either from this or the definition of the Hamiltonian, we find

$$H = \frac{p_r^2}{2m} + \frac{p_\theta^2}{2mr^2} - \frac{k}{r}.$$

The Hamilton's equations are

$$\dot{r} = \partial H/\partial p_r = p_r/m, \quad \dot{\theta} = \partial H/\partial p_\theta = p_\theta/mr^2,$$

$$\dot{p}_r = -\partial H/\partial r = p_\theta^2/mr^2, \quad \dot{p}_\theta = -\partial H/\partial \theta = 0.$$

The first two equations reproduce the definitions of the momenta. The last equation indicates that p_θ is a constant of the motion as we expected because θ is a cyclic coordinate. Integrating the last equation first, and then using the second equation, we find

$$p_\theta = mr^2\dot{\theta} = \text{constant} = l.$$

Example 5.3

Find the Hamiltonian for a charged particle moving in an electromagnetic field.

Solution:

The Lagrangian L is given by Equation 4.58.

$$L = T - e\phi + e\vec{v} \cdot \vec{A}$$

and the generalized momentum is given by Equation 4.49

$$\vec{p} = m\vec{v} + e\vec{A}$$

or, in component forms,

$$p_x = m\ddot{x} + eA_x, \quad p_y = m\ddot{y} + eA_y, \quad p_z = m\ddot{z} + eA_z$$

from which we obtain

$$\ddot{x} = (p_x - eA_x)/m, \quad \ddot{y} = (p_y - eA_y)/m, \quad \ddot{z} = (p_z - eA_z)/m.$$

The Lagrangian L now becomes

$$L = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - e\phi + e(\dot{x}A_x + \dot{y}A_y + \dot{z}A_z) = \frac{1}{2m}\left(p^2 - \frac{e^2}{A^2}\right) - e\phi$$

and the Hamiltonian H is

$$H = \sum_j p_j q_j - L = p_x x + p_y y + p_z z - L$$

$$= \frac{1}{2m}(\vec{p} - e\vec{A})^2 + e\phi.$$

5.3 INTEGRALS OF MOTION AND CONSERVATION THEOREMS

5.3.1 Energy Integrals

In the preceding section, we showed that if the Hamiltonian H does not depend on time explicitly, then H is a constant of the motion:

$$H = \sum_j p_j \dot{q}_j - L = h \text{ (constant).} \tag{5.13}$$

The quantity h is called the Jacobian integral of the motion. Further, if the potential energy depends only on the coordinates, and the holonomic constraints are time independent, H is also the total energy of the system:

$$H = h = E. \tag{5.14}$$

5.3.2 Cyclic Coordinates and Integrals of Motion

A cyclic coordinate is defined as one that does not appear explicitly in L. It is obvious that a coordinate that is cyclic will also be absent from H for

$$\frac{\partial H}{\partial q_j} = \frac{\partial}{\partial q_j}\left(\sum_j p_j \dot{q}_j - L\right) = 0.$$

Combining this result with Hamilton's equations, we obtain

$$\dot{p}_j = \frac{\partial H}{\partial q_j} = 0$$

from which it follows that

$$p_j = b_j \text{ (constant).} \tag{5.15}$$

Thus, we get the same result: the generalized momentum conjugate to a cyclic coordinate is conserved, that is, it is an integral of the motion. In the following section, we will show a more general momentum conservation theorem.

When some coordinates, say, $q_1, q_2, \ldots, q_m$ ($m < n$), are cyclic, the Lagrangian of the system is of the form

$$L = L(q_{m+1}, \ldots, q_{n,}\ p_1,\ p_2, \ldots, p_n).$$

We still have to solve the problem of n degrees of freedom even though m of them correspond to m cyclic coordinates. But the Hamiltonian of the system is of the form

$$H = H(q_{m+1}, \ldots, q_n,\ p_{m+1}, \ldots, p_n,\ b_1,\ b_2, \ldots, b_m; t).$$

Thus, $(n - m)$ coordinates and momenta remain, and the problem is essentially reduced to $(n - m)$ degrees of freedom. Hamilton's equations corresponding to each of the $(n - m)$ degrees of freedom can be obtained while completely ignoring the cyclic coordinates. The cyclic coordinates themselves can be found by integrating the equations of motion $\dot{q}_j = \partial H/\partial b_j, j = 1, 2, \ldots m.$ Routh has devised a procedure that combines the advantage of the Hamiltonian formulation in handling cyclic coordinates with the Lagrangian formulation. We refer the students to the book by Goldstein about Routh's procedure.

5.3.3 Conservation Theorems of Momentum and Angular Momentum

In the preceding section, we found that, for a conservative system, the Hamiltonian is a constant if it does not depend on time explicitly. This represents the constancy of the energy of the system. By examining the Hamiltonian, it is possible to establish two other conservation theorems, namely, conservation of momentum and conservation of angular momentum, just as we did in Chapter 4 by examining the Lagrangian of the system.

We first show that if the Hamiltonian H is invariant with respect to an arbitrary infinitesimal translation of the coordinates, the total momentum of the system is conserved. For simplicity, consider the Hamiltonian depending on the difference of two coordinates $\left|\vec{r}_1 - \vec{r}_2\right|$. If the whole system is translated by a small amount $\mathrm{d}\vec{r}$, then the difference of two coordinates is not affected:

$$\left|\vec{r}_1 - \mathrm{d}\vec{r} - (\vec{r}_2 - \mathrm{d}\vec{r})\right| = \left|\vec{r}_1 - \vec{r}_2\right|.$$

Hence,

$$H(\vec{r} + \mathrm{d}\vec{r}, \vec{p}) = H(\vec{r}, \vec{p}). \tag{5.16}$$

Expanding the left-hand side, we obtain

$$H(\vec{r}, \vec{p}) + \sum_{j=1}^{3N} \mathrm{d}r_j \cdot \frac{\partial H}{\partial r_i} = H(\vec{r}, \vec{p})$$

or

$$\sum_{j=1}^{3N} \mathrm{d}r_j \cdot \frac{\partial H}{\partial r_i} = 0. \tag{5.17}$$

By Hamilton's equations, Equation 5.17 reduces to

$$\sum_{j=1}^{3N} \mathrm{d}r_j \left(-\frac{\mathrm{d}p_j}{\mathrm{d}t}\right) = 0.$$

Because $\mathrm{d}\vec{r}$ is arbitrary, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\sum_j p_j = 0$$

which is the conservation of the total momentum.

We can also show that if the Hamiltonian is invariant with respect to an arbitrary infinitesimal rotation of the coordinate axes, the total angular momentum of the system is conserved. Consider a vector $\vec{r}$ in the x_1x_2 plane; we rotate the coordinates counterclockwise through an angle θ about the x_3 axis (Figure 5.2). In the old coordinate system, the point P is at

$$x_1 = r\cos\phi, \quad x_2 = r\sin\phi$$

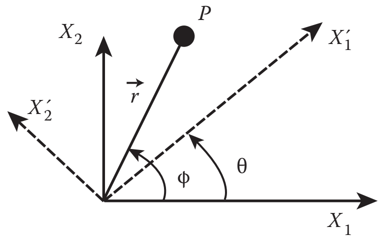


FIGURE 5.2 Infinitesimal rotation of the coordinate axes.

while in the new coordinate system, P is at

$$x_1' = r\cos(\phi - \theta), \quad x_2' = r\sin(\phi - \theta).$$

Using the trigonometric identities

$$\sin(A \pm B) = \sin A \cos B \pm \cos A \sin B$$

$$\cos(A \pm B) = \cos A \cos B \pm \sin A \sin B$$

we obtain

$$x_1' = r\cos\phi\cos\theta + r\sin\phi\sin\theta = x_1\cos\theta + x_2\sin\theta$$

and

$$x_2' = r\sin\phi\cos\theta - r\cos\phi\sin\theta = -x_1\sin\theta + x_2\cos\theta.$$

For infinitesimal rotations $\delta\theta$, $\cos\delta\theta \approx 1$, and $\sin\delta\theta \approx \delta\theta$, we obtain

$$x_1' = x_1 + x_2\delta\theta, \quad x_2' = -x_1\delta\theta + x_2$$

and so

$$x_1' - x_1 = \delta x = x_2\delta\theta, \quad x_2' - x_2 = \delta x_2 = -x_1\delta\theta \tag{5.18}$$

where $\delta\theta$ has only a z component. Equation 5.18 can be written as a vector equation:

$$\delta\vec{r} = \vec{r} \times \delta\theta. \tag{5.19}$$

All vectors will transform according to Equation 5.19 under an infinitesimal rotation about an axis.

If the Hamiltonian H is invariant with respect to an infinitesimal rotation, we then have

$$H\left([\vec{r} + \vec{r} \times \delta\theta],\ [\vec{p} + \vec{p} \times \delta\vec{p}]\right) = H(\vec{r},\ \vec{p}).$$

Expanding the left-hand side, we have

$$H(\vec{r},\vec{p}) + \sum_{j=1}^{3N}\left[(\delta\theta \times \vec{r})_j \frac{\partial H}{\partial r_j} + (\delta\theta \times \vec{p})_j \frac{\partial H}{\partial p_j}\right] = H(\vec{r},\vec{p})$$

or

$$\sum_{j=1}^{3N}\left[(\delta\theta\times\vec{r})_j\frac{\partial H}{\partial r_j}+(\delta\theta\times\vec{p})_j\frac{\partial H}{\partial p_j}\right]=0$$

from which we obtain

$$\delta\vec{\theta}\cdot\left[\sum_{j=1}^{3N}\left\{\left(\vec{r}\times\frac{\partial H}{\partial\vec{r}}\right)_j+\left(\vec{p}\times\frac{\partial H}{\partial\vec{p}}\right)_j\right\}\right]=0 \tag{5.20}$$

where we have used the vector identity $(\vec{A}\times\vec{B})\cdot\vec{C}=(\vec{C}\times\vec{A})\cdot\vec{B}$.

Using Hamilton's equations, Equation 5.20 can be further simplified:

$$\delta\vec{\theta}\cdot\left[\sum_{j=1}^{3N}\left\{(\vec{r}\times\dot{\vec{p}})_j+(\vec{p}\times\dot{\vec{r}})_j\right\}\right]=-\delta\vec{\theta}\cdot\frac{\mathrm{d}}{\mathrm{d}t}\sum_{j=1}^{3N}(\vec{r}\times\vec{p})_j=-\delta\vec{\theta}\cdot\frac{\mathrm{d}\vec{L}}{\mathrm{d}t}=0$$

where $\vec{L}$ is the total angular momentum of the system. Because $\delta\vec{\theta}$ is arbitrary, we have

$$\frac{\mathrm{d}\vec{L}}{\mathrm{d}t}=0,\quad\text{or}\quad\vec{L}=\vec{\alpha}\ \text{(a constant vector).}$$

5.4 CANONICAL TRANSFORMATIONS

As shown in the previous section, there is some advantage in using cyclic coordinates. However, in general, it is impossible to obtain more than a limited number of such coordinates by means of coordinate transformations. On the other hand, we can employ a more general class of transformations that involve both generalized coordinates and momenta. If the equations of motion are simpler in the set of new variables Q_j and P_j than in the original old set q_j and p_j, we then have a clear gain. We will not be able to consider all possible transformations but only the so-called canonical transformations that preserve the canonical form of Hamilton's equations of motion; that is, given that the q's and p's satisfy Hamilton's equations

$$\dot{q}_j=\frac{\partial H}{\partial p_j},\quad\dot{p}_j=-\frac{\partial H}{\partial q_j}$$

for some H, then the transformation

$$Q_j=Q_j(q_k,p_k,t),\quad P_j=P_j(q_k,p_k,t) \tag{5.21}$$

is canonical if and only if there exists a function K such that the time evolutions of the Q's and P's are still governed by Hamilton's equations

$$\dot{Q}_j=\frac{\partial K}{\partial P_j},\quad\dot{P}_j=-\frac{\partial K}{\partial Q_j}. \tag{5.22}$$

Here, $K(Q, P, t)$ is the new Hamiltonian that may be different from the old Hamiltonian $H(q, p, t)$.

In the old variables q_j and p_j, we can derive Hamilton's equations from the modified Hamilton's principle:

$$\delta\int_{t_1}^{t_2}\left[\sum_j p_j\dot{q}_j - H(p,q,t)\right]\mathrm{d}t = 0. \tag{5.23}$$

In the new variables Q_j and P_j, the modified Hamilton's principle

$$\delta\int_{t_1}^{t_2}\left[\sum_j P_j\dot{Q}_j - H(P,Q,t)\right]\mathrm{d}t = 0 \tag{5.24}$$

should hold. Both requirements can be satisfied if we require a relationship

$$\sum_j p_j\dot{q}_j - H = \alpha\left(\sum_j P_j\dot{Q}_j - K\right) + \frac{\mathrm{d}F}{\mathrm{d}t} \tag{5.25}$$

or

$$\sum_j p_j\mathrm{d}q_j - H\mathrm{d}t = \alpha\left(\sum_j P_j\mathrm{d}Q_j - K\mathrm{d}t\right) + \mathrm{d}F \tag{5.26}$$

where F is some function of the phase space coordinates through continuous second derivatives, and α is a constant independent of coordinates, momenta, and time, and it is related to a simple type of transformation: the scale transformation. It is always possible to select $\alpha = 1$ in Equation 5.25. We shall do so in the following discussion. The function F is termed the generating function, and it may be a function of q_j, p_j, Q_j, P_j, and t.

It is often stated in some textbooks on theoretical mechanics that because both of the variations δq_j and δQ_j vanish at the endpoints t_1 and t_2, the variation of F, δF, would also vanish at t_1 and t_2. So the total time derivative of F in Equation 5.25 will not contribute to the modified Hamilton principle. Caution should be exercised here. The vanishing of δq_j alone would not be sufficient for the vanishing of δQ_j. This follows directly by carrying out the variation in Equation 5.21:

$$\delta Q_j = \sum_k \frac{\partial Q_j}{\partial q_k}\delta q_k + \sum_k \frac{\partial Q_j}{\partial p_k}\delta p_k.$$

In order to make $\delta Q_j = 0$, the variations δq_j and δp_j must all vanish at these endpoints. This is different from the practice employed to obtain canonical equations from the modified Hamilton's principle, where q_j was varied subject to $\delta q_j(t_1) = \delta q_j(t_2) = 0$, but no such restriction was set on the variation of p_j.

Now, as $\delta q_j(t_1) = \delta p_j(t_1) = 0$ and $\delta q_j(t_2) = \delta p_j(t_2) = 0$, Equation 5.21 implies that the variations of the new variables will likewise vanish where $\delta Q_j(t_1) = \delta P_j(t_1) = 0$ and $\delta Q_j(t_2) = \delta P_j(t_2) = 0$. Thus, the total time derivative of F in Equation 5.26 will not contribute to the modified Hamilton's principle because the integral of the total time derivative is just the function evaluated at the endpoints where the variations of all the canonical variables vanish.

The function F must be some function of both the old and the new canonical variables in order for a transformation to be effected. It is obvious that we have the following four choices:

$$F_1(q, Q, t), \quad F_2(q, P, t), \quad F_3(p, Q, t), \quad \text{and} \quad F_4(p, P, t).$$

The circumstances of the problem will dictate which form is the best choice. It may be shown that F_2, F_3, and F_4 can be generated from F_1. As an example, let us consider F_1 and rewrite Equation 5.25) as

$$\sum_i (p_i\dot{q}_i - P_i\dot{Q}_i) + (K - H) = \frac{dF_1}{dt} \tag{5.27}$$

which becomes, by multiplying through by dt,

$$\sum_i (p_i dq_i - P_i dQ_i) + (K - H)dt = \sum_i \left(\frac{\partial F_1}{\partial q_i} dq_i + \frac{\partial F_1}{\partial Q_i} dQ_i\right) + \frac{\partial F_1}{\partial t} dt$$

from which it immediately follows that

$$\left.\begin{aligned} p_i = \frac{\partial F_1}{\partial q_i}, \quad P_i = -\frac{\partial F_1}{\partial Q_i} \\ K = H + \frac{\partial F_1}{\partial t} \end{aligned}\right\}. \tag{5.28}$$

When F_1 is known, Equation 5.28 gives n relations between q, p and Q, P as well as H and K. The function F_1 thus acts as a bridge between the two sets of canonical variables and is called the "generating function" of the transformation. As an example of such a generating function, we take

$$F_1 = \sum_i q_i Q_i. \tag{5.29}$$

For this special case, Equation 5.28 gives

$$Q_i = p_i, \quad P_i = -q_i, \quad \text{and} \quad K = H \tag{5.30}$$

which shows clearly that generalized coordinates and their conjugate momenta are not distinguishable, and the nomenclature for them is arbitrary. Therefore, q and p should be treated equally, and we simply call them "canonically conjugate variables" or "canonical variables."

Earlier, we mentioned that F_2, F_3, and F_4 may be generated from F_1. Now let us consider a simple example in which the independent arguments of F are to be q_i and P_i. Then, the generating function is of the type F_2. Equation 5.28 will give us help for the transition from q, Q as independent variables to q, P because

$$\frac{\partial F_1}{\partial Q_i} = -P_i.$$

This suggests that the generating function F_2 can be defined in terms of F_1 according to the relationship

$$F_2(q, P, t) = F_1(q, Q, t) + \sum_i P_i Q_i. \tag{5.31}$$

We now rewrite Equation 5.25 in terms of F_1:

$$\sum_j p_j\dot{q}_j - H = \left(\sum_j P_j\dot{Q}_j - K\right) + \frac{\mathrm{d}F_1}{\mathrm{d}t}.$$

Solving Equation 5.31 for F_1 and substituting, the last equation becomes

$$\begin{aligned}\sum_j p_j\dot{q}_j - H &= \sum_j P_j\dot{Q}_j - K + \frac{\mathrm{d}}{\mathrm{d}t}\left(F_2(q,P,t) - \sum_i Q_iP_i\right)\\ &= -\sum_i Q_i\dot{P}_i - K + \frac{\mathrm{d}}{\mathrm{d}t}F_2(q,P,t)\end{aligned}$$

where

$$\frac{\mathrm{d}F_2}{\mathrm{d}t} = \sum_i\left(\frac{\partial F_2}{\partial q_i}\dot{q}_i + \frac{\partial F_2}{\partial P_i}\dot{P}_i\right) + \frac{\partial F_1}{\partial t}\mathrm{d}t.$$

Combining the last two equations and collecting coefficients of $\dot{q}_i$, $\dot{P}_i$ we obtain the transformation equations:

$$p_i = \frac{\partial F_2}{\partial q_i} \tag{5.32a}$$

$$Q_i = \frac{\partial F_2}{\partial P_i} \tag{5.32b}$$

$$K = H + \frac{\partial F_2}{\partial t}. \tag{5.32c}$$

As an example of such a generating function, we take

$$F_2 = \sum_i q_iP_i. \tag{5.33a}$$

For this special case, Equations 5.32a through 5.32c gives

$$p_i = \frac{\partial F_2}{\partial q_i} = P_i, \quad Q_i = \frac{\partial F_2}{\partial P_i} = q_i, \quad K = H. \tag{5.33b}$$

The new and old coordinates are the same; hence, F_2 merely generates the identity transformation from Equations 5.28 and 5.32 and observing that the difference $K - H$ is the partial derivative of the generating function F with respect to time. This is also true for the other two generating functions F_3 and F_4. Thus, if the generating function does not contain the time explicitly, then $K = H$.

We also note that time t is unchanged by the transformation and that it may be regarded as an independent parameter. Because t is not directly involved, we may consider a contemporaneous variation with dt set equal to zero. Then, Equation 5.26 with $\alpha = 1$ becomes

$$\sum_i (p_i\delta q_i - P_i\delta Q_i) = \delta F \tag{5.34}$$

which is a criterion for a canonical transformation without reference to the Hamiltonian function. Thus, it is more convenient to use Equation 5.34 in testing whether or not a given transformation is canonical. The functions Q_j and P_j from Equation 5.21 are used in expressing each P_j, Q_j in terms of the old variables. If the differential form on the left-hand side of Equation 5.34 is exact and if the functions $Q_j(q, p, t)$, $P_j(q, p, t)$ are at least twice differentiable, then the given transformation is canonical, and a function $\phi(q, p, t) = F$ exists such that

$$\delta\phi = \sum_j p_j \delta q_j - \sum_j P_j \delta Q_j$$

or equivalently

$$\left.\begin{aligned} p_j - \sum_j P_j \frac{\partial Q_j}{\partial q_j} &= \frac{\partial \phi}{\partial q_j} \\ -\sum_j P_j \frac{\partial Q_j}{\partial P_j} &= \frac{\partial \phi}{\partial P_j} \end{aligned}\right\}. \tag{5.35}$$

Upon integrating to obtain $\phi(q, p, t)$, the new Hamiltonian K is found by equating the coefficients of dt in Equation 5.26. This leads to

$$K = H + \frac{\partial \phi}{\partial t} + \sum_j P_j \frac{\partial Q_j}{\partial t}. \tag{5.36}$$

It is different in form from the expression $K = H + \partial F/\partial t$. This is because different variables are held constant in the two cases when taking partial derivatives with respect to time.

We can get a better understanding of the usefulness of canonical transformations by examining a specific problem. We choose the simple harmonic oscillator. Of course, using such a powerful method as canonical transformation is scarcely necessary for such a simple problem. But an example with familiar physics and uncomplicated algebra will help us to gain a better understanding of the procedures employed.

Example 5.4: Simple Harmonic Oscillator

Consider a linear harmonic oscillator for which we have the Hamiltonian

$$H = \frac{1}{2m} p^2 + \frac{1}{2} kq^2$$

and the Hamilton equations of motion

$$\dot{q} = \partial H/\partial p = p/m, \; \dot{p} = -\partial H/\partial q = -kq.$$

Suppose that we do not know the solution to these equations and that we wish to simplify them by a canonical transformation. For the generating function, we select a function of the type

$$F_1 = \mu\, q^2 \cot Q. \tag{5.37}$$

Then, from Equation 5.28, we find that

$$p = \partial F_1/\partial q = 2\,\mu q \cot Q, \quad P = -\partial F_1/\partial Q = \mu q^2 \cos ec^2\, Q.$$

Hence,

$$p = \sqrt{4\mu P}\cos Q, \quad q = \sqrt{P/\mu}\ \sin Q. \tag{5.38}$$

Now, we can evaluate the new Hamiltonian K. Because the generating function F_1 does not depend on time explicitly, we have

$$\begin{aligned} K = H &= \frac{1}{2m}p^2 + \frac{1}{2}kq^2 \\ &= \frac{kP}{2\mu}\left(\sin^2 Q + \frac{4\mu^2}{mk}\cos^2 Q\right). \end{aligned}$$

If $\mu = \sqrt{mk}/2$, this reduces to

$$K = kP/2\mu = P\sqrt{k/m} \tag{5.39}$$

which is of a particularly simple form. Because the new coordinate Q is a cyclic coordinate, the new momentum P conjugated to Q is a constant of the motion:

$$\dot{P} = -\partial K/\partial Q = 0$$

and

$$P = \beta \text{ (a constant of the motion).} \tag{5.40}$$

Hamilton's equations of motion for the new coordinate Q gives

$$\dot{Q} = \partial K/\partial P = \sqrt{k/m}$$

from which we obtain

$$Q = \sqrt{k/mt} + \alpha \tag{5.41}$$

where α is the integration of the constant. The desired expression for p and q can be obtained by substituting Equations 5.40 and 5.41 into Equation 5.38.

You might wonder where we obtained the generating function F_1. Unfortunately, it is not always easy to find a generating function that leads to a convenient solution, and there is no simple standard procedure for doing so. Sometimes, the desired transformation can be found by an intuitive method or by solving Equation 5.28 that connects the generating function to the old and new Hamiltonians. However, there are two unknown functions in Equation 5.28: One of the two is F, which is needed to generate the coordinate transformation equations. The other is K, which is needed to provide the equations of motion. Thus, given K, we can work backward with Equation 5.28 until the generating function F is reached. A detailed discussion goes beyond our syllabus. Fortunately, the generating function for a linear harmonic oscillator $F_1 = \mu\, q^2 \cot Q$ can be constructed by the recognition that F_1 transforms oscillatory motion into uniform rectilinear motion. See Chow (1997).

It should be pointed out that, in practice, we rarely solve a dynamic problem by canonical transformations but rather study these transformations as a means of gaining a deeper understanding of the Hamiltonian formalism and of phase space.

5.5 POISSON BRACKETS

The Poisson brackets were originally introduced into the framework of theoretical mechanics in 1809 by Simeon Denis Poisson (1781–1840) in the study of planetary motion. The Poisson brackets do not assist materially in a complete solution of a system's equations of motion, but they are a great help in discussing and finding the integrals of motion. They also provide the most direct transition between theoretical mechanics and quantum mechanics (in Heisenberg's picture). Here, we shall content ourselves with the definition and some of the relevant properties of the Poisson brackets.

Given a dynamic system with a set of coordinates q_j and a set of conjugate momenta p_j, the Poisson bracket of any two dynamic variables $F(q, p, t)$ and $G(q, p, t)$, written as $[F, G]$, is defined by

$$[F, G] = \sum_j \left(\frac{\partial F}{\partial q_j} \frac{\partial G}{\partial p_j} - \frac{\partial F}{\partial p_j} \frac{\partial G}{\partial q_j} \right). \tag{5.42}$$

5.5.1 Fundamental Properties of Poisson Brackets

The following properties follow immediately from definition 5.42:

1. $[F, F] = 0$ (5.43a)
2. $[F, G] = -[G, F]$ (antisymmetry) (5.43b)
3. $[F, aG + bX] = [F, G] + [F, X]$ (linearity, for constants a and b) (5.43c)
4. $[F, GX] = [F, G]X + G[F, X]$ (5.43d)
5. $[F, [G, X]] + [G, [X, F]] + [X, [F, G]] = 0$. (Jacobi's identity) (5.43e)

where F, G, and X are functions of canonical variables and time. The Jacobi's identity (5.43e) states that the sum of the cyclic permutation of the double Poisson bracket of three functions is zero. Some of the applications of this equation are discussed later.

Any pair of functions for which the Poisson bracket vanishes, $[F, \text{G}] = 0$, are said to commute with each other.

5.5.2 Fundamental Poisson Brackets

An important Poisson bracket is that between a generalized coordinate and its conjugate momentum:

$$[q_j, p_{\text{k}}] = \sum_i \left(\frac{\partial q_j}{\partial q_i} \frac{\partial p_k}{\partial p_i} - \frac{\partial q_j}{\partial p_i} \frac{\partial p_k}{\partial q_i} \right).$$

The second term on the right-hand side vanishes, and we are left with the first term only. Now $\partial q_j/\partial q_i$ is zero unless $i = j$, in which case it is equal to unity. A similar argument holds for $\partial p_j/\partial p_i$; hence, we obtain the result

$$[q_j, p_k] = \delta_{jk}. \tag{5.44a}$$

Similarly,

$$[q_j, q_k] = 0 \quad [p_j, p_k] = 0. \tag{5.44b}$$

These are known as the fundamental Poisson brackets.

5.5.3 Poisson Brackets and Integrals of Motion

As mentioned earlier, the Poisson brackets do not assist materially in the complete solution of a system's equations of motion but are of great help in finding the integrals of motion. Let us take a close look at this. To this end, let us put G equal to the Hamiltonian H of the system. Then Equation 5.42 becomes, with the help of Hamilton's equations,

$$[F,\ H]=\sum_i\left(\frac{\partial F}{\partial q_i}\frac{\partial H}{\partial p_i}-\frac{\partial H}{\partial q_i}\frac{\partial F}{\partial p_i}\right)=\sum_i\left(\frac{\partial F}{\partial q_i}\dot{q}_i+\dot{p}_i\frac{\partial F}{\partial p_i}\right)=\frac{\mathrm{d}F}{\mathrm{d}t}-\frac{\partial F}{\partial t}$$

from which we have

$$\frac{\mathrm{d}F}{\mathrm{d}t}=[F,\ H]+\frac{\partial F}{\partial t}. \tag{5.45}$$

This result gives us an easy way to find the integrals of motion. From Equation 5.45, we see that the condition for the quantity F to be an integral of the motion becomes

$$\partial F/\partial t + [F, H] = 0. \tag{5.46}$$

If the integral of motion F does not explicitly depend on time, then Equation 5.46 reduces to

$$[F, H] = 0. \tag{5.47}$$

That is, when the integral of motion F does not contain t explicitly, its Poisson bracket with the Hamiltonian of the system vanishes.

Another important property of Poisson brackets is that, if F and G are two integrals of motion, then the Poisson bracket of F and G, $[F, G]$, is also an integral of motion; that is,

$$[F, G] = \text{constant} \tag{5.48}$$

during the motion. This is known as Poisson's theorem.

The proof of the Poisson theorem starts by noting that, because F and G are integrals of motion, we have

$$\frac{\mathrm{d}F}{\mathrm{d}t}=\frac{\partial F}{\partial t}+[F,\ H]=0,$$

and

$$\frac{\mathrm{d}G}{\mathrm{d}t}=\frac{\partial G}{\partial t}+[G,\ H]=0$$

or

$$\frac{\partial F}{\partial t}=-[F,\ H]$$

and

$$\frac{\partial G}{\partial t}=-[G,\ H]. \tag{5.49}$$

Now, from Equation 5.45, we have

$$\frac{\mathrm{d}[F,G]}{\mathrm{d}t} = \big[[F,G],H\big] + \frac{\partial[F,G]}{\partial t}$$
$$= \big[[F,G],H\big] + \left[\frac{\partial F}{\partial t},G\right] + \left[F,\frac{\partial G}{\partial t}\right]$$

which, with the aid of Equation 5.49, becomes

$$\frac{\mathrm{d}[F,G]}{\mathrm{d}t} = \big[[F,G],H\big] - \big[[F,\ H],G\big] - \big[F,\ [G,\ H]\big]$$
$$= \big[[F,G],H\big] + \big[[H,\ F],G\big] + \big[[G,\ H],F\big] = 0.$$

In the last step, Jacobi's identity was used. Hence, we have

$$[F,\ G] = \text{constant}.$$

Example 5.5

A particle of mass m is moving in a central potential V that does not depend on velocity. Find the integrals of motion.

Solution:

The kinetic energy of the particle, in spherical coordinates, is

$$T = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\ \dot{\phi}^2\right)$$

and its Lagrangian function is $L = T - V$. Because V does not depend on V, hence,

$$p_j = \partial L/\partial\dot{q}_j = \partial T/\partial\dot{q}_j$$

from which we obtain

$$p_r = m\dot{r},\quad p_\theta = mr^2\dot{\theta},\quad p_\phi = mr^2\sin^2\theta\dot{\phi}.$$

The Hamiltonian H is

$$H = \sum_j p_j\dot{q}_j - L = \frac{1}{2m}\left(p_r^2 + \frac{p_\theta^2}{r^2} + \frac{p_\phi^2}{r^2}\sin^2\theta\right) + V.$$

Because V is central, it depends on r only, so

$$[p_\theta,\ H] = 0,$$

that is, p_θ is an integral of the motion. The following Poisson bracket also vanishes:

$$\left[p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}, H\right] = \left[p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}, \frac{p_r^2}{2m}\right] + \left[p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}, \frac{1}{2mr^2}\left(p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}\right)\right]$$
$$= 0 + \frac{1}{2mr^2}\left[p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}, p_\theta^2 + \frac{p_\phi^2}{\sin^2\theta}\right] = 0$$

which shows that the quantity $p_\theta^2 + p_\phi^2/\sin^2\theta$ is also an integral of the motion.

5.5.4 Equations of Motion in Poisson Bracket Form

If we set $G = H$ and $F = q$ in the defining Equation 5.37, we have

$$\left[q_j, H\right] = \sum_k \left(\frac{\partial q_j}{\partial q_k}\frac{\partial H}{\partial p_k} - \frac{\partial q_j}{\partial p_k}\frac{\partial H}{\partial q_k}\right) = \frac{\partial H}{\partial p_j} = \dot{q}_j \tag{5.50a}$$

because $\partial q_j/\partial q_k = \delta_{jk}$ and $\partial q_j/\partial p_k = 0$ for all k. Similarly, we have

$$\left[p_j, H\right] = \dot{p}_j. \tag{5.50b}$$

These are the equations of motion in Poisson bracket form. We can also obtain them from Equation 5.40 with F replaced by the canonical variables q_j and p_j, respectively. As an example, consider a charged particle moving in an electromagnetic field. The Hamiltonian of the particle has been shown to be

$$\frac{1}{2}m(p_i - eA_i)(p_i - eA_i) + e\phi$$

where we have used the result

$$p_i = mv - eA_i.$$

Now, it is easy to verify that

$$\left[x_j, H\right] = \frac{1}{m}(p_j - eA_j) = \dot{x}_j.$$

Similarly, we have

$$\left[p_j, H\right] = \dot{p}_j$$

in accordance with Equation 5.50.

5.5.5 Canonical Invariance of Poisson Brackets

Like Hamilton's equations of motion, Poisson brackets are canonical invariants. This means that if (q, p) and (Q, P) are two canonically conjugating sets, then

$$[F, G]_{q,p} = [F, G]_{Q,P} \tag{5.51}$$

for any pair of functions F and G, where the q, p and the Q, P are related by a canonical transformation, such as Equation 5.21. The proof is straightforward, but it is tedious. We shall not pursue it here.

It is easy to show that the fundamental Poisson brackets are canonical invariants. Suppose we make the canonical transformation from a set of variables (q, p) to a new set of (Q, P). Now, any canonical transformation preserves the form of Hamilton's equations so that Equations 5.45 to 5.51 still hold for the new variables as do Equations 5.44a and 5.44b.

The Poisson-bracket description of mechanics is preserved by a canonical transformation. Equivalently, a canonical transformation can be defined as one that preserves the Poisson-bracket description of mechanics. So the fundamental Poisson brackets provide a convenient way to decide whether a given transformation of the form Equation 5.21 is canonical. In fact, they are sufficient conditions for a canonical transformation.

Example 5.6

Show that, using the Poisson brackets, the following transformation

$$Q = \sqrt{e^{-2q} - p^2}, \quad p = \cos^{-1}(pe^q)$$

is canonical.

Solution:

It is obvious that

$$[Q, Q] = 0, \text{ and } [P, P] = 0.$$

The Poisson bracket for $[Q, P]$ is

$$[Q, P] = \frac{\partial Q}{\partial q}\frac{\partial P}{\partial p} - \frac{\partial Q}{\partial p}\frac{\partial P}{\partial q}.$$

Now,

$$\frac{\partial Q}{\partial q} = \frac{-e^{-2q}}{\sqrt{e^{-2q} - p^2}}, \quad \frac{\partial Q}{\partial p} = \frac{-p}{\sqrt{e^{-2q} - p^2}}$$

$$\frac{\partial P}{\partial q} = \frac{-p}{\sqrt{e^{-2q} - p^2}}, \quad \frac{\partial P}{\partial p} = \frac{-1}{\sqrt{e^{-2q} - p^2}}.$$

Substituting these into the expression for the Poisson bracket for $[Q, P]$, we find

$$[Q, P] = 1.$$

Thus, the transformation is canonical.

5.6 POISSON BRACKETS AND QUANTUM MECHANICS

Hamiltonian dynamics has close connections with quantum mechanics. We now briefly discuss such connections. There are two formulations of quantum mechanics. One is by Heisenberg who worked with matrices, and one is by Schrödinger who developed a differential equation. Later, Dirac unified the two seemingly different systems.

We start with Heisenberg's approach. He relied on the fact that operators represent observable dynamic quantities in quantum mechanics and that operators can be represented by matrices. Operators and matrices, in general, do not commute. This non-commutability of two operators is vital to Heisenberg's work. The commutator of two operators $\underset{\sim}{A}$ and $\underset{\sim}{B}$ that represent the dynamic variables A and B is defined as

$$\left[\underset{\sim}{A},\underset{\sim}{B}\right]=\underset{\sim}{A}\underset{\sim}{B}-\underset{\sim}{B}\underset{\sim}{A}.$$

From this definition, we can see that the commutator of two operators possesses the same properties of the Poisson brackets as indicated by Equation 5.38. Dirac was the first physicist who realized this and was inspired enough to postulate that the commutator was the quantum mechanical Poisson bracket and made the connection

$$\left[A,B\right]_{P.B.}\to\frac{1}{i\hbar}\left[\underset{\sim}{A},\underset{\sim}{B}\right]$$

where $i=\sqrt{-1}$ is necessary to ensure that observed quantities are real and the $\hbar$ ($h/2\pi$, h is the Planck constant) keeps things dimensionally correct. Accordingly, once a theoretical Poisson-bracket equation of a system is written, the quantum equation may be written down directly. Thus, from Equation 5.44, we have

$$[q_k, p_j] = i\hbar\delta_{jk}, \quad [p_j, p_k] = [q_j, q_k] = 0.$$

Similarly, from Equation 5.45, we can write down the quantum equation for the operator $\underset{\sim}{O}$ as

$$\frac{\mathrm{d}\underset{\sim}{O}}{\mathrm{d}t}=\frac{\partial\underset{\sim}{O}}{\partial t}+\frac{1}{i\hbar}\left[\underset{\sim}{O},\underset{\sim}{\mathrm{H}}\right].$$

This is Heisenberg's operator equation of motion. In Heisenberg's approach (known as the Heisenberg picture of quantum mechanics), all of the time dependence is put into the operators while the state vectors of a system are time independent.

In contrast to Heisenberg picture, an observable (such as position, energy, or momentum, etc. that can be measured are called observables) is represented in the Schrödinger picture by a constant operator that does not contain the time explicitly; the time dependence enters the state vector (called the wave function or the state function) through a partial differential equation, known as the Schrödinger wave equation. Schrödinger noted that geometrical optics required extension to include wave optical effects, such as diffraction, and he argued that perhaps in analogy, mechanics could also be extended. This leads to his partial differential wave equation, which is closely connected with the Hamilton–Jacobi equation. Unfortunately, it is beyond our scope to reproduce his detailed treatment. The Schrödinger picture of quantum mechanics is known as wave mechanics.

In 1948, Feynman proposed an alternative formulation, using a Lagrangian operator and the action S. Given a system at the initial state A, theoretically, only one path is important to reach state B, the path for which the action S is least. Quantum mechanically, various paths are possible, and all we can say is that there is a certain probability we shall get to state B. Feynman suggested that all possible paths must be considered, and they can be considered from the point of view of the action, and he postulated that the probability amplitude associated with a particular path was related to the action by the expression $\exp(iS/\hbar)$. The quantum mechanical amplitude is then obtained by summing over all paths. Feynman's path integral formulation was finally laid out in clear textbook fashion in 1966 (Feynman and Hibbs 1966).

5.7 PHASE SPACE AND LIOUVILLE'S THEOREM

Hamilton's canonical formalism is of no great assistance in getting an actual solution of the trajectories of the particles, but it is very valuable in picturing the motion of a dynamic system in phase space. The motion in phase space takes place along a surface given by $H(q_1, q_2,\ldots, q_n, p_1, p_2,\ldots, p_n) =$ energy (constant). As an example, the motions in phase space for the harmonic oscillators are along the elliptical curves (Figure 5.3):

$$\frac{p^2}{2m} + \frac{x^2}{(2/k)} = E.$$

We could picture the motion of a dynamic system in an n-dimensional configuration space of the q_n's in the Lagrangian scheme. We are often interested in the motion of a group of systems having different initial conditions. At any point q_n, there are many possible curves intersecting with different values of the velocity at that point. This is not the case in the Hamiltonian scheme, where there is only one possible path through each phase point because for given $2n$ initial conditions $q_j(0)$ and $p_j(0)$, the solution of the Hamilton's equations of motion is uniquely determined.

An important property of the $2n$-dimensional phase space is described by Liouville's theorem, which states that the phase points move like an incompressible fluid. More precisely, the phase volume occupied by a set of phase points is constant.

To prove Liouville's theorem, we shall first define a few terms, such as phase volume. The product of differentials $\mathrm{d}V = \mathrm{d}q_1\ldots\mathrm{d}q_n\ \mathrm{d}p_1\ldots\mathrm{d}p_n$ may be regarded as an element of volume in phase space. If ρ is the density of representative points in phase space, then $N = \rho\ \mathrm{d}V$ is the number of representative points within the element volume $\mathrm{d}V$. We next consider an element of area in the q_j and p_j planes in phase space (Figure 5.4). The number of representative points moving across the left-hand edge into the area per unit time is

$$\rho\dot{q}_j\mathrm{d}p_j.$$

By a Taylor series expansion, the number of representative points moving out of the area through its right-hand edge is

$$\left[\rho\dot{q}_i + \frac{\partial}{\partial q_i}(\rho\dot{q}_i)\mathrm{d}q_i\right]\mathrm{d}p_i.$$

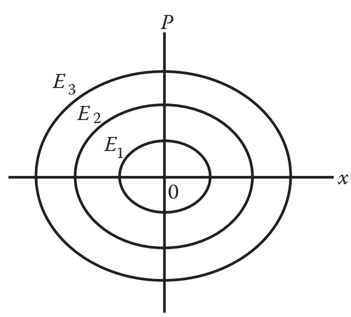


FIGURE 5.3 Phase space for the harmonic oscillators.

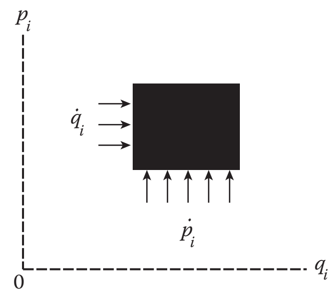


FIGURE 5.4 Element area in phase space.

Hence, the net increase in ρ in the element $\mathrm{d}q_i\,\mathrm{d}p_i$ because of flow in the horizontal direction is

$$-\frac{\partial}{\partial q_i}(\rho\dot{q}_i)\mathrm{d}q_i\mathrm{d}p_i.$$

In a similar way, we find the net gain resulting from flow in the vertical direction to be

$$-\frac{\partial}{\partial p_i}(\rho\dot{p}_i)\mathrm{d}q_i\mathrm{d}p_i.$$

The total increase in density in the element $\mathrm{d}q_i\,\mathrm{d}p_i$ per unit time is, therefore,

$$-\left[\frac{\partial}{\partial q_i}(\rho\dot{q}_i)+\frac{\partial}{\partial p_i}(\rho\dot{p}_i)\right]\mathrm{d}q_i\mathrm{d}p_i.$$

This should equal the net changes in ρ in $\mathrm{d}q_i\,\mathrm{d}p_i$ per unit time, which is

$$(\partial\rho/\partial t)\,\mathrm{d}q_i\,\mathrm{d}p_i.$$

Summing over all possible values of i, we find

$$\frac{\partial\rho}{\partial t}+\sum_i\left[\frac{\partial}{\partial q_i}(\rho\dot{q}_i)+\frac{\partial}{\partial p_i}(\rho\dot{p}_i)\right]=0$$

or

$$\frac{\partial\rho}{\partial t}+\sum_i\left(\dot{p}_i\frac{\partial\rho}{\partial p_i}+\dot{q}_i\frac{\partial\rho}{\partial q_i}\right)+\rho\sum_i\left(\frac{\partial\dot{p}_i}{\partial p_i}+\frac{\partial\dot{q}_i}{\partial q_i}\right)=0.$$

The last parentheses vanish because of Hamilton's equations of motion, leaving

$$\frac{\partial\rho}{\partial t}+\sum_i\left(\dot{p}_i\frac{\partial\rho}{\partial p_i}+\dot{q}_i\frac{\partial\rho}{\partial q_i}\right)=0.$$

But this is just the total time derivative of ρ with respect to time t, so we conclude that

$$\mathrm{d}\rho/\mathrm{d}t=0$$

or, in Poisson bracket notation,

$$\frac{d\rho}{dt} = \frac{\partial \rho}{\partial t} + \left[\rho, H\right] = 0. \tag{5.52}$$

It says that the total time rate of change of ρ, the density of representative points in phase space (briefly, phase-point density), vanishes. This result is often referred to as the principle of conservation of phase-point density.

Equation 5.52 means that to one following the motion of the phase points in an ensemble (a group of a large number of similar systems), the phase-point density does not change with the time. The phase-point density may change at any given place in phase space, but what may be called the "motional change" vanishes. Thus, if we consider a certain region of phase space containing a certain number of phase points, in the course of time, these points move in such a way as to occupy an equal phase volume at every instant, even though the shape of the phase volume may alter.

Liouville's theorem attains considerable importance for aggregates of microscopic particles where the concepts of statistical physics are important as in the statistical mechanics of particle systems, focusing properties of charged particle accelerators, in the study of the motion of electrons in the Earth's magnetic field, or even in stellar dynamics and galactic dynamics where stars or galaxies are treated as particles. It is often impractical to calculate an exact solution for such complex systems or even impossible because of a lack of complete information on the initial conditions. Statistical physics makes no attempt to obtain a complete solution for systems containing many particles. Instead, its aim is to make predictions about certain average properties at a given time by examining the motion of a group of a large number of similar systems. The group of similar systems is called an ensemble of systems and is to be seen as an intellectual construction to simulate and represent, at one time, the properties of the actual system as developed in the course of time. The statistical properties of an ensemble of systems can be specified at any time t by giving the density ρ in the phase space of system points per unit volume.

Example 5.7

As an illustration of Liouville's theorem, consider a system consisting of a large number of charged particles, each of mass m and charge e moving in a uniform electric field D. The Hamiltonian for such a particle is

$$H = p^2/2m - eDq = E \quad (E_1 < E < E_2)$$

where the electric field D is assumed to be in the direction of the positive q-axis. At a given time, say, $t = 0$, an ensemble representing such a system occupies the region A_1 shown in Figure 5.5. As time advances, these points will move to an adjacent region A_2 bounded by the momentum values p_1' and p_2', where

$$p_1' = p_1 + pt, \quad p_2' = p_2 + pt$$

so

$$p_1' - p_2' = p_1 - p_2.$$

Now, $\dot{p} = -\partial H/\partial q = eD$, and the phase "volume" occupied by the ensemble at $t = 0$ is the area

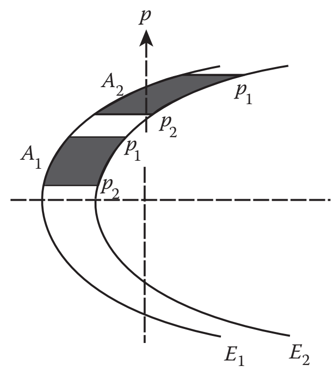


FIGURE 5.5 Motion of a "volume" in phase space.

$$A_1 = \frac{(p_2 - p_1)(E_2 - E_1)}{eD}.$$

Obviously, the phase "volume" occupied by the ensemble at time t is equal to A_1:

$$A_2 = \frac{(p_2' - p_1')(E_2 - E_1)}{eD} = \frac{(p_2 - p_1)(E_2 - E_1)}{eD} = A_1.$$

As time passes, the same set of phase points occupies an equal volume of phase space. Therefore the phase-point density must remain the same as Liouville's theorem requires. It is easy to note that the shape of the successive phase region occupied by the ensemble changes, although the phase volume is invariant.

5.8 TIME REVERSAL IN MECHANICS (OPTIONAL)

We saw earlier and in Chapter 4 that conservation of energy, momentum, and angular momentum are well established in mechanics; they are the consequences of translation invariance in time, space, and the rotational symmetry of space, respectively. These conservation laws can be extended to the whole field of physics if all the physical laws (such as electromagnetic) have the same invariance properties as in mechanics. In all experiences accumulated up to now, no evidence has been found to cause us to doubt the validity of these laws. Thus, we believe that the nature of our physical world has the fundamental symmetries represented by the invariance properties above.

Symmetries in variables other than spatial coordinates are also significant in many parts of physics, especially in subnuclear physics (or high-energy particle physics). We now take a brief look at the symmetry of physical laws under a reversal of the direction of time. The question as to how a direction of time is to be defined is in itself a difficult one. But ignoring this point for the present, we may ask whether we should expect physical laws to be invariant under the simple time reversal operation T:

$$t' = Tt = -t. \tag{5.53}$$

A little reflection shows that, apart from an intuitive feeling or attitude, there is no a priori reason for requiring all physical laws to be invariant under time reversal. In actual fact, mechanics and electromagnetic laws are invariant under the time reversal in Equation 5.53.

Upon the simple time reversal Equation 5.53, we have, for Equation 5.4a,

$$q' = q, \quad \dot{q}' = -\dot{q}, \quad p' = -p, \quad \dot{p}' = \dot{p}\,. \tag{5.54}$$

If we assume that the Hamiltonian H or Lagrangian L is invariant under Equation 5.53, then it is easily seen that Hamilton's Equation 5.4a is invariant under the time reversal (Equation 5.53).

With this connection, it is of some interest to ask about the relationship (if any) between the time reversal transformation (Equation 5.53) and the canonical transformations of mechanics. We shall see that there are two aspects of this relationship.

Now, the condition, in mechanics, for a transformation from the (q_k, p_k) to another set (Q_k, P_k) is

$$p, q \rightarrow P, Q. \tag{5.55}$$

To leave the equations of motion (Equation 5.54) invariant is usually expressed in one of the following two completely equivalent forms:

$$\sum p_k dq_k - Hdt = \sum P_k dQ_k - Kdt + dF \tag{5.56}$$

invariance of Poisson brackets

$$[F, G]_{p,q} = [F, G]_{P,Q}. \tag{5.57}$$

In terms of Poisson brackets, the canonical Equations 5.4a and 5.4b become Equation 5.45:

$$\dot{q}_k = \left[q_k, H\right]_{p,q} \quad \dot{p}_k = \left[p_k, H\right]_{p,q}. \tag{5.58}$$

For the time reversal (Equation 5.53), it is easily seen from the definition of the Poisson brackets that Equation 5.58 goes over into

$$\dot{q}'_k = \left[q'_k, H\right]_{p',q'} \quad \dot{p}'_k = \left[p'_k, H\right]_{p',q'}$$

and is invariant. On account of Equation 5.54, however, the Poisson bracket is not invariant under time reversal (Equation 5.53), but

$$[F, G]_{p,q} \rightarrow [F, G]_{p',q'}.$$

Thus, the two criteria (Equations 5.26 and 5.51) are not equivalent, and the time-reversal transformation is the difference. The Poisson bracket condition excludes the time reversal from being canonical. Time reversal would become a canonical transformation if the equation-of-motion criterion were adopted for all cases.

5.9 PASSAGE FROM HAMILTONIAN TO LAGRANGIAN

In modern physics, the Hamiltonian H is often regarded as more basic than the Lagrangian L, and so H is given in most cases. Then a natural question arises: given a Hamiltonian H, how do we find L? The motion of the system now satisfies Hamilton's equations of motion:

$$\dot{q}_i = \frac{\partial H}{\partial p_i}, \quad \dot{p}_i = -\frac{\partial H}{\partial q_i}\,. \tag{5.59}$$

To find a Lagrangian L so that the motion also satisfies Lagrange's equations of motion, we solve the first set of equations in Equation 5.59 for the p's in terms of the q's, the $\dot{q}$'s, and t. Then we write

$$L = \sum_i p_i \dot{q}_i - H$$

and express L as a function of the q's, the $\dot{q}$'s, and t. It is obvious that the essential relationship is Equation 5.1, which follows no matter from which end we start.

PROBLEMS

1. A particle of mass m moves in a plane under the influence of a central force that depends only on its distance from the origin. Write the Hamiltonian and Hamilton's equations.
2. Set up the Hamiltonian and obtain Hamilton's equations for a mass–spring system that consists of a mass m and a linear spring of spring constant k as shown in Figure 5.6.
3. A particle of mass m moves in a force field whose potential in spherical coordinates is $V = -(K \cos \theta)/r^2$. Obtain the canonical equations of motion.
4. A particle of mass m moves in a central field of attractive force that has a magnitude $(k/r^2)\, e^{-\alpha t}$, where k and α are constants, t is the time, and r is the distance of m from the force's center. Find the Lagrangian and Hamiltonian functions. Compare the Hamiltonian and the total energy of the system and discuss the conservation of energy for the system.
5. A bead of mass m slides on a frictionless wire under the influence of gravity. The wire has a parabolic shape and rotates with a constant angular velocity ω. Set up Hamilton's equations of motion.
6. A mass m is suspended by a massless spring of spring constant k and unstretched length b. The suspension point has a constant upward acceleration a_0. Gravity is acting vertically downward. Find the Lagrangian and Hamiltonian functions and obtain Hamilton's equations of motion. What is the period of motion?
7. The governor for an engine, shown in Figure 5.7, consists of two balls, each of mass m, attached by light arms to sleeves on a rotating rod. The upper sleeve is fixed to the rod, and the lower one of mass M is free to move up and down. Assume the arms to be massless and the angular velocity ω to be constant. Find the Lagrangian and Hamiltonian functions and obtain Hamilton's equations of motion describing the system. Solve for the angle θ at which the arms would stop waving, and obtain the frequency of small oscillations about the steady value.
8. As shown in Problem 4.18, the Lagrangian of a relativistic particle is given by

$$L = -m_0 c^2 \sqrt{1 - v^2/c^2} - V$$

where m_0 is the rest mass of the particle, v is its velocity, and V is not velocity dependent. Find the generalized momentum and the Hamiltonian H. It may be shown that the relativistic kinetic energy T is $-m_0 c^2 (1 - v^2/c^2)^{-1/2}$. Check that $H = T + V$.

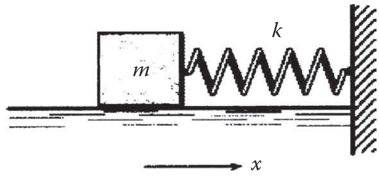


FIGURE 5.6 Mass–spring system.

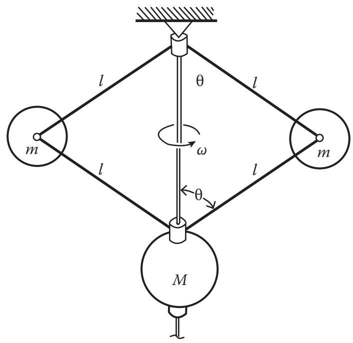


FIGURE 5.7 Engine governor.

9. Show that the following transformation is canonical:

$$Q = \sqrt{2q}e^{t}\cos p \quad P = \sqrt{2q}e^{-t}\sin p.$$

10. If $[F, G]$ is the Poisson bracket, show that
    (a) $[F_1F_2, G] = F_1[F_2, G] + F_2[F_1, G]$
    (b) $\dfrac{\partial}{\partial t}[F,G] = \left[\dfrac{\partial F}{\partial t}, G\right] + \left[F, \dfrac{\partial G}{\partial t}\right]$
    (c) $\dfrac{\mathrm{d}}{\mathrm{d}t}[F,G] = \left[\dfrac{\mathrm{d}F}{\mathrm{d}t}, G\right] + \left[F, \dfrac{\mathrm{d}G}{\mathrm{d}t}\right]$
11. Determine the Poisson brackets formed from the Cartesian components of the momentum $\vec{p}$ and the angular momentum $\vec{L} = \vec{r} \times \vec{p}$.
12. Determine the Poisson brackets formed from the components of the angular momentum $\vec{L}$.
13. Show that if the Hamiltonian and a quantity F are constants of motion, then $\partial F/\partial t$ is also a constant.
14. Consider the uniform motion of a free particle of mass m. The Hamiltonian is conserved, and there is a constant of motion $F = x - pt/m$. Show by direct computation that the constant of the motion $\partial F/\partial t$ agrees with $[H, F]$.
15. A particle of mass m and charge e is moving freely in an electromagnetic field with a velocity $\vec{v}$. The Lagrangian is

$$L = T - V + (e/c)(\vec{v} \cdot \vec{A})$$

    where T and V are the kinetic and potential energies, c is the velocity of light, and $\vec{A}$ is the vector potential with rectangular components A_x, A_y, and A_z. Find the generalized momenta conjugate to x, y, z and the Hamiltonian function.
16. Consider a large number of simple pendulums each of length b and mass m, which possess energies lying between E_1 and E_2.
    (a) Show that the Hamiltonian for each pendulum is an ellipse. Sketch the phase paths for $E = E_1$ and $E = E_2$.
    (b) The representative points corresponding to these pendulums define an area A between the two ellipses E_1 and E_2 shown in Figure 5.8. Show by direct calculation that the area A is a time invariant (i.e., $\mathrm{d}A/\mathrm{d}t = 0$).
17. An electron beam of circular cross section (radius r) is directed along the z-axis. The density of the electron beam across the beam is constant, and the momentum components transverse to the beam are distributed uniformly over a circle of radius p_0 in momentum space. If some focusing system is used to reduce the beam radius from r_0 to r_1, use the

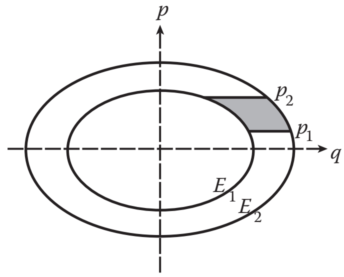


FIGURE 5.8 Area in phase space of simple pendulum.

Liouville theorem to find the resulting distribution of the transverse momentum components. What is the physical meaning of this result?

REFERENCES

Chow, T.L. Generating function for a linear harmonic oscillatore, *Eur. J. Phys.*, 18, 466, 1997.
Feynman, R.P. and Hibbs, A.R. *Quantum Mechanics and Path Integrals*, 1966, McGraw-Hill, New York.

6 Motion Under a Central Force

A central force is a force whose line of action passes through a single point or center (fixed or in motion with constant velocity) and whose magnitude depends only on the distance from the center. Forces, such as gravity and electrostatic force, are all central forces. Perhaps the first example of central force motion to be recognized was that of the planets about the sun. In old quantum theory, Bohr's hydrogen atom was described in terms of a classical two-body central picture. Certain two-body nuclear interactions, such as the scattering of alpha particles by nuclei, undoubtedly have a central character.

6.1 TWO-BODY PROBLEM AND REDUCED MASS

Consider a conservative system of two particles of mass m_1 and m_2. We shall limit ourselves to the case where the only forces acting are equal and opposite, directed along the line connecting the masses. Such a system has six degrees of freedom and, hence, six independent generalized coordinates. We can select these to be the three components of $\vec{r}_1$ and the three components of $\vec{r}_2$. Alternatively, we can choose the three components of the position vector $\vec{R}$ of the center of mass (CM) of the system and the three components of $\vec{r}\,(=\vec{r}_1-\vec{r}_2)$. Components $\vec{r}_1'$ and $\vec{r}_2'$ are position vectors of m_1 and m_2 with respect to the CM. The relationships among the various position vectors are illustrated in Figure 6.1.

The position vector $\vec{R}$ of the center mass is given by

$$\vec{R}=\frac{m_1\vec{r}_1+m_2r_2}{m_1+m_2} \tag{6.1}$$

and so in the CM system,

$$m_1\vec{r}_1'+m_2\vec{r}_2'=0$$

from which we have

$$\vec{r}_2'=-(m_1/m_2)\vec{r}_1'. \tag{6.2}$$

Substituting this into the expression for $\boldsymbol{r}$,

$$\vec{r}=\vec{r}_1'-\vec{r}_2'=\vec{r}_1'+(m_1/m_2)\vec{r}_1'=\frac{m_1+m_2}{m_2}\vec{r}_1'$$

or

$$\vec{r}_1'=\frac{m_2}{m_1+m_2}\vec{r} \tag{6.3}$$

then, from Equation 6.2, we have

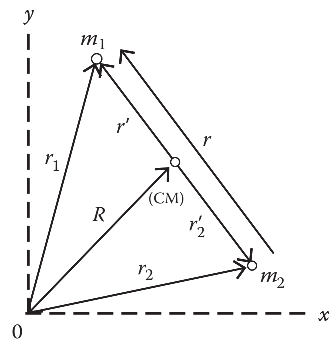


FIGURE 6.1 Coordinates of the two-body system.

$$\vec{r}_2' = -\frac{m_1}{m_1 + m_2}\vec{r}. \tag{6.4}$$

We can also express $\vec{r}_1$ and $\vec{r}_2$ in terms of $\vec{r}$:

$$\vec{r}_1 = \vec{R} + \vec{r}_1' = \vec{R} + \frac{m_2}{m_1 + m_2}\vec{r} \tag{6.5}$$

$$\vec{r}_2 = \vec{R} + \vec{r}_2' = \vec{R} - \frac{m_1}{m_1 + m_2}\vec{r}. \tag{6.6}$$

The Lagrangian L of the system is

$$L = \frac{1}{2}m_1\dot{r}_1^2 + \frac{1}{2}m_2\dot{r}_2^2 - V(r_1 - r_2) \tag{6.7}$$

or

$$L = \frac{1}{2}M\dot{R}^2 + \frac{1}{2}m_1\dot{r}_1'^2 + \frac{1}{2}m_2\dot{r}_2'^2 - V(r) \tag{6.8}$$

where $M = m_1 + m_2$. Differentiating Equations 6.3 and 6.4 and substituting the results into Equation 6.8, we obtain

$$L = \frac{1}{2}M\dot{R}^2 + \frac{1}{2}\mu\dot{r}^2 - V(r) \tag{6.9}$$

where μ is given by

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2} \tag{6.10}$$

and is called the reduced mass of the system.

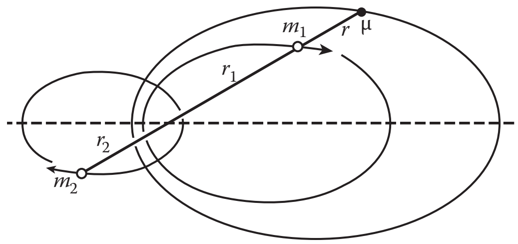


FIGURE 6.2 Orbits of the two-body system.

Note that $\vec{R}$ is cyclic; thus, the CM of the system is either at rest or moving at a constant velocity, and we drop the term involving $\vec{R}$ from the Lagrangian L. The effective Lagrangian L is now simply given by

$$L = \frac{1}{2}\mu\left|\dot{\vec{r}}\right|^2 - V(r). \tag{6.11}$$

It is the Lagrangian for a particle of mass μ moving in a central force field described by the potential function $V(r)$. Thus, we can always reduce the problem of two-body central force motion to an equivalent one-body problem moving about a fixed-force center. If the path $\boldsymbol{r} = \boldsymbol{r}(t)$ is found, we can immediately find the paths of m_1 and m_2 by means of Equations 6.3 and 6.4. The path $\boldsymbol{r}(t)$ is in the form of conic sections (see Section 6.3). As the position vectors of m_1 and m_2 are proportional to the position vector $\boldsymbol{r}$, each of these masses also describes, in the CM frame, geometrically similar conics with a common focus at the CM (Figure 6.2). The particles m_1 and m_2 are, at every instant, at the ends of a line through the common focus, and their distances from the common focus are inversely proportional to their masses.

6.2 GENERAL PROPERTIES OF CENTRAL FORCE MOTION

A central force motion has the following two general properties:

1. The total angular momentum of the system is conserved.
2. The motion lies entirely in one plane. This orbital plane is perpendicular to the fixed direction of the total angular momentum.

These are consequences of the spherical symmetry properties of the problem. Because the potential V is a function of r only, any rotation about any fixed axis passing through the center of force can have no effect on the solution; that is, the system is spherical symmetric. We now proceed to show these properties explicitly. The equation of motion is, in Newtonian form:

$$\vec{f}(r) = f(r)\hat{r} = f(r)\frac{\vec{r}}{r} = \mu\ddot{\vec{r}}.$$

Taking the cross product of both sides with $\boldsymbol{r}$, we obtain

$$\mu\ddot{\vec{r}} \times \vec{r} = f(r)\frac{\vec{r}}{r} \times \vec{r} = 0. \tag{6.12}$$

Next, differentiate the angular momentum $\vec{L} = \mu\dot{\vec{r}} \times \vec{r}$ with respect to t:

$$\frac{d\vec{L}}{dt} = \mu\left(\ddot{\vec{r}} \times \vec{r} + \dot{\vec{r}} \times \dot{\vec{r}}\right) = \mu\ddot{\vec{r}} \times \vec{r} = 0$$

from which it follows that

$$\vec{L} = \text{constant (a constant vector).}$$

That is, the angular momentum $\vec{L}$ of a body moving in a central force field is conserved.

As for the second property, it is a result of the fact that $\vec{L} \cdot \vec{r} = 0$. Thus, the radius vector $\boldsymbol{r}$ of the particle is normal to the constant angular momentum vector $\vec{L}$. This means that the motion of the particle lies in a plane perpendicular to the angular momentum $\vec{L}$.

As a consequence, the Lagrangian L can be expressed in plane polar coordinates (r, θ):

$$L = \frac{1}{2}\mu\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - V(r). \tag{6.13}$$

As θ is a cyclic coordinate, its corresponding momentum, the orbital angular momentum, p_θ is conserved:

$$p_\theta = \frac{\partial L}{\partial \dot{\theta}} = \mu r^2\dot{\theta} = l \tag{6.14}$$

where l is a constant; it is the first integral of the motion. This result has a simple geometric interpretation. As indicated in Figure 6.3, we see that the radius vector $\boldsymbol{r}$ sweeps out an area dA in a time dt, where

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

The rate at which the radius vector sweeps out an area is

$$\frac{dA}{dt} = \frac{1}{2}r^2\dot{\theta} = \frac{l}{2\mu} = \text{constant}. \tag{6.15}$$

Therefore, conservation of angular momentum implies the constancy of the area velocity. Equation 6.15 is known as Kepler's second law of planetary motion, deduced empirically by Kepler

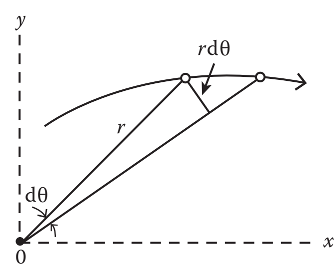


FIGURE 6.3 Area swept out by the radius vector in time dt.

in 1709 after studying Tycho Brahe's observational records of the motion of Mars for many years. The conservation of area velocity is a general result of central motion, and it is not only for the inverse square law characteristic of planetary motion.

Kepler's second law implies that a planet will move faster at a point closer to the sun than at a point farther from it, a fact that was first pointed out by Copernicus.

Because the central forces are conserved, the total energy of the system is also conserved:

$$E = \frac{1}{2}\mu\left(\dot{r}^2 + r^2\dot{\theta}^2\right) + V(r) = \text{constant}$$

which becomes, after expressing θ in terms of l from Equation 6.14,

$$\frac{1}{2}\mu\dot{r}^2 + \frac{l^2}{2\mu r^2} + V(r) = E. \tag{6.16}$$

6.3 EFFECTIVE POTENTIAL AND CLASSIFICATION OF ORBITS

Equation 6.16 is equivalent to the motion of a particle of mass μ and total energy E under a force given by the effective potential energy U, where

$$U(r) = V(r) + \frac{l^2}{2\mu r^2}. \tag{6.17}$$

In physical terms, Equation 6.17 describes the motion of a particle of mass μ under a central force as seen by an observer at 0 (Figure 6.3), who is always being rotated to face toward the particle. Such an observer is in a rotating frame of reference, and the additional potential energy term in Equation 6.17, in fact, leads to an inertial force:

$$F_c = -\frac{\mathrm{d}}{\mathrm{d}r}\left(\frac{l^2}{2\mu r^2}\right) = \frac{l^2}{\mu r^3}. \tag{6.18}$$

With the help of Equation 6.14, we can eliminate l and put F_c into a more familiar form:

$$F_c = \mu r\dot{\theta}^2.$$

This inertial force F_c is always a consideration when we look at a motion from the point of view of a rotating frame of reference. Note that, while the inertial force F_c is a central force, it depends on the angular velocity of the frame of reference relative to an inertial frame. Only in a special case, when the real force is central so that the angular momentum is constant, can we eliminate the angular velocity and derive the inertial force F_c from a potential energy.

Without solving the equation of motion for a specific force law, we can learn about the motion in the general case, using only the first integrals of the motion. Now solving Equation 6.16 for $\dot{r}$, we have

$$\dot{r} = \sqrt{\frac{2}{\mu}\left(E - V - \frac{l^2}{2\mu r}\right)} = \sqrt{\frac{2}{\mu}(E-U)}. \tag{6.19}$$

Because $\dot{r}$ must be positive, that is, $\dot{r} \geq 0$, the possible values of r for the given E and 1 are determined by the inequality

$$U(r) = V(r) + \frac{l^2}{2\mu r^2} \leq E, \tag{6.20}$$

and the maximum and minimum values of r, for which $\dot{r} = 0$, are given by the equality sign in Equation 6.20. They are the turning points or the apsidal distances of the orbit. When $U(r)$ has a minimum and is equal to this minimal value, then $\dot{r} = 0$ during the entire motion. The motion is therefore circular. No motion is possible for values of F less than this.

To illustrate the preceding points, consider motion in an attractive inverse square force field. We see that, without solving the equations of motion, the shapes of the orbits are conic sections. The force law now has the following form:

$$f(r) = -\frac{k}{r^2}, \quad V(r) = -\frac{k}{r}, \quad k > 0.$$

Hence,

$$U(r) = -\frac{k}{r} + \frac{l^2}{2\mu r^2}.$$

To sketch $U(r)$ as a function of r, all that is necessary to note is that for large r, the first term is dominant, so that

$$U(r) < 0 \text{ for large } r\text{; and } U(r) \to 0 \text{ as } r \to \infty,$$

while for small r, the second term is dominant, so that

$$U(r) \to +\infty \text{ as } r \to 0.$$

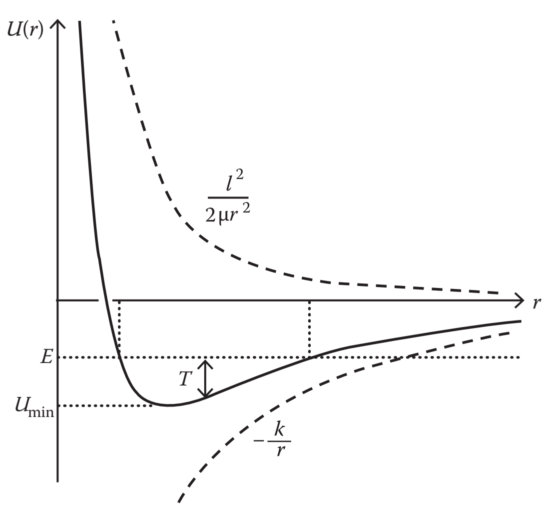


FIGURE 6.4 Effective potential $U(r)$ for attractive inverse square force.

Clearly, then, $U(r)$ must be of the form shown in Figure 6.4. The same graph can also be used to show the constant value of E, the energy for a given motion. The kinetic energy T at any distance r can then be seen at once as can the values of r for which the motion is possible.

We now consider motions for different values of E. In order to show these more clearly, we plot $U(r)$ both to the left and to the right, that is, for $\theta = 0°$ and $\theta = 180°$ (Figure 6.5). We then have four distinct cases:

(1) $E > 0$. There is a minimum radial distance but no maximum. The motion is unbounded. The particle moves toward the center of force from ∞, strikes the potential barrier at the turning point, and is reflected back to ∞. The path is a hyperbola.

(2) $0 > E > U_{\text{min}}$. There is both a maximum and a minimum radial distance at which $\dot{r} = 0$, so $V(r) + l^2/2\mu r^2 = E$. This does not mean that the particle comes to rest because the angular velocity is not zero. These maximum and minimum distances are turning points of the path. A possible shape of the path for the attractive inverse square law force is an ellipse with the focus at the center of attraction. With other forces, the orbits may not have such a simple form. During the time in which r varies from r_{max} to r_{min} and back, the radius vector turns through an angle $\Delta\theta$, which is given by

$$\Delta\theta = 2\int_{r_{\text{min}}}^{r_{\text{max}}} \frac{lr^{-2}\mathrm{d}r}{\sqrt{2\mu(E-V) - l^2/r^2}}. \tag{6.21}$$

To obtain Equation 6.21, we rewrite Equation 6.14 as

$$\mathrm{d}\theta = \frac{l}{\mu r^2}\mathrm{d}t. \tag{6.22}$$

From Equation 6.19, we have

$$\mathrm{d}t = \left[\frac{2}{\mu}\left(E - V - \frac{l^2}{2\mu r^2}\right)\right]^{-1/2}\mathrm{d}r.$$

Substituting this into Equation 6.22 and then integrating, we find

$$\theta = \int \frac{lr^{-2}\mathrm{d}r}{\left[2\mu(E-\mathrm{V}) - l^2/r^2\right]^{1/2}} + \text{constant} \tag{6.23}$$

which follows Equation 6.21.

What is the appearance of the actual orbits in space? We first notice that the radial motion is bounded between certain minimum and maximum values of r, and it is periodic with a certain period T_r. Thus, we know that the particle always moves within the area between two circles as shown in Figure 6.6, and the radial distances r_{min} and r_{max} represent turning points of the radial motion. The orbit must be such that it is tangential to both these circles because, at these points, the radial velocity is zero, but the tangential velocity cannot be zero given that the particle has angular momentum. Consider the particle after it has reached point A of Figure 6.6. It moves in as indicated, its trajectory becoming tangential to the inner circle at point C and continuing until it again becomes tangential to the outer circle at point B. The time it takes for this part of the motion is the radial period T_r. On the other hand, the radius vector is continually changing direction, always in the same sense

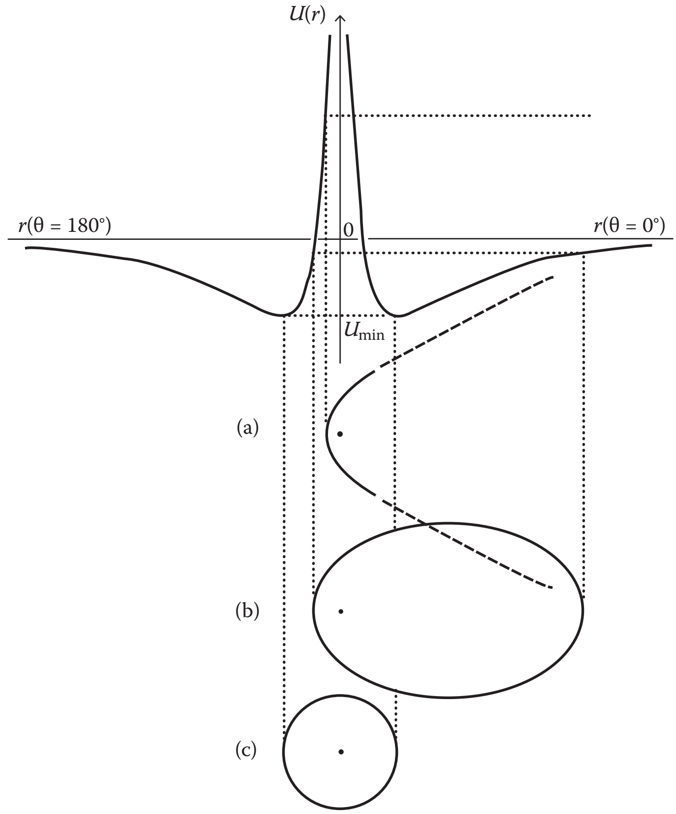


FIGURE 6.5 Schematic picture of the orbits for different. (a) Hyperbola, (b) ellipse, and (c) circle.

(either clockwise or counterclockwise) and will have turned through 2π after a characteristic period T_θ. In Figure 6.6, the line *OA* is the vector position of the particle at some instant, and the line *OA′* represents the vector position at the time T_θ later.

It is clear that the character of the orbit depends strongly on the ratio of the two periods, T_r and T_θ, of the doubly periodic motion. If the periods are commensurable (i.e., if their ratio can be expressed as the ratio of two integers), the moving particle will ultimately (after a time equal to the lowest common multiple of T_r and T_θ) find itself in exactly the same position as initially, and the orbit will thus have been closed. If the two periods happen to be exactly equal, this closure will happen after only one radial period and one increment of 2π in θ. In Figure 6.6, this would mean that the points *A* and *B* would coincide.

If the radial and angular periods are comparable but definitely different, then we have just the kind of situation shown in Figure 6.6. This corresponds to a case in which T_r is somewhat greater than T_θ, so that the radius vector rotates through rather more than 2π before r completes its variation from r_{max} to r_{min} and back again. In a situation where the path is near to being a closed curve but is, in effect, also turning (either forward or backward), we say that the orbit is precessing. The study of orbital precession is important in astronomical systems.

If the radial and angular periods are incommensurable, the orbit will never close and will eventually fill up the whole region between r_{max} and r_{min}.

In summary, we see that the path is a closed curve when the angle $\Delta\theta$ is a rational fraction of 2π: $\Delta\theta = 2\pi(m/n)$, where m and n are integers. In this case, after n periods, the radius vector of the particle will have made m complete revolutions and will occupy its original

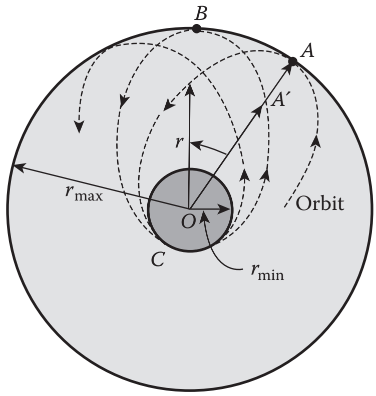


FIGURE 6.6 Schematic illustration of a rosette.

position, so that the path is closed. When the angle $\Delta\theta$ is not a rational fraction of 2π, the path has the shape of a rosette (Figure 6.6).

(3) $E = U_{min}$. This is the limiting case when the path is a circle.

(4) $E < U_{min}$. No motion is possible.

Classically, we cannot always treat the case in which the particle finds itself in a physically accessible region that includes the origin. If the particle can pass through the origin, then its angular momentum must necessarily be zero; otherwise, $\dot{\theta}$ and, hence, $\dot{r}$ become infinite. For example, consider the motion of a particle in an attractive force field that has a singularity at the origin such that the effective potential energy becomes negatively infinite as $r \to 0$:

$$U(r) = V(r) + \frac{l^2}{2\mu r^2} \to -\infty \text{ as } r \to 0.$$

Such a particle will necessarily acquire a very large velocity on approaching the origin, and therefore, its motion requires a relativistic description. Yet a classical relativistic description is also insufficient. Highly energetic particles radiate energy, and this radiation cannot be treated classically. Furthermore, potential energies that have singularities at the origin yield a force that may vary considerably over the dimensions of the body on which it acts. Bodies moving in such rapidly varying fields cannot be treated as point particles. We, therefore, do not consider such a problem.

6.4 GENERAL SOLUTIONS OF CENTRAL FORCE PROBLEM

The complete solution of the problem of the motion of a particle in a central force field can be obtained either by starting from the laws of conservation of energy and angular momentum or by writing out the differential equations of motion. We shall discuss the energy method first and then the Lagrangian analysis.

6.4.1 Energy Method

If we are interested in the path of the particle $\boldsymbol{r} = \boldsymbol{r}(t)$, we can solve Equation 6.17 for $\dot{r}$ and obtain Equation 6.19:

$$\dot{r} = \sqrt{\frac{2}{\mu}\left(E - V - \frac{l^2}{2\mu r^2}\right)}$$

from which we obtain by integration

$$t = \int \frac{\mathrm{d}r}{\sqrt{\frac{2}{\mu}(E - V) - \frac{l^2}{\mu^2 r^2}}} + \text{constant} \tag{6.24}$$

which yields the general solution of the problem in the form $t = t(r)$. An inversion of this result then gives the solution in the standard form $\boldsymbol{r} = \boldsymbol{r}(t)$.

If we are interested in the equation of path $\boldsymbol{r} = \boldsymbol{r}(\theta)$, we can use Equation 6.23:

$$\theta = \int \frac{l r^{-2}\mathrm{d}r}{\left[2\mu(E - \mathrm{V}) - l^2/r^2\right]^{1/2}} + \text{constant}.$$

Because $\dot{\theta} = l/\mu r^2$ and l is constant in time, $\dot{\theta}$ can never change sign; hence, the angle θ always varies monotonically with time. The integral can be put in a more standard form by changing the variable of integration: Let $u = 1/r$; then $\mathrm{d}u = (-1/r^2)\mathrm{d}r$, and then the above integral, Equation 6.23, becomes

$$\theta = \frac{l}{\sqrt{2\mu}} \int \frac{-\mathrm{d}u}{\sqrt{E - V - (l^2/2\mu)u^2}} + \text{constant}. \tag{6.25}$$

If the central force varies as $f(r) = \alpha r^n$, then $V = -\int f(r)\mathrm{d}r = -\alpha r^{n+1}/(n+1)$, where α and n are constants. In terms of u, V has the form $V = -\alpha u^{-(n+1)}/(n+1)$, and Equation 6.25 becomes

$$\theta = \frac{l}{\sqrt{2\mu}} \int \frac{-\mathrm{d}u}{\sqrt{E + \alpha u^{-(n+1)}/(n+1) - (l^2/2\mu)u^2}} + \text{constant} \tag{6.26}$$

where $n \neq -1$. It may be shown that when $n = 1, -2, -3$, the integral on the left may be evaluated in terms of trigonometric functions; when $n = 5, 3, 0, -4, -5, -7$, the integral may be evaluated in terms of elliptic functions.

6.4.2 Lagrangian Analysis

We next discuss the solution to the problem by writing out the equations of motion themselves via Lagrange's equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{r}} - \frac{\partial L}{\partial r} = 0 \quad \text{and} \quad \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{\theta}} - \frac{\partial L}{\partial \theta} = 0$$

where L is given by Equation 6.13.

The Lagrange's equation for coordinate θ gives the angular momentum conservation $\mu r^2\dot{\theta} = l$, and the Lagrange's equation for coordinate r gives

$$\mu\left(\ddot{r} - r\dot{\theta}^2\right) = -\partial V/\partial r = f(r). \tag{6.27}$$

We note that Equation 6.27 is Newton's second law written in the form of equations of motion. It is more convenient to use the new variable u (= $1/r$). In terms of u, we have

$$\dot{\theta} = \frac{l}{\mu r^2} = lu^2/\mu \tag{6.28}$$

and

$$\frac{dr}{dt} = \frac{d}{dt}\frac{1}{u} = -u^{-2}\frac{du}{dt} = -u^{-2}\frac{du}{d\theta}\dot{\theta} = -l\mu^{-1}\frac{du}{d\theta}. \tag{6.29}$$

If we differentiate dr/dt once more, we obtain

$$\frac{d^2r}{dt^2} = -\frac{l}{\mu}\frac{d}{dt}\left(\frac{du}{d\theta}\right) = -\frac{l}{\mu}\frac{d^2u}{d\theta^2}\dot{\theta} = -\left(\frac{lu}{\mu}\right)^2\frac{d^2u}{d\theta^2}. \tag{6.30}$$

Substituting Equations 6.28 through 6.30 into Equation 6.27, we obtain

$$\frac{d^2u}{d\theta^2} + u = -\frac{\mu}{l^2}\frac{1}{u^2}f(1/u). \tag{6.31}$$

This equation is particularly useful if we wish to find the force law when the equation of orbit $\boldsymbol{r} = \boldsymbol{r}(\theta)$ is given. To illustrate this, we consider the following simple example.

Example 6.1

A particle of mass m under the action of a force describes an orbit $r = r_0 e^{\theta}$. What is the force function that leads to this spiral orbit?

Solution:

Now,

$$u = \frac{1}{r} = \frac{1}{r_0}e^{-\theta}$$

and so

$$\frac{du}{d\theta} = -\frac{1}{r_0}e^{-\theta} \text{ and } \frac{d^2u}{d\theta^2} = \frac{1}{r_0}e^{-\theta}.$$

Then, Equation 6.31 gives

$$\frac{1}{r_0}e^{-\theta} + \frac{1}{r_0}e^{-\theta} = -\frac{\mu}{l^2}r_0^2 e^{2\theta} f(1/u)$$

from which it follows that

$$f(1/u) = -2l^2\mu^{-1}r_0^{-3}e^{-3\theta}$$

or

$$f(r) = -\frac{2l^2}{\mu r^3}\ .$$

Alternatively, we can solve the problem with the energy method. The equation of orbit indicates clearly that the particle is in a central force field. Thus, the total energy $T + V$ of the particle is conserved. Now, substituting u and $\mathrm{d}u/\mathrm{d}\theta$ into Equation 6.17, we have

$$E = \frac{l^2u^2}{2\mu} + \frac{l^2u^2}{2\mu} + V\,.$$

Solving for the potential V,

$$V = E - \frac{l^2u^2}{\mu} = E - \frac{l^2}{\mu r^2}$$

and the force $f(r)$ is, therefore,

$$f(r) = -\frac{\mathrm{d}V}{\mathrm{d}r} = -\frac{2l^2}{\mu r^3}.$$

There are many simple situations in which the required information may be determined without the details of the orbits. One such situation is the following simple case.

Example 6.2

A particle of mass m moves in a central repulsive force field that varies inversely as the cube of the radial distance: $f(r) = k/r^3$, where the constant k is positive. As shown in Figure 6.7, point O is the center of the force. The particle m moves in from a very great distance with an initial velocity v_0, and the impact parameter is b. Find the closest distance of approach of the particle to point O.

Solution:

As shown in Figure 6.7, the closest point of approach is P, at which point the radial velocity $\dot{r}$ vanishes and the direction of motion is perpendicular to OP. If the distance between points O and P is a, then the potential energy at P is

$$V_P = -\int_{\infty}^{a} \frac{k}{r^3}\,\mathrm{d}r = \frac{k}{2a^2}$$

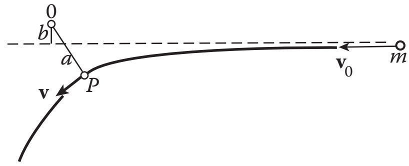


FIGURE 6.7 Particle moving in a central repulsive force field.

and the energy conservation gives the following relationship:

$$\frac{mv_0^2}{2} = \frac{mv^2}{2} + \frac{k}{2a^2} \tag{6.32}$$

where v is the velocity of the particle at point P.

In order to find the closest distance of approach a, we need a second relationship. This is provided by the conservation of angular momentum:

$$mv_0 b = mva. \tag{6.33}$$

Eliminating v from Equations 6.32 and 6.33, we obtain a:

$$a = \left(b^2 + \frac{k}{mv_0^2} \right)^{1/2}.$$

6.5 INVERSE SQUARE LAW OF FORCE

The most important type of central force is the one in which the force varies inversely as the square of the radial distance:

$$f(r) = -\frac{k}{r^2} \quad \text{or} \quad V(r) = -\frac{k}{r}$$

where the constant k is positive for an attractive force (e.g., $k = Gm_1m_2$ for attractive gravitational force), and k is negative for a repulsive force.

The equation of the orbit can be obtained from Equation 6.31. The advantage of doing this is to avoid the complicated integration. For inverse square force $f(1/u) = -ku^2$, Equation 6.31 becomes

$$\frac{d^2u}{d\theta^2} + u = \frac{\mu k}{l^2}$$

which can be rewritten as

$$\frac{d^2(u - \mu k/l^2)}{d\theta^2} + (u - \mu k/l^2) = 0.$$

Its solution is

$$u - \mu k/l^2 = b\cos(\theta - \theta_0)$$

or

$$r = \frac{\alpha}{1 + p\cos(\theta - \theta_0)} \tag{6.34}$$

where

$$\alpha = \frac{l^2}{\mu k}, \quad p = \frac{bl^2}{\mu k}$$

and b and θ_0 are integration constants.

Equation 6.34 is the equation of a conic section in polar coordinates. By definition, a conic section in polar coordinate representation is the curve traced by a point *P* (Figure 6.8) that moves so that the ratio of its distance from a fixed point *F* (the focus) to its distance from a straight line *D* (the directrix) is constant, denoted by ε, called the eccentricity of the cone. A conic section is an ellipse, parabola, or hyperbola according to whether $\varepsilon < 1$, $\varepsilon = 1$, or $\varepsilon > 1$. Appendix 4 is prepared for readers who need to review conic sections.

If *r* is the distance from the focus to *P*, and *d* is the distance of the focus from the directrix, then

$$\varepsilon = \frac{r}{d - r\cos\theta} \quad \text{or} \quad r = \frac{d\varepsilon}{1 + \varepsilon\cos\theta}. \tag{6.35}$$

Comparing this with Equation 6.34, we see that the path of the particle is a conic section of eccentricity

$$\varepsilon = p = \frac{bl^2}{\mu k}.$$

Equation 6.34 can also be obtained from Equation 7.25, which now takes the form

$$\theta = \theta_0 - \int \frac{\mathrm{d}u}{[(2\mu E/l^2 + (2\mu k u/l^2) - u^2]^{1/2}} \tag{6.36}$$

where θ_0 is a constant of integration determined by the initial condition. The integration on the right-hand side is of the standard form:

$$\int \frac{\mathrm{d}x}{(a + bx + cx^2)^{1/2}} = \frac{1}{\sqrt{-c}}\cos^{-1}\left(-\frac{b + 2cx}{\sqrt{q}}\right)$$

where $q = b^2 - 4ac$. To apply this to Equation 6.35, we set

$$a = 2\mu E/l^2, \quad b = 2\mu k/l^2, \quad c = -1, \quad q = \left(\frac{2\mu k}{l^2}\right)^2\left(1 + \frac{2El^2}{\mu k^2}\right)$$

and obtain

$$\frac{1}{r} = \frac{\mu k}{l^2}\left[1 + \sqrt{1 + \frac{2El^2}{\mu k}}\cos(\theta - \theta_0)\right]$$

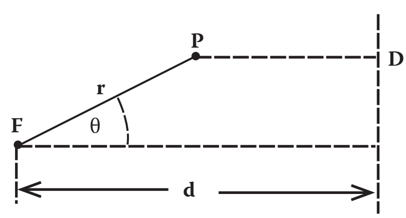


FIGURE 6.8 Conic sections.

or

$$r = \frac{\alpha}{1 + \beta \cos(\theta - \theta_0)} \tag{6.37}$$

where

$$\alpha = \frac{l^2}{\mu k} \quad \text{and} \quad \beta = \sqrt{1 + \frac{2El^2}{\mu k^2}}. \tag{6.38}$$

Comparing this with Equation 6.35, we see that Equation 6.37 is a conic section of eccentricity

$$\varepsilon = \beta = \sqrt{1 + \frac{2El^2}{\mu k^2}}. \tag{6.38a}$$

It remains to be shown that Equation 6.37 is identical to Equation 6.34. If we can show that p and β are identical in these two equations, then the two equations are also identical. To this end, we start with Equation 6.17, which now takes the form

$$\frac{1}{2}\mu \dot{r}^2 + \frac{l^2}{2\mu r^2} - \frac{k}{r} = E.$$

When r has its minimum value $r_{\min}$, the particle is turning so that $\dot{r} = 0$, and the energy equation reduces to

$$E = \frac{l^2}{2\mu r_{\min}^2} - \frac{k}{r_{\min}}.$$

Now, from Equation 6.34, we have, by setting $\cos(\theta - \theta_0) = 1$,

$$r_{\min} = \alpha/(1 + p).$$

Substituting this into the preceding equation for E and solving for p, we obtain

$$p = \sqrt{1 + \frac{2El^2}{\mu k^2}} = \beta,$$

and Equation 6.34 is identical to Equation 6.37.

The value of θ_0 merely determines the orientation of the orbit, so without loss of generality in discussing the form of the orbit, we can choose $\theta_0 = 0$. Then, the equation of the orbit (Equation 6.37) becomes

$$r = \frac{\alpha}{1 + \varepsilon \cos\theta} \tag{6.39}$$

where α and ε are given by Equation 6.38.

It is obvious that $r = r_{min}$ when cos θ is a maximum, that is, when θ = 0. Thus, the choice of $\theta_0 = 0$ corresponds to measuring θ from r_{min}. This position is called the pericenter. The position of r_{max} is called the apocenter. The corresponding terms for motion about the sun are perihelion and aphelion and, for motion about the earth, perigee and apogee. The general term for these turning points is apsides.

Let us return to the equation of orbit (Equation 6.39), a conic section. For attractive force, $k > 0$, and so α (= $l^2/\mu k$) > 0; it is then evident from Equation 6.39 that the conditions $\varepsilon < 1$ and $\varepsilon = 1$ are possible because $|\cos\theta| \le 1$ so that r will be positive. Thus, we can have parabolic and elliptic orbits. If $\varepsilon > 1$, we must have $1 + \varepsilon\cos\theta > 0$, or

$$-1/\varepsilon < \cos\theta < 1$$

and the orbit is a hyperbola with the center of force at the interior focus. In summary, the character of the orbit depends on the value of ε, according to the following scheme as shown in Figure 6.9:

$\varepsilon > 1\ E = E_h > 0$	hyperbola
$\varepsilon = 1\ E = E_p = 0$	parabola
$0 < \varepsilon < 1\ E = E_e < 0$	ellipse
$\varepsilon = 0\ E = E_c = -\mu k^2/2l^2$	circle

This classification agrees with the qualitative results of Section 6.4.

For a repulsive force, k is negative, and therefore, $\alpha < 0$. Under this condition, the denominator $(1 + \varepsilon\cos\theta)$ in Equation 6.39 is always less than zero because r must be positive. This cannot happen for $\varepsilon \le 1$. Therefore, parabolic and elliptic orbits are not possible for a repulsive inverse square force.

Now, for a repulsive force, k is negative, and therefore, we can write $V(r) = -k/r = |k|/r$. Then, by following the same procedures we employed for the attractive force case ($k > 0$), the orbit has the following form:

$$r = \frac{\alpha}{-1 + \varepsilon\cos\theta} \tag{6.40}$$

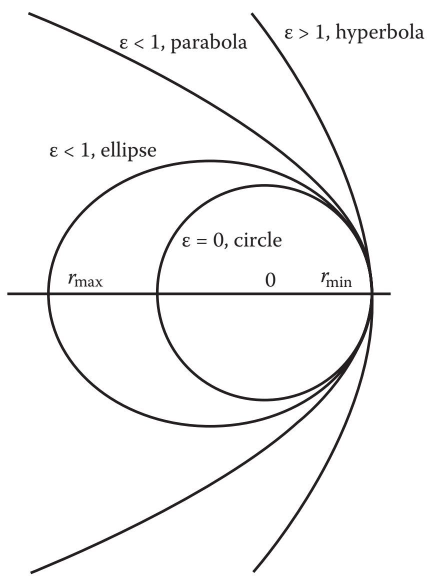


FIGURE 6.9 Family of central conics.

where α and ε are given by Equation 6.38 with k replaced by $|k|$. Equation 6.40 is similar to Equation 6.39 except with 1 replaced by −1. The orbits are therefore given by

$$r = \frac{\alpha}{\pm 1 + \varepsilon\cos\theta} \tag{6.41}$$

with the positive sign for $k > 0$ and the negative sign for $k < 0$. Relation 6.41 represents the two branches of the same hyperbola. In terms of Cartesian coordinates $x = r\cos\theta$ and $y = r\sin\theta$, Equation 6.41 becomes

$$\frac{(x - a\varepsilon)^2}{a^2} - \frac{y^2}{b^2} = 1 \tag{6.42}$$

where

$$a = \frac{\alpha}{\varepsilon^2 - 1} = \frac{|k|}{2E} \tag{6.43}$$

$$b = \frac{\alpha}{\sqrt{\varepsilon^2 - 1}} = a\sqrt{\varepsilon^2 - 1} = \frac{l}{\sqrt{2\mu E}}\,. \tag{6.44}$$

Equation 6.42 represents a hyperbola with its center C at $x = a\varepsilon$, $y = 0$, and asymptotes $y = \pm\, b(x - a\varepsilon)/a$ (Figure 6.10).

The path followed by an alpha particle in the vicinity of a nuclear charge provides an instance of a hyperbolic orbit of repulsion. The problem arises from an observation made by Geiger and Marsden in 1909. The alpha particles, which usually pass through thin foils with only very small angle deflections, on rare occasions suffer large angle deflections that are unaccountable without the supposition that a very strong force can be exerted by an atom on an alpha particle. It was this that led Rutherford to suggest his nuclear model of the atom, in which electrons, occupying a sphere of about 10^{-8} cm in radius, circle around a tiny, concentrated positive nucleus with a radius perhaps

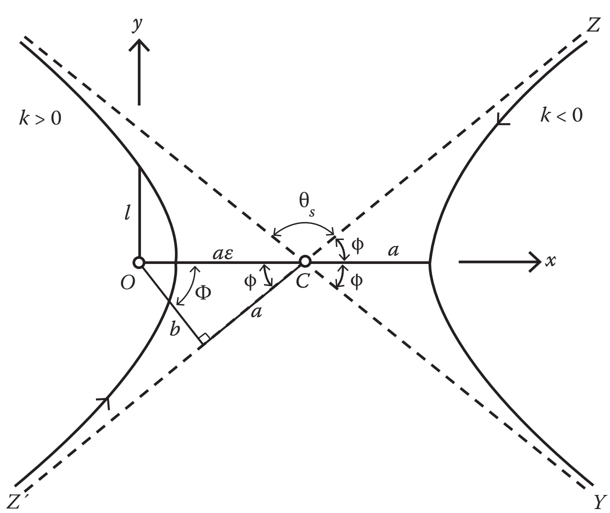


FIGURE 6.10 Hyperbolic orbit for $k < 0$ and $k > 0$. Point O is the force center.

10 times smaller. The rare scatterings of alpha particles through large angles are then to be interpreted as being a result of an alpha particle passing very close to the nucleus. Rutherford's α particle scattering will be discussed in Chapter 10 on collision and scattering.

6.6 KEPLER'S THREE LAWS OF PLANETARY MOTION

Kepler's three laws of planetary motion are the following:

1. Each planet moves in an elliptical orbit with the sun at one focus.
2. The radius vector drawn from the sun to a planet describes equal areas in equal times.
3. The square of the period of revolution about the sun is proportional to the cube of the semi-major axis of the orbit.

These laws were formulated on an empirical basis and preceded the discovery of the law of gravity by more than half a century. The second law comes about from the fact that the gravitational field of the sun is central. The other two laws are consequences of the fact that the gravitational force varies as the inverse square of the distance. The planets are trapped by the gravitational force of the sun, and their orbits are ellipses with semi-major and semi-minor axes a and b (Figure 6.11).

By the definition of eccentricity, we have

$$\varepsilon = \frac{\text{interfocal distance}}{\text{major axis}} = \frac{2(OF_1)}{2a}$$

or

$$OF_1 = a\varepsilon = OF_2.$$

An inspection of Figure 6.11 reveals that the minimum value of r occurs at $\theta = 0$:

$$r_{\min} = a - OF_2 = a(1 - \varepsilon)$$

while Equation 6.39 gives

$$r_{\min} = \frac{\alpha}{1+\varepsilon}.$$

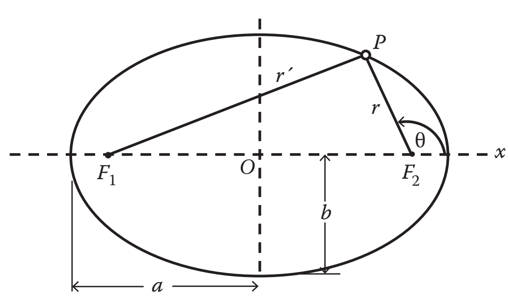


FIGURE 6.11 Elliptical orbit of a plane.

From these two equations, we have

$$\frac{\alpha}{1+\varepsilon} = a(1-\varepsilon).$$

Solving for the major axis a, we find that it depends on the energy E of the particle:

$$a = \frac{\alpha}{1-\varepsilon^2} = -\frac{k}{2E} = \frac{k}{2|E|}.\ (E_e < 0). \tag{6.45}$$

The minor axis b is given by

$$b = a\sqrt{1-\varepsilon^2} = l/\sqrt{2\mu|E|}. \tag{6.46}$$

To show Equation 6.46, we start with the fact that for any point P on the ellipse we have

$$r' + r = 2a$$

or

$$\sqrt{[x-(-a\varepsilon)]^2 + y^2} + \sqrt{[x+(-a\varepsilon)]^2 + y^2} = 2a$$

from which we have

$$\frac{x^2}{a^2} + \frac{y^2}{a^2(1-\varepsilon^2)} = 1$$

which shows that the minor axis b is given by Equation 6.46. It depends on both E and l, but the major axis a depends only on the energy E. Thus, ellipses with the same major axis have the same energy.

The period τ of the elliptic motion can be easily calculated from the following relationship:

area A of the ellipse = [rate at which the radial vector sweeps out area]
× [period of the motion]

or

$$\tau = \frac{A}{\mathrm{d}A/\mathrm{d}t}. \tag{6.47}$$

Now, the area is given by the integral

$$A = \int_0^{2\pi} \frac{r^2}{2}\mathrm{d}\theta = \int_0^{2\pi}\left(\frac{a(1-\varepsilon^2)}{1+\varepsilon\cos\theta}\right)^2 \mathrm{d}\theta = \pi a^2\sqrt{1-\varepsilon^2} = \pi ab.$$

The rate at which the radius vector from the sun to the planet sweeps out the area is dA/dt and it is given by Equation 6.15:

$$\mathrm{d}A/\mathrm{d}t = r^2\dot{\theta}/2 = l/2\mu.$$

Substituting these into Equation 6.47, we obtain

$$\tau = \frac{A}{dA/dt} = \frac{\pi ab}{l/2\mu} = 2\pi(\mu/k)^{1/2} a^{3/2}$$

or if we square τ,

$$\tau^2 = \frac{4\pi^2\mu}{k} a^3.$$

This is Kepler's third law, which he announced in 1719, 10 years after he announced the first two laws. We can put the third law in a more familiar form by noting that $k = Gm_1m_2$ and $\mu = m_1m_2/(m_1 + m_2)$. Replacing k and μ with these values, we find

$$\tau^2 = \frac{4\pi^2}{G(m_1 + m_2)} a^3. \quad (6.48)$$

The orbital eccentricities of the nine major planets are quite small, ranging from 0.007 for Venus to 0.249 for Pluto. For Earth's orbit, $\varepsilon = 0.017$ and $r_{min} = 91 \times 10^7$ mi. or 147×10^7 km and $r_{max} = 95 \times 10^7$ mi. or 153×10^7 km. The comets generally have highly elongated orbits and, therefore, large orbital eccentricities. The famous Halley's Comet has an orbital eccentricity of 0.977 with a perihelion distance of only 55×10^7 mi. ($\sim 89 \times 10^7$ km), but its aphelion is beyond the orbit of Neptune. The nonreturning type of comets have either parabolic or hyperbolic orbits.

In the special case $\varepsilon = 0$, an elliptical orbit reduces to a circle; the semi-major axis a equals the semi-major axis b and is just the radius of the circle. From Equation 6.40, we find

$$E = -k/2a \quad (6.49)$$

and, in this case, we also have

$$V = -k/a = 2E. \quad (6.50)$$

Then the kinetic energy of the particle is

$$T = E - V = k/2a = -E. \quad (6.51)$$

6.7 APPLICATIONS OF CENTRAL FORCE MOTION

6.7.1 Satellites and Spacecraft

The orbits of satellites and spacecraft are an interesting application of central force motion. First, we use Equation 6.51, the circular-orbit relationship, to resolve the so-called "satellite paradox." The effect of the slight atmospheric drag on a satellite in a circular orbit at a height of several hundred kilometers above the earth is to increase the speed of the satellite; this seems to contradict our intuition. Is there a paradox? Actually, there is no paradox. The atmospheric drag converts mechanical energy into heat; hence, the energy E of the satellite decreases, and so, by Equation 6.50, the kinetic energy T increases and the satellite speeds up. The increase in kinetic energy is accompanied by a decrease in the radius a of the orbit. Thus, as the radius of the orbit of the satellite decreases, the satellite speeds up. The atmosphere is quite thin at an altitude of several hundred kilometers, and its drag on a satellite is not very strong; the satellite makes many orbits before its orbital height is changed appreciably.

Next, we determine the orbit parameters of a satellite from conditions at its closest approach. From the following two relationships:

$$r_{min} = \alpha/(1 + \varepsilon) \text{ and } \alpha = l^2/\mu k$$

we find

$$\varepsilon = \frac{\alpha}{r_{min}} - 1 = \frac{l^2}{\mu k r_{min}} - 1.$$

If v_0 is the velocity of the satellite at r_{min}, then by the conservation of angular momentum, we have

$$l = \mu r^2\dot{\theta} = \mu r_{min} v_0$$

and the eccentricity ε can now be expressed as

$$\varepsilon = \frac{\mu r_{min} v_0^2}{k} - 1. \tag{6.52}$$

If we denote the quantity $k/\mu r_{min}$ by v_c^2, we see that $v_0 = v_c$ for a circular orbit ($\varepsilon = 0$). And for $v_0 \geq v_c$, the eccentricity ε can be written as

$$\varepsilon = (v_0/v_c)^2 - 1$$

and the equation of the orbit becomes

$$r = r_{min} \frac{(v_0/v_c)^2}{1 + [(v_0/v_c)^2 - 1]\cos\theta}. \tag{6.53}$$

The velocity at any point on the elliptical orbit can be calculated from $E = T + V$ and Equation 6.3:

$$E = mv^2/2 + (-k/r) = -k/2a \tag{6.54}$$

where a is the semi-major axis of the ellipse. Because $k = GmM$ and $GmM/R^2 = mg$, we get

$$k = mgR^2$$

where R is the radius of the Earth, and g is the gravitational acceleration at Earth's surface. In terms of g and R, the magnitude of the velocity of the satellite at any point on its orbit is then given by

$$v = \sqrt{\left(\frac{2}{r} - \frac{1}{a}\right) g R_e^2} = v_{esp}\sqrt{\left(\frac{1}{r} - \frac{1}{2a}\right) R_e} \tag{6.55}$$

where $v_{esp} = (2gR)^{1/2}$ is the escape velocity from the Earth. The velocity reaches its maximum value at perigee, and at apogee, the velocity reaches its minimum value.

A satellite in a circular orbit ($\varepsilon = 0$) of radius r_c around the earth can be economically sent into an elliptical orbit with a distance of closest approach r_c by a sudden blast of rockets at the proposed perigee. A rocket blast at perigee increases the velocity perpendicular to the radius vector $\boldsymbol{r}$ without change in the $\boldsymbol{r} \cdot \boldsymbol{v} = 0$ condition as a turning point on the orbit. The increase in velocity is accompanied by an increase both in energy E ($= V + mv^2/2$) and in angular momentum, so ε becomes positive, and the orbit is thereby changed from circular to elliptical. The procedure can be used in reverse to convert an elliptical orbit to a circular one by firing retrorockets at perigee. This technique was followed in the Apollo moon missions. The Apollo spacecraft arrived in an elliptical lunar orbit. After two orbits, a retrograde rocket blast was used to convert the orbit to a circular one prior to the landing on the lunar surface.

Example 6.3

A spacecraft in a circular orbit of radius r_c around the Earth was inserted into an elliptical orbit by firing a rocket (Figure 6.12). If the speed of the spacecraft was increased 10% by the sudden blast of the rocket motor, what is the equation of the new orbit? Also calculate the apogee distance.

Solution:

If v_c is the circular orbital speed, the speed right after the sudden firing of the rocket motor is

$$v_0 = v_c + 0.1v_c = 1.1v_c.$$

The equation of the new orbit is given by Equation 6.53 with r_{min} replaced by r_c:

$$r = \frac{r_c(1.1)^2}{1+[(1.1)^2-1]\cos\theta} = \frac{1.21r_c}{1+0.21\cos\theta}.$$

At the apogee, $\theta = \pi$, so $\cos\theta = -1$ and $r_{max} = r_1$; thus, the apogee distance r_1 is

$$r_1 = \frac{1.21r_c}{1-0.21} = 1.53r_c.$$

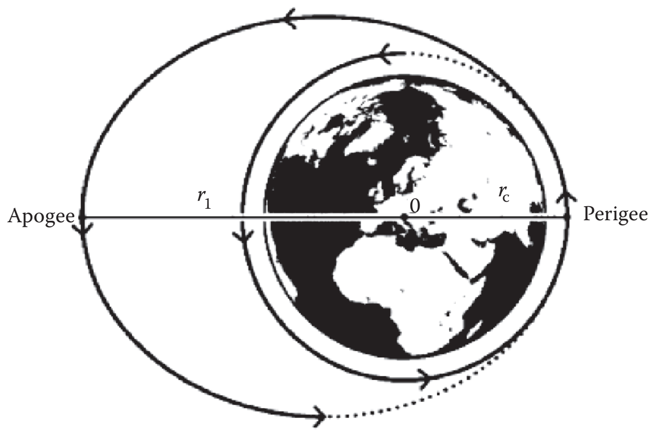


FIGURE 6.12 Orbit of a spacecraft in space.

Example 6.4

A satellite is launched into orbit under the following conditions: When all fuel has been consumed, the satellite will be 743 km from Earth's surface and will have a velocity of 9144 m/s perpendicular to the perigee distance (Figure 6.13). What is the eccentricity of the orbit? Also calculate the apogee distance.

Solution:

The orbit followed by the satellite satisfies the general equation of the orbit:

$$r = \frac{\alpha}{1 + \varepsilon\cos\theta}$$

where

$$\alpha = l^2/\mu k,\ l = \mu r^2\dot{\theta},\ k = GmM,\ \mu = mM/(m+M) \cong m,$$

$$m = \text{mass of the satellite},\ M = \text{mass of the Earth} = 5.977\times 10^{27}\,\text{g}.$$

We now rewrite the general equation of the orbit in the following form:

$$r = \frac{\left(r^2\dot{\theta}\right)^2/GM}{1+\varepsilon\cos\theta} = \frac{C^2/GM}{1+\varepsilon\cos\theta}$$

where $C = r^2\dot{\theta}$, and $G = 7.778 \times 10^{-8}$ dyn cm^2 g^{-2}.

The initial conditions are shown in Figure 6.13. Because θ = 0 at the perigee, cos θ = 1;

$$v = r\dot{\theta}$$

and so

$$C = r^2\dot{\theta} = rv$$

$$= (6373 + 643)\times 1000 \times 9144 = 64.154\times 10^9\,\text{m}^2/\text{s}.$$

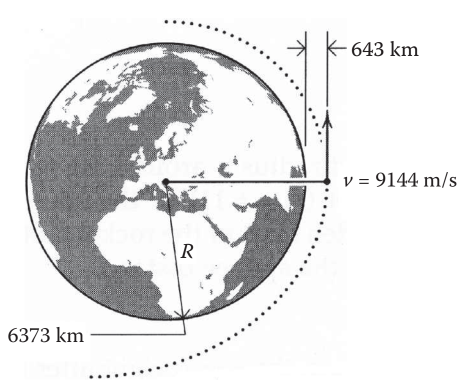


FIGURE 6.13 Satellite in orbit.

Substituting these values into the general equation of the orbit, we find

$$\varepsilon = 0.477.$$

The apogee distance can be calculated from the general equation of the orbit, by setting $\theta = \pi$, and so $\cos\theta = -1$:

$$r_1 = \frac{C^2/GM}{1-\varepsilon} = 19{,}746 \text{ km}.$$

6.7.2 Communication Satellites

We are using two types of communication satellites for transmitting signals from one part of the Earth's surface to another: the passive system (Figure 6.14a) and the active system (Figure 6.14b). In the passive system, the signals are reflected from the transmitting station T to the receiving station R. But in the active system, the signal from T is received at r on the surface of the satellite and undergoes amplification by equipment on board the satellite, after which it is transmitted at t to the ground receiving station R. The active system is of greater practical importance. In either case, the satellites may be in motion or at rest relative to the Earth. If the satellite is at rest relative to the Earth, it revolves around the Earth in a circular orbit at a unique height H about the Earth's surface. Height H can be expressed in terms of the Earth's radius R and the period P of the satellite in the following form:

$$H = (GM)^{1/3}(P/2\pi)^{2/3} - \mathrm{R} \tag{6.56}$$

where M is the mass of the Earth. To derive this relationship, we start with Equation 6.52. For a circular orbit, the eccentricity $\varepsilon = 0$, so Equation 6.52 gives the orbital speed v_0 of the satellite:

$$v_0 = (k/\mu)^{1/2}(R + H)^{-1/2}.$$

Next, we express the period P of the satellite in terms of v_0:

$$P = \frac{2\pi(R+H)}{v_0} = 2\pi(R+H)^{3/2}(\mu/k)^{1/2}$$

where $\mu = mM/(m + M)$ and $k = GmM$. Solving for H, we obtain Equation 6.56.

The period P is one day ($= 8.74 \times 10^4$ s), $M = 5.977 \times 10^{24}$ kg, $R = 7378$ km, and $G = 7.77 \times 10^{-11}$ m^3 g^{-1} s^2; substituting these values into Equation 6.56, we obtain

$$H = 3.58 \times 10^4 \text{ km}.$$

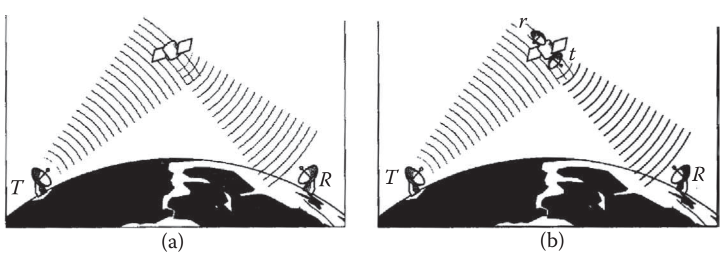


FIGURE 6.14 Communication satellites. (a) Passive system and (b) active system.

6.7.3 Flyby Missions to Outer Planets

Flyby missions to the outer planets provide another interesting practical application of the inverse square theory. The energy requirements would be far beyond the capabilities of rocket technology today if it were not possible to utilize gravity boost along the way.

We first estimate the minimum launch speed from the Earth for a spacecraft getting into orbit to an outer planet. For a spacecraft of mass m at distance r from the sun to completely escape the sun's gravitational pull, the minimum launch speed is given by

$$E = 0 = \frac{mV_{\text{esp}}^2}{2} - \frac{GmM_s}{r}. \tag{6.57}$$

Solving for V_{esp}, we find

$$V_{\text{esp}} = \sqrt{2GM_s/r} = 42 \text{ km/s}$$

where r is the distance of the Earth from the sun, and M_s is the sun's mass. However, if the launch is made in the direction of the Earth's orbital speed about the sun (Figure 6.15), the minimum launch speed for getting into orbit can be reduced significantly. The Earth's orbital speed V_e can be determined by equating centripetal and gravitational force:

$$GmM_s/r^2 = mV_e^2/r$$

from which we find

$$V_e = \sqrt{GM_s/r} = \frac{1}{\sqrt{2}} V_{\text{esp}} = 30 \text{ km/s}.$$

Thus, by making the launch in the direction of Earth's orbital velocity, the initial velocity required for escape from the sun's gravitational pull can be reduced to 12 km/s. Note that the spacecraft must have additional initial velocity to escape from the Earth's gravitational pull. Moreover, the launch must be so timed that the spacecraft and the outer planet will meet at aphelion on the spacecraft

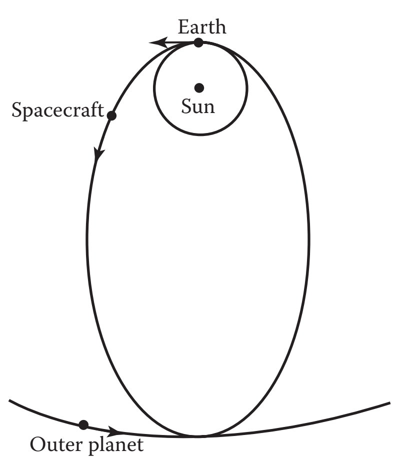


FIGURE 6.15 Flyby missions to the outer planets.

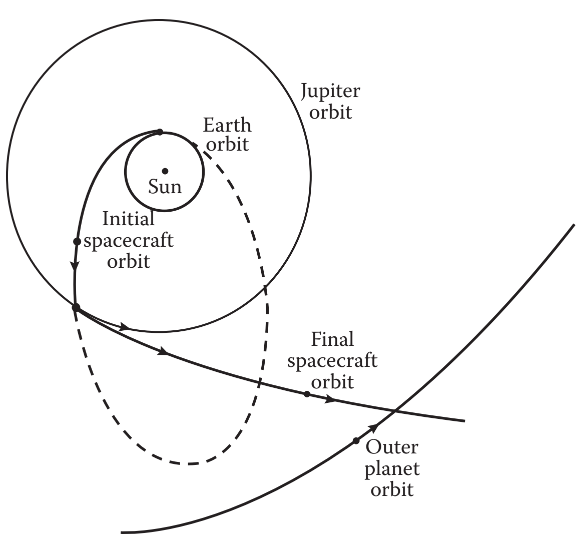


FIGURE 6.16 Orbits illustrating the gravity-assistance effect.

orbit as illustrated in Figure 6.16. The orbit of the spacecraft is an elliptical one about the sun with perihelion at the Earth's orbit, and the planet and the spacecraft meet at the aphelion. The equation of the orbit is given by Equation 6.39:

$$r = \frac{\alpha}{1 + \varepsilon \cos\theta}$$

where α and ε are determined from the maximum and minimum values of r. For planet Uranus,

$$r_{min} = 1\ AU \text{ (at Earth)},\ r_{max} = 19.2\ AU \text{ (at Uranus)}$$

where AU is the astronomical unit that is the mean Earth–sun distance (~93 × 10^7 mi. or 1.5 × 10^8 km). Thus,

$$r_{min} = \frac{\alpha}{1+\varepsilon} = 1\ AU, \quad r_{max} = \frac{\alpha}{1-\varepsilon} = 19.2\ AU.$$

Solving for α and ε, we find ε = 0.9 and α = 1.9. The orbit of the spacecraft to Uranus is

$$r = \frac{1.9}{1 + 0.9\cos\theta}$$

with the semi-major axis given by

$$a = (r_{min} + r_{max})/2 = 10.1\ AU.$$

The velocity of the spacecraft at any point on its orbit can be found by using Equation 6.49 or Equation 6.51. The time T at which the spacecraft reaches Uranus at the aphelion is just half of the period given by Kepler's third law (Equation 6.43):

$$T = \frac{\tau}{2} = \frac{1}{2}\frac{2\pi}{\sqrt{GM_s}}a^{3/2} = \frac{1}{2}(10.1)^{3/2} \text{ years} \approx 16 \text{ years}.$$

This is quite a long flight time. Fortunately, it can be considerably shortened by means of gravitational assists as the spacecraft swings by Jupiter, known as the "slingshot effect" or "gravitational whiplash." Jupiter and its gravity field are moving around the sun at a speed of approximately 1300 m/s, and any probe passing behind the planet will be accelerated by this moving gravity field much as a surfer is pushed forward by a wave. The energy does not come from the gravitational field but from the kinetic energy of the moving planet, which is slowed by the tiniest amount in its orbit, causing it to drop ever so slightly closer to the sun. In order to utilize the gravitational assistance of Jupiter, we have to select the launch time such that the spacecraft will pass close to Jupiter as it flies by the planet. Because the duration of the encounter is much shorter than Jupiter's period of revolution, we can neglect the slight change in the direction of Jupiter's velocity during the encounter. Let $\vec{p}$ and $\vec{p}'$ be the spacecraft's momenta in the sun-centered inertial frame just before and just after the Jupiter encounter, and Jupiter is moving with velocity $\vec{v}_J$ in the sun-centered inertial frame. Furthermore, $\vec{p}_J$ and $\vec{p}'_J$ are the corresponding quantities in the Jupiter rest frame. Then Galilean transformation gives

$$\vec{p} = \vec{p}_J + m_J\vec{v}_J \quad \text{and} \quad \vec{p}' = \vec{p}'_J + m_J\vec{v}_J. \tag{6.58}$$

We can see that the change in momentum during the encounter is the same in both frames:

$$\Delta\vec{p} \equiv \vec{p}' - \vec{p} = \vec{p}'_J - \vec{p}_J \equiv \Delta\vec{p}_J.$$

But the change in kinetic energy depends on the frame of reference in which the spacecraft is observed. In the sun-centered setting, we have

$$\Delta(KE) = \frac{p'^2}{2m} - \frac{p^2}{2m}$$

which can be rewritten with the help of Equation 6.58 as

$$\begin{aligned}\Delta(KE) &= \frac{(\vec{p}'_J + m_J\vec{v}_J)^2}{2m} - \frac{(\vec{p}_J + m_J\vec{v}_J)^2}{2m} \\ &= \left(\frac{p'^2_J}{2m} - \frac{p^2_J}{2m}\right) + \vec{v}_J\cdot(\vec{p}'_J - \vec{p}_J) = \Delta(KE)_J + \vec{v}_J\cdot\Delta\vec{p}_J.\end{aligned}$$

Now, the encounter with Jupiter is elastic in the Jupiter-centered frame of reference, and so $\Delta(KE)_J = 0$, and $\Delta(KE)$ reduces to

$$\Delta(KE) = \vec{v}_J\cdot\Delta\vec{p}_J = \vec{v}_J\cdot(\vec{p}'_J - \vec{p}_J).$$

That is, in the sun-centered frame, the encounter with Jupiter produces a gain in the kinetic energy of the spacecraft. This energy gain by the spacecraft is at Jupiter's expense. Because the initial orbit fixes the value $\vec{v}_J\cdot\vec{p}_J$, the gain in kinetic energy $\Delta(KE)$ is maximized when $\vec{p}'_J$ is parallel to $\vec{v}_J$. Before the spacecraft reaches the vicinity of Jupiter, the orbit of the spacecraft is governed by the sun's gravity. However, as the spacecraft approaches Jupiter, Jupiter's gravitational field essentially governs its orbit relative to Jupiter. A detailed analysis of the kinematics of the spacecraft orbit near Jupiter in the Jupiter rest frame will reveal that the final spacecraft velocity can be almost twice its initial velocity.

6.8 NEWTON'S LAW OF GRAVITY FROM KEPLER'S LAWS

We saw earlier that Kepler's three laws can be deduced from Newton's law of gravity. Historically, Kepler's three laws were deduced empirically from Tycho Brahe's observational records of Mars' motion. It was Newton who employed these laws 60 years later to show that the law of gravity between the planets and the sun must be that of the inverse square. We now proceed to show this.

Kepler's second law, $\mathrm{d}A/\mathrm{d}t = r^2\dot{\theta} = l/2\mu$, gives assurance that the gravitational force between the sun and the planets is a central one. Differentiating the second law with respect to time can easily show this:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{r^2\dot{\theta}}{2}\right) = \frac{ra_\theta}{2} = 0$$

where we have made use of the following relationship:

$$a_\theta = r\ddot{\theta} + 2\dot{r}\dot{\theta} = \frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}t}\left(r^2\dot{\theta}\right).$$

Thus, the transverse component of the force acting on the planet is zero, and this means that the gravitational force is a central one. To determine the force law, we start with Equation 6.31, which can be written as

$$\mu\left[\frac{\mathrm{d}^2}{\mathrm{d}\theta^2}\left(\frac{1}{r}\right) + \frac{1}{r}\right] = -\frac{\mu^2}{l^2}r^2 f(r)$$

in which r is given by Kepler's first law (Equation 6.39), and $c = l/2\mu$, m = mass of the planet, and $f(r)$ is the gravitational force of the sun acting on the planet.

Substituting Equation 6.39 into Equation 6.31 and solving for $f(r)$, we find

$$f(r) = -\frac{l^2}{\alpha\mu}\frac{1}{r^2} = -\frac{k}{r^2} \tag{6.59}$$

where $k = l^2/\alpha\mu$, and the negative sign indicates that the force is an attractive one.

Kepler's third law will assure that $k = GmM$. Equation 6.47 gives the period τ of the elliptical motion

$$\tau = \frac{\pi ab}{\mathrm{d}A/\mathrm{d}t} = \frac{\pi ab}{l/2\mu}.$$

But from Equations 6.43 and 6.44, we find that $b^2 = \alpha a$, so

$$\tau^2 = (4\pi^2\mu^2\alpha/l^2)a^3 = 4\pi^2\mu k a^3$$

where we have made use of the relation $\alpha = l^2/\mu k$. Comparing this expression for τ^2 with Equation 6.48, we obtain the desired result

$$k = GmM.$$

6.9 STABILITY OF CIRCULAR ORBITS (OPTIONAL)

It was pointed out in Section 6.3 that, for all central forces, when the effective potential $U(r)$ has a minimum and the total energy E is equal to this minimal value, the orbit will be circular. A circular orbit is also possible when $U(r)$ has a maximum value. Obviously, it is unstable. In general, a circular orbit is allowed for any central, attractive potential because the centripetal force required for circular motion can be supplied by the potential. However, such orbits are not necessarily stable. What do we mean by a "stable orbit"? If a particle moving in circular orbit returns to the circular path or oscillates about it when the disturbance is removed, the circular orbit is said to be stable. On the other hand, if the particle, when disturbed, deviates from the circular path more and more as time passes, the orbit is then unstable. What is the stability of a circular orbit? To find this condition, we start with the effective potential $U(r)$:

$$U(r) = V(r) + \frac{l^2}{2\mu r^2}$$

which will have a minimum at some value $r = \rho$ provided that

$$\left(\frac{\mathrm{d}U}{\mathrm{d}r}\right)_{r=\rho} = 0 \quad \text{and} \quad \left(\frac{\mathrm{d}^2U}{\mathrm{d}r^2}\right)_{r=\rho} > 0. \tag{6.60}$$

From the first equation of Equation 6.60, we find

$$\left(\frac{\mathrm{d}V}{\mathrm{d}r}\right)_{r=\rho} - \frac{l^2}{\mu\rho^3} = 0$$

or

$$f(\rho) = -\frac{l^2}{\mu\rho^3} \tag{6.61}$$

which is the condition for a circular orbit. From the second equation, we obtain

$$f'(\rho) - 3\left(\frac{l^2}{\mu\rho^4}\right) < 0$$

which yields

$$f'(\rho) + \frac{3}{\rho} f(\rho) < 0. \tag{6.62}$$

This is the condition of stability for a circular orbit. However, if we want to know the frequency of oscillation about the stable circular orbit, the stability condition (Equation 6.62) cannot help us at all, and we have to use a different general procedure for this purpose. For convenience, we write the force as

$$f(r) = -\mu g(r). \tag{6.63}$$

For attractive central force, $f(r) < 0$, so $g(r) > 0$ always. In terms of $g(r)$, Equation 6.27 becomes

$$\ddot{r} - \frac{l^2}{\mu^2 r^3} = -g(r) \tag{6.64}$$

where we have eliminated the term $r\dot{\theta}^2$ by using Equation 6.14, the equation of conservation of angular momentum. For a circular orbit, $\ddot{r} = 0$; then, calling ρ the radius of the circular orbit, Equation 6.64 reduces to Equation 6.62, the condition for circular orbit:

$$l^2/\mu^2\rho^3 = g(\rho).$$

Now, consider a particle initially in a circular orbit of radius ρ and slightly disturbed in the radial direction so that

$$\rho + x = r$$

where x is the small radial deviation. In terms of r, the equation of the orbit is in the form of Equation 6.64. Now, $\dot{r} = \dot{x}$ and $\ddot{r} = \ddot{x}$, so Equation 6.64 becomes

$$\ddot{x} - \frac{l^2}{\mu^2\rho^3(x/\rho + 1)^3} = -g(x + \rho)$$

where $x/\rho \ll 1$. Expanding the quantity $(1 + x/\rho)^{-3}$ by means of the binomial theorem and performing a Taylor series expansion of $g(x + \rho)$ about the point $r = \rho$, we obtain

$$\ddot{x} - \frac{l^2}{\mu^2\rho^3}\left[1 - 3\frac{x}{\rho} + 0\left(\frac{x}{\rho}\right)^2\right] = -\left[g(\rho) + x\left(\frac{\mathrm{d}g}{\mathrm{d}r}\right)_{r=\rho} + \dots\right]$$

which, upon using the condition for a circular path Equation 6.62, reduces to

$$\ddot{x} + \omega_0^2 x = 0 \tag{6.65}$$

where

$$\omega_0^2 = \frac{3g(\rho)}{\rho} + g'(\rho), \quad g'(\rho) = (\mathrm{d}g/\mathrm{d}r)_{r=\rho}. \tag{6.66}$$

If ω_0^2 is positive, then the equation is the same as that for the simple harmonic oscillator. In this case, the perturbed particle oscillates harmonically about the circle with radius ρ, so the circular orbit is a stable one. On the other hand, if ω_0^2 is negative, the motion after perturbation is non-oscillatory with the result that the particle recedes further and further from the circular path, so the orbit becomes unstable. If ω_0^2 is zero, higher-order terms must be included in the expansion in order to determine the stability.

In view of the preceding discussion, we can state that the condition of stability for circular orbits is

$$\omega_0^2 = \frac{3}{\rho}g(\rho) + g'(\rho) > 0$$

or because $f(r) = -\mu g(r)$

$$\frac{3}{\rho} f(\rho) + f'(\rho) < 0 \tag{6.67}$$

which is identical to Equation 6.62. In particular, if the force law is a power law

$$f(r) = -cr^n$$

where c is a positive constant, then the condition for stability reads

$$\rho^{n-1}(3 + n) > 0. \tag{6.68}$$

In order for the condition 6.68 to hold, n is restricted to values greater than –3. Hence, we conclude that, of the central forces of the form $-cr^n$, only those for which $n > -3$ will provide stable circular orbits. We notice that the inverse square law ($n = -2$) gives stable circular orbits as does the law of direct distance ($n = 1$). The latter case is that of the two-dimensional harmonic oscillator. The problem of stable circular orbits is of more than academic interest; it has many practical applications as well. As an example, it has to be considered in the design of instruments involving the motion of charged particles in electromagnetic fields, such as synchrotrons used in high-energy physics to accelerate charged particles.

Example 6.5

A particle of mass m is connected to another particle of equal mass by a light, inextensible string (of length b) that passes through a small hole in the table. Find the path of the mass on the table.

Solution:

It is more convenient to use a cylindrical coordinate system (r, θ, z) with the origin at the hole in the table. The positive z-axis is taken upward. As coordinates r and z are restricted by

$$r + (-z) = b$$

there are only two independent coordinates, and because we are only interested in the path of the particle on the table, we select r and θ as the two independent coordinates.

If we measure potential energy from the table surface, then the potential energy of the particle on the table is zero, and that of the suspended particle is

$$V = -mg(-z) = \mathrm{mg}(r - b)$$

which depends on r only. The force is, therefore, central and conservative.

The Lagrangian of the system is

$$L = \frac{1}{2} m\dot{r}^2 + \frac{1}{2} m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - mg(r - b).$$

The first term on the right-hand side is the kinetic energy of the suspended mass:

$$T = \frac{1}{2} m\dot{z}^2 = \frac{1}{2}\dot{r}^2.$$

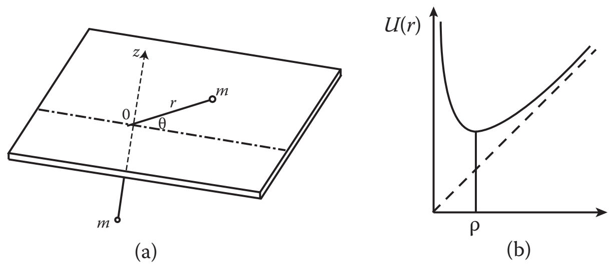


FIGURE 6.17 (a) Two equal masses connected by a light spring and (b) variation of effective potential $U(r)$ with r.

As θ is cyclic, its corresponding momentum, the angular momentum, is conserved:

$$\frac{\partial L}{\partial \dot{\theta}} = mr^2\dot{\theta} = l \text{ or } r^2\dot{\theta} = l/m = h$$

where $l = mh$ is the constant angular momentum of the particle on the table. Energy conservation gives

$$\frac{m}{2}\left(\dot{r}^2 + r^2\dot{\theta}^2\right) + \frac{1}{2}m\dot{r}^2 + mgr = E, \text{ constant}.$$

Eliminating $\dot{\theta}$ from the last two equations, the angular momentum conservation and energy conservation equations, we obtain

$$\frac{1}{2}m\dot{r}^2 + U(r) = E'$$

where $E' = E/2$ and the effective potential $U(r)$ is

$$U(r) = \frac{mh^2}{4r^2} + \frac{mg(r)r}{2} \tag{6.69}$$

and Figure 6.17b shows its variation with r. There is one stationary value, when $r = \rho$, given by

$$U'(r)\Big|_{r=\rho} = -\frac{h^2}{2\rho^3} + \frac{g}{2} = 0$$

from which it follows that

$$h^2 = g\rho^3. \tag{6.70}$$

Thus, for the particle of mass m on the table to move in a circular orbit of a given radius ρ, its angular momentum must satisfy condition 7.75. Furthermore, because

$$U''(\rho) = \frac{3h^2}{2\rho^4} = \frac{3g}{2\rho} > 0,$$

this circular path is stable; a small perturbation will make the particle oscillate harmonically about the stable circular path with a frequency $(3g/2\rho)^{1/2}$.

Example 6.6

According to Yukawa's theory of nuclear forces, the force of attraction between a proton and a neutron has the potential

$$V(r) = \frac{Ke^{-\alpha r}}{r}, \quad K < 0 \text{ and } \alpha > 0.$$

Find the force and investigate the stability of circular orbits for mass m under such a force.

Solution:

The force $F(r)$ is given by

$$F(r) = -\frac{\partial V(r)}{\partial r} = K(\alpha r + 1)\frac{e^{-\alpha r}}{r^2}$$

and it is central and conservative. The condition of stability is

$$\frac{3}{\rho} + \frac{F'(\rho)}{F(\rho)} < 0.$$

Now,

$$F'(r) = \frac{\partial F}{\partial r} = -K\left(\frac{\alpha^2}{r} + \frac{2\alpha}{r^2} + \frac{1}{r^3}\right)e^{-\alpha r}.$$

Substituting these into the stability condition leads to the inequality

$$3 + \rho\frac{-K(\alpha^2/\rho + 2\alpha/\rho^2 + 2/\rho^3)e^{-\alpha r}}{K(\alpha/\rho + 1/\rho^2)e^{-\alpha\rho}} < 0$$

which simplifies to

$$1 + \alpha\rho - \alpha^2\rho^2 < 0.$$

If we let $y = \alpha\rho$, this equation becomes

$$y^2 - y - 1 > 0.$$

Stability will result for all y that exceed the value that satisfies the equation

$$y^2 - y - 1 = 0.$$

The physically meaningful solution is

$$y = \left(-1 + \sqrt{5}\right)/2 = 1.618.$$

Thus, if the energy and the angular momentum of the particle are such that they allow a circular orbit of radius ρ, the motion will be stable provided that

$$y = \alpha\rho \geq 1.618$$

or

$$\rho \geq 1.718\alpha^{-1}.$$

6.10 APSIDES AND ADVANCE OF PERIHELION (OPTIONAL)

In Section 7.3, it was pointed out that if a particle performs bounded, noncircular motion in a central force field, the radial distance from the force center to the particle is also bounded in the range $r_{max} \geq r \geq r_{min}$. The values r_{max} and r_{min} are the turning points of the orbits; they are also called the apocenters and pericenters, respectively. The general term "apsis" (or apse, pl. apsides) is applied for both apocenters and pericenters. The radius vector to an apse is called an apsidal radius; its length is called an apsidal distance, and the angle between adjacent apsides is called an apsidal angle.

The apsidal radius divides the orbit symmetrically. That is, every apsidal radius is an axis of symmetry of the orbit. This is a consequence of the symmetry properties of the differential equation of the orbit. Reasoning as follows may also see it: As shown in Figure 6.18, point *O* is the force center, and *Q* is an apsis on the orbit. If we imagine two identical particles projected from *Q* with equal speeds but traveling in opposite directions perpendicular to the apsidal radius *OQ*, then because the force is the same at the same distance from point *O*, their subsequent path must be symmetrical. Thus, the apsidal radius *OQ* divides the orbit symmetrically.

It is also very clear that, regardless of the number of apsides in an orbit, there are, at most, two apsidal distances; furthermore, all the apsidal angles are equal. Referring to Figure 6.18 again, we see that *P*, *Q*, and *R* are consecutive apsides. As *OQ* divides the orbit symmetrically, *OP* must be equal to *OR*, and angle *POQ* must be equal to angle *QOR*. Similarly, *OR* must divide the segment of the orbit in its vicinity symmetrically, and therefore, there exists a subsequent apsis *S* (not shown) such that *OS* is equal to *OR* and for which angle *ROS* is equal to angle *QOR*. The ultimate result is that there may be, at most, two different magnitudes of apsidal distance in any orbit and one magnitude of apsidal angle. The apsidal angle for elliptical motion, for example, is π; semi-major and semi-minor axes are the two apsidal distances.

The apsidal distances may be determined from Equation 6.16 by setting $\dot{r}$ equal to zero. After we have found the apsidal distances r_1 and r_2, we are able to find the apsidal angle β between r_1 and r_2 by means of Equation 7.23:

$$\theta = \int_{r_1}^{r_2} \frac{l r^{-2} \mathrm{d}r}{[2\mu(E - \mathrm{V}) - l^2/r^2]^{1/2}}. \quad (6.71)$$

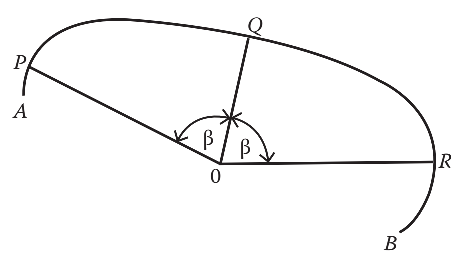


FIGURE 6.18 Apsides and the advance of perihelion.

6.10.1 Advance of Perihelion and Inverse-Square Force

The inverse-square force requires that all elliptical orbits be exactly closed, which means the apsides must stay fixed in space for all time. Even the smallest progressive motion of the apsides would indicate a deviation from the inverse-square law. Newton realized this important fact three centuries ago and stated that observation of an advance or regression of a planet's perihelion would be a very sensitive test of the inverse-square law of gravity. To see this point, let us consider planetary orbits as disturbed circular orbits. For a small disturbance, the radial motion is expected to perform simple harmonic oscillations about the radial value of the stable circular orbit. The frequency ω_0 of this oscillation is given by Equation 6.66. The time required to pass from the maximum value of the apsidal distance to the succeeding minimum value is just half the period τ_0:

$$t_a = \frac{\tau_o}{2} = \frac{1}{2}\frac{2\pi}{\omega_0} = \frac{\pi}{\omega_0}. \tag{6.72}$$

If we neglect the small variation in the orbital angular frequency ω, the corresponding apsidal angle is

$$\beta = \omega t_a = \frac{\pi\omega}{\omega_0}. \tag{6.73}$$

Now, $\omega = v/\rho$ approximately, where v is the linear velocity. And $g(\rho) = l^2/\mu^2\rho^3 = v^2/\rho$. Elimination of v from these two relations yields

$$\omega = \sqrt{g(\rho)/\rho}\,.$$

Substituting this into Equation 6.73, the apsidal angle β becomes

$$\beta = \frac{\pi}{\omega_0}\sqrt{g(\rho)/\rho} = \frac{\pi}{\sqrt{3 + \rho g'(\rho)/g(\rho)}}. \tag{6.74}$$

If we choose $g(r)$ of the form

$$g(r) = k/r^n \tag{6.75}$$

where k and n are constants, then the apsidal angle becomes especially simple:

$$\beta = \pi(3 - n)^{-1/2}. \tag{6.76}$$

The dependence on the size of the orbit disappears. It is clear that the orbit is re-entrant, or repetitive, in the case of the inverse-square force ($n = 2$) for which $\beta = \pi$ and also in the case of the linear law ($n = -1$) for which $\beta = \pi/2$. If, however, say $n = -2$, then $\beta = \pi/\sqrt{5}$, which is an irrational multiple of π, and so the motion does not repeat itself. In the case of an inverse-square force, if n is just slightly different from 2, the orbit will be no longer re-entrant and the apsides will either advance or regress steadily in the plane of the orbit, depending on whether the apsidal angle is slightly greater or slightly less than π.

The planetary orbits demonstrate such progressive motions, although these movements are very slight. Part of them can be explained by the presence of perturbations resulting from other bodies in the solar system. However, some of the motions cannot be accounted for in this manner. For example, the perihelion of the planet Mercury, which shows the largest effect, advances approximately

574 seconds of arc per century; approximately 531 seconds of arc can be explained by the presence of perturbations resulting from other bodies in the solar system. The rest (42.57 seconds of arc) can be accounted for almost exactly by the modification that is introduced into the equation of motion of a planet by the general theory of relativity. This small modification term behaves effectively as a small additional inverse fourth-power component to the Newtonian gravitational force. We shall indicate below qualitatively how to use the modified equation of motion to calculate the advance of the perihelion.

As a first approximation, we still view planetary orbits as disturbed circular orbits. Now, with the small modification term, the force $g(r)$ is of the form

$$g(r) = \frac{k}{r^2} + \frac{b}{r^4} = \frac{k}{r^2}\left(1 + \frac{b}{kr^2}\right) \tag{6.77}$$

where k and b are constants and $b/kr^2 \ll 1$. Then,

$$g(\rho) = \frac{k}{\rho^2}\left(1 + \frac{b}{k\rho^2}\right)$$

and

$$\rho g'(\rho) = -\frac{2k}{\rho^2}\left(1 + \frac{2b}{k\rho^2}\right)$$

from which we obtain

$$\frac{\rho g'(\rho)}{g(\rho)} = -2\left(1 + \frac{2b}{k\rho^2}\right)\left(1 + \frac{b}{k\rho^2}\right)^{-1} = -2\left(1 + \frac{2b}{k\rho^2}\right)\left(1 - \frac{b}{k\rho^2} + \ldots\right) \cong -2\left(1 + \frac{b}{k\rho^2}\right)$$

where we have neglected powers of b/k higher than the first. Thus, from Equation 6.74, the apsidal angle β is

$$\beta = \frac{\pi}{\sqrt{3 - 2(1 + b/k\rho^2)}} \cong \pi\left(1 + \frac{b}{k\rho^2}\right). \tag{6.78}$$

We see that the orbit will not be re-entrant and that the apsides advance if b is positive, whereas they regress if b is negative. The advance or regress of apsides in each revolution varies inversely as the square of the radius of the orbit, so the effect decreases rapidly the farther a planet is from the sun. Thus, Mercury is the only planet to demonstrate a large effect.

6.10.2 Method of Perturbation Expansion

When the eccentricity of the orbit is large, the approximation of the disturbed circular orbit becomes a poor one. In this case, the precession of perihelion can be calculated directly from Equation 6.71:

$$\beta = 2\int_{r_1}^{r_2} \frac{lr^{-2}\mathrm{d}r}{\sqrt{2\mu(E - V) - l^2 r^{-2}}}$$

where r_1 and r_2 are the apsidal distances. In order to avoid spurious divergences at the limits of integration r_1 and r_2, we write the preceding expression for β as

$$\beta = -2\frac{\partial}{\partial r}\int_{r_1}^{r_2}\sqrt{2\mu(E-V)-l^2r^{-2}}\,\mathrm{d}r \tag{6.79}$$

where $V = V_0 + \delta V = -k/r + \delta V$, and δV is the small correction term to the unperturbed potential V_0 for which the exact solution of the motion is known. Now, expanding the integrand in powers of δV, we find the changes of the apsidal angle in various orders:

$$\beta_0 = \int_{r_1}^{r_2}\frac{2lr^{-2}\mathrm{d}r}{\sqrt{2\mu(E-V_0)-l^2r^{-2}}}$$

$$\beta_1 = \frac{\partial}{\partial l}\int_{r_1}^{r_1}\frac{2\mu\delta V\mathrm{d}r}{\sqrt{2\mu(E+k/r)-l^2r^2}} = \frac{\partial}{\partial l}\int_0^{\pi}2\mu l^{-1}r^2\,\delta\mathrm{d}\theta$$

$$\beta_2 = \frac{\partial}{\partial l}\frac{1}{l}\frac{\partial}{\partial l}\left[\frac{\mu^2}{l}\int_0^{\pi}r^4(\delta V)^2\,\mathrm{d}\theta\right].$$

In general,

$$\beta_k = \frac{2}{2^k k!}\left(\frac{\partial}{\partial l}\frac{1}{l}\right)^k\int_0^{\pi}r^{2k}(2\mu\,\delta V)^k\,\mathrm{d}\theta,\quad k = 1,2,3,\ldots \tag{6.80}$$

where we have changed from integration over r to one over θ along the path of the unperturbed motion.

For $\delta V = b/r^3$, we have $r^2\delta V = b/r$ and $r = \alpha/(1 + \varepsilon\cos\theta)$; the zero-order term in the expansion gives 2π, and the first-order term gives the required changes:

$$\beta_1 = -6\pi\alpha kb\mu^2 l^{-4}.$$

Now, $\alpha = l^2\mu k$, $l^2 = \mu ka(1 - \varepsilon^2)$, and $k = GmM$; with these substitutions, β_1 becomes

$$\beta_1 = -\frac{6\pi b}{Gm^2Ma^2(1-\varepsilon^2)} \tag{6.81}$$

which is a familiar expression in terms of a (semi-major axis) and ε (eccentricity). We can further simplify this if we know the exact value of b. It is obvious from Equation 6.81 that the effect is enhanced if the semi-major axis a is small and if the eccentricity ε is large. Mercury, which is the planet closest to the sun and which has the most eccentric orbit, therefore, shows the largest effect. The calculated value of the precessional rate for Mercury by relativity is 43.03 + 0.03 second of arc per century. The observed value (corrected for the influence of the other planets) is 43.11 + 0.45 second of arc. This striking agreement is one of the major triumphs of the theory of relativity.

The gravitational field near the Earth departs slightly from the inverse-square law. This is because the Earth is not a perfect sphere. Consequently, the perigee of a satellite whose orbit lies near the Earth's equatorial plane will advance steadily in the direction of the satellite's motion as the satellite moves around the Earth. We can use the observation of this advance to determine the shape of the

Earth. Such observations have shown that the Earth is slightly pear-shaped. The oblateness of the Earth also causes the orbital plane of a satellite to process if it is not in the Earth equatorial plane.

6.11 LAPLACE–RUNGE–LENZ VECTOR AND THE KEPLER ORBIT (OPTIONAL)

The equation of the orbit for the Kepler problem can be obtained by a simple algebraic method. This method requires only vector analysis techniques and does not involve solving a differential equation or performing any integration. Apart from its simplicity, the method is intrinsically interesting because it makes use of an unusual conserved quantity, the Laplace–Rung–Lenz (LRL) vector. In the corresponding quantum-mechanical case, the hydrogen atom, eigenvalues, and eigenfunctions are readily found from knowledge of the LRL vector. Knowledge of the LRL vector also allows one to study the symmetry characteristics in a simple way. Thus, the reader may encounter generalizations of the ideas considered here in more advanced courses.

We first show that the LRL vector is a constant of the motion and then use this fact to obtain the equation of the orbit. Now the equation of motion of a particle of mass m in the potential energy $V(r) = -k/r$ is

$$m\ddot{\vec{r}} = -\frac{k\hat{r}}{r^2}.$$

Taking the cross product of both sides with $\vec{l} = m\vec{r} \times \dot{\vec{r}}$, we obtain

$$m\ddot{\vec{r}} \times \vec{l} = -\frac{k\hat{r}}{r^2} \times \vec{l} = -mk\hat{r} \times \frac{\vec{r} \times \dot{\vec{r}}}{r^2}.$$

Using the vector identity that holds for any three vectors $\boldsymbol{a}$, $\boldsymbol{b}$, $\boldsymbol{c}$

$$\vec{a} \times (\vec{b} \times \vec{c}) = \vec{b}(\vec{a} \cdot \vec{c}) - \vec{c}(\vec{a} \cdot \vec{b})$$

in the preceding equation, we obtain

$$\ddot{\vec{r}} \times \vec{l} = -\frac{k}{r^2}\left[-\dot{\vec{r}}(\hat{r} \cdot \vec{r}) + \vec{r}(\hat{r} \cdot \dot{\vec{r}})\right]. \tag{6.82}$$

The right-hand side reduces to $-k\dot{\hat{r}}$. To prove this, we start with

$$\vec{r} = r\hat{r}$$

from which it follows that

$$\dot{\vec{r}} = \dot{r}\hat{r} + r\dot{\hat{r}}$$

and

$$\dot{\vec{r}}(\hat{r} \cdot \vec{r}) = \dot{\vec{r}}r = r(\dot{r}\hat{r} + r\dot{\hat{r}}) = \dot{r}\vec{r} + r^2\dot{\hat{r}}$$

$$\vec{r}(\hat{r} \cdot \dot{\vec{r}}) = \vec{r}\left[\hat{r} \cdot (\dot{r}\hat{r} + r\dot{\hat{r}})\right] = \vec{r}(\dot{r} + r\hat{r} \cdot \dot{\hat{r}}) = \vec{r}\dot{r}$$

where we have used the fact that $\hat{r} \cdot \dot{\hat{r}} = 0$. Thus,

$$-\frac{k}{r^2}\left[-\dot{\vec{r}}\left(\hat{r}\cdot\vec{r}\right)+\vec{r}\left(\hat{r}\cdot\dot{\vec{r}}\right)\right]=-\frac{k\left(\dot{r}\vec{r}-r^2\dot{\hat{r}}-\dot{r}\vec{r}\right)}{r^2}=k\dot{\hat{r}}.$$

Equation 6.82 now reduces to

$$\ddot{\vec{r}}\times\vec{l} = k\dot{\hat{r}}. \tag{6.83}$$

Because $\boldsymbol{l}$ is a constant in time, integration of Equation 6.83 with respect to time yields

$$\dot{\vec{r}}\times\vec{l} = k\hat{r}+\vec{a}$$

or

$$\dot{\vec{r}}\times\vec{l} - \frac{k\vec{r}}{r} = \vec{a}$$

where $\vec{a}$ is a vector independent of time and is in the plane of motion. The vector quantity

$$\begin{aligned}\vec{A} &= \vec{a}/k \\ &= \dot{\vec{r}}\times\frac{\vec{l}}{k}-\frac{\vec{r}}{r} = -\frac{\vec{l}\times\vec{p}}{mk}-\hat{r}\end{aligned} \tag{6.84}$$

is, therefore, a constant of the motion, and its magnitude is given by

$$\begin{aligned}A^2 = \vec{A}\cdot\vec{A} &= \left(-\frac{\vec{l}\times\vec{p}}{mk}-\hat{r}\right)\cdot\left(-\frac{\vec{l}\times\vec{p}}{mk}-\hat{r}\right) \\ &= \frac{2l^2}{m^2k^2}\left(\frac{p^2}{2m}-\frac{k}{r}\right)+1\end{aligned} \tag{6.85}$$

where we have used the vector identity

$$\left(\vec{a}\times\vec{b}\right)\cdot\left(\vec{c}\times\vec{d}\right)=\left(\vec{a}\cdot\vec{c}\right)\left(\vec{b}\cdot\vec{d}\right)=\left(\vec{a}\cdot\vec{d}\right)\left(\vec{b}\cdot\vec{c}\right).$$

The quantity in the parentheses in Equation 6.89 is the energy E of the particle; thus, the magnitude of $\vec{A}$ can be expressed in terms of $\vec{l}$ and E:

$$A=\left(1+\frac{2El^2}{mk^2}\right)^{1/2}$$

which is just the eccentricity ε of the orbit (see Equation 6.38a). With this association, the vector $\vec{A}$ is sometimes called the eccentricity vector. It first appeared in Laplace's *Traite de mecanique celeste* in 1799 and was reintroduced by Runge (1919) and Lenz (1924), so $\vec{A}$ is now known as the LRL vector.

To find the equation of the orbit, we take the dot product of $\vec{A}$ with $\vec{r}$:

$$\vec{A}\cdot\vec{r} = \vec{r}\cdot\vec{p}\times\frac{\vec{l}}{mk} - r.$$

By interchange of dot and cross products according to the following rule

$$\vec{a}\cdot\left(\vec{b}\times\vec{c}\right) = \left(\vec{a}\times\vec{b}\right)\cdot\vec{c}$$

and use of $\vec{l} = \vec{r}\times\vec{p}$, we obtain

$$\vec{A}\cdot\vec{r} = \frac{l^2}{mk} - r$$

or

$$r = -\vec{A}\cdot\vec{r} + \frac{l^2}{mk}\,. \tag{6.86}$$

If we define the direction of $\boldsymbol{r}$ relative to $\vec{A}$ by an angle θ, then

$$\vec{A}\cdot\vec{r} = Ar\cos\theta.$$

Substituting this into Equation 9.90 leads to

$$r = \frac{l^2/mk}{1 + A\cos\theta} = \frac{l^2/mk}{1 + \varepsilon\cos\theta} \tag{6.87}$$

which is just the result found previously in Equation 6.39 with $\theta_0 = 0$. The orientation of the orbit is specified relative to the LRL vector by Equation 6.86, and the LRL vector lies along the symmetry axis of the conic section as illustrated in Figure 6.19.

We can also obtain the frequency of a small oscillation about a stable circular path that results when the particle is given a small radial impulse that changes its energy but not its angular momentum. Now, in a cylindrical coordinate system (with the z-axis perpendicular to the orbit plane), we have

$$\vec{l} = \mu r^2\dot{\theta}\,\hat{e}_z$$

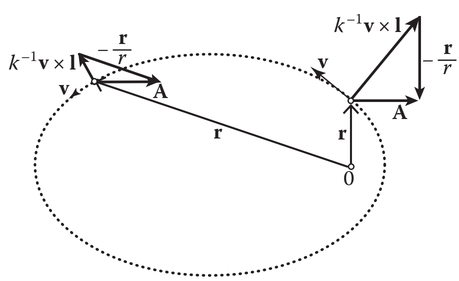


FIGURE 6.19 LRL vector.

and

$$\vec{A} = k^{-1}\left(\mu r^3\dot{\theta}^2 - k\right)\hat{e}_r - k^{-1}\mu r^2 \dot{r}\dot{\theta}\hat{e}_\theta.$$

From these two equations, we have

$$A^2 = \vec{A}\cdot\vec{A} = k^{-2}\left(\frac{l^2}{\mu r} - k\right)^2 + k^{-2}l^2\dot{r}^2. \tag{6.88}$$

For a circular orbit, A and $\dot{r}$ must vanish at $r = r_0$, and then Equation 6.88 reduces to

$$\frac{l^2}{\mu r_0} = k. \tag{6.89}$$

We now perturb the orbit, letting $r \to r_0 + x$ and $\dot{r} = \dot{x}$ in Equation 6.88:

$$k^2A^2 = \left(\frac{l^2}{\mu(r_0 + x)} - k\right)^2 + l^2\dot{x}^2. \tag{6.90}$$

Assuming that $x/r_0 \ll 1$ and using Equation 6.89 twice, we approximate Equation 6.94 as

$$k^2A^2 \approx k^2(x/r_0)^2 + k\mu\ r_0\dot{x}^2\ . \tag{6.91}$$

Because the particle is still moving in the same force field, we have $\dot{\vec{A}} = 0$; thus, from Equation 6.91, we have

$$2k^2A\dot{A} \approx 2k\mu r_0\dot{x}\left(\ddot{x} + kx/\mu r_0^3\right) = 0$$

or

$$\ddot{x} + \frac{k}{\mu r_0^3}x = 0 \tag{6.92}$$

which gives

$$\omega^2 = \frac{k}{\mu r_0^3}. \tag{6.93}$$

Another application of the LRL vector is that the precession of the perihelia can also be obtained from computing the rate of change of the LRL vector. A discussion of this interesting application would be very lengthy and goes beyond our syllabus. We refer the interested reader to the works of Davies (1982), Sivardiere (1984), and Garavaglia (1987).

PROBLEMS

1. Consider a particle moving in a field whose potential is $x^{-2} - x^{-4}$. Show by a graphical method that for small energies the motion is oscillatory and for large energies the motion is non-periodic and extends to infinity. Find the energy that forms the dividing line between

these two curves. Also compute the limiting frequency as the amplitude of the oscillatory motion gets smaller and smaller.

2. Determine the motion of a two-dimensional linear oscillator of potential energy

$$V = \frac{1}{2}kr^2.$$

3. Show that, in terms of Cartesian coordinates

$$x = r\cos\theta \; y = r\sin\theta,$$

   (a) the equation

$$\frac{(x+ae)^2}{a^2} + \frac{y^2}{b^2} = 1$$

   is the Cartesian equation for an ellipse with center C at $x = -ae$ and $y = 0$, semi-major axis a, semi-minor axis b, and eccentricity e.

   (b) The equation

$$\frac{(x-ae)^2}{a^2} - \frac{y^2}{b^2} = 1$$

   represents a hyperbola with its center C at $x = ae$, $y = 0$, and asymptotes $y = \pm b(x - ae)/a$.

4. A particle of mass m under the action of a central force describes an orbit that is a circle of radius a passing through the center of force. Find the law of force.
5. A particle in a central field moves in the orbit

$$r = r_0\theta^2.$$

   Determine (a) the force function and the potential energy function; and (b) how the angle θ varies with time.

6. A comet is seen at a distance r_0 from the sun. It is moving with a speed v_0, and its direction of motion makes an angle φ with the radius vector from the sun. Determine the eccentricity of the comet's orbit.
7. A particle is constrained to move on the inner surface of a smooth right circular cone of half-angle α. The axis of the cone is vertical, so gravity is acting vertically downward (Figure 6.20). (a) Find the Lagrangian function for the motion in cylindrical coordinates (r, θ, z). (b) Show that at any instant, there is a central force directed toward a point on the z-axis on the same horizontal line with the particle. (c) Suppose the particle is now in steady motion in a horizontal circle at a vertical height h above the vertex. Show that this motion is dynamically stable and find the angular frequency of small oscillations about this motion.
8. Show that when a satellite is in elliptical orbit, its speed V is given by

$$v^2 = k(2/r - 1/a)$$

   and when in parabola orbit

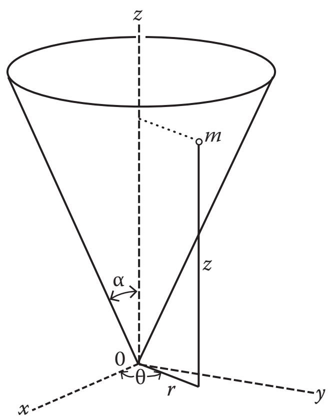


FIGURE 6.20 Particle moving on the inner surface of a smooth right circular cone.

$$v^2 = 2k/r.$$

9. Obtain the conditions of the k's for stable circular orbits and the period of small oscillations about the stable circular orbit for the following central force fields:

$$(a)\ f(r) = kr^2/2;\ (b)\ f(r) = -a/r^k,\ a > 0;\ (c)\ f(r) = (k_1/r^2) + (k_2/r^4).$$

10. Show that particles moving in a central force field move in closed non-processing orbits only if the force law is of the inverse square (Coulomb) or linear (harmonic oscillator) type.
11. A satellite is moving in a circular orbit of $2r_e$ around the Earth, where r_e is the Earth's radius. At an instant of time, the direction of motion of the satellite is changed through an angle α toward the Earth without a change in speed. Find the angle α in order that the satellite just touches the Earth. (The condition for the satellite to just touch the Earth is that the apsidal distance $r_1 = r_e$.)
12. (a) A satellite of mass m, initially in a circular orbit of radius r_1 around the Earth, is maneuvered to change to another circular orbit of radius $r_2(>r_1)$ as shown in Figure 6.21. In order to accomplish this, the rocket motor should be first fired at point A on the orbit nearest to the origin (the force center) in such a way that the satellite goes into an elliptical orbit. When the satellite reaches point B, which is furthest from the origin, the motor is fired again to bring the velocity to the necessary value for circular motion with a radius of r_2. What is the advantage of firing the motor at A and B? Determine the velocity v at A after the motor is fired and the velocity at B before the motor is fired. What is the gain in energy at A (it is the increase in speed for the required elliptical orbit)? And what is the gain in energy at B?

    (b) Two satellites, S_1 and S_2, are describing circular orbits in the same plane of radii r_1 and r_2 around the Earth. We want to effect a rendezvous by boosting the velocity of S_1. The most effective way to do this is to increase the velocity of S_1 at point A so as to put S_1 into an elliptical orbit whose apsidal distances are r_1 and r_2. Find the time taken for S_1 to arrive at the orbit of S_2 and the angular distance traveled by S_2 during this time interval.

    In order to effect a rendezvous, S_1 should be shot into orbit when S_2 is in advance of S_1 by a certain proper angular distance. Find this angular distance.

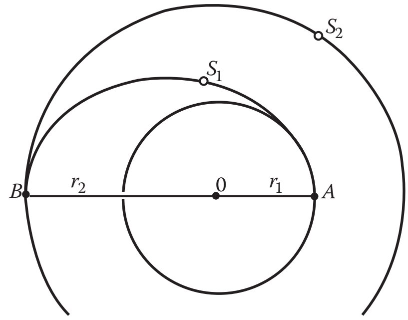


FIGURE 6.21 Satellite orbit in space.

13. Derive an expression for the differential scattering cross section of a mass m subject to the force

$$\vec{F} = \frac{k}{r^3}\hat{e}_r .$$

14. Show that
    (a) the energy of an elliptical orbit can be written as

$$E = \frac{GmM}{r_{\max} + r_{\min}} .$$

    (b) The period of an elliptic orbit can be written as

$$T = \frac{2\pi GM}{(-2E/m)^{3/2}} .$$

15. The interaction between two atoms in a diatomic molecule is described approximately by the potential

$$V(x) = -ax^{-6} + bx^{-12},\ a > 0,\ b > 0$$

where x is the distance between the atoms. Find the force between the two atoms. If one of the two atoms is heavy and can be considered at rest, what is the possible motion for the light atom? Describe the possible motion qualitatively first, and then find the equilibrium distance and the period of small oscillations about the equilibrium distance.

REFERENCES

Davies, B. Derivation of perihelion precession, *Am. J. Phys.*, 51, 909, 1982.
Garavaglia, T. The Runge-Lunz vector and Einstein perihelion precession, *Am. J. Phys.*, 55, 164, 1987.
Sivardiere, J. Precession of elliptic orbits, *Am. J Phys.*, 52, 909, 1984.

7 Harmonic Oscillator

Harmonic oscillators, along with damped and driven oscillators, will be treated in considerable detail in this chapter not merely because harmonic motion is a good approximation of many physical processes but also because a thorough understanding of this process aids comprehension of the other types of oscillations. By Fourier analysis, complicated oscillations often may be regarded as consisting of a number of simple harmonic oscillations.

Simple harmonic motion (SHM) arises whenever a system vibrates around an equilibrium position. It is caused by a force that is directed toward the equilibrium position and that is proportional to the displacement of the particle from the equilibrium position, which causes its motion. Examples of simple harmonic motions are found in the motion of a weight on the end of a perfect elastic spring, the bob of a simple pendulum swinging through a very small arc, atoms in a crystal lattice, the nuclei of atoms in molecules, and so on.

7.1 SIMPLE HARMONIC OSCILLATOR

We first consider two examples of simple harmonic motion.

7.1.1 Motion of Mass *m* on the End of a Spring

The spring has a natural length b and spring constant k when the system is in equilibrium (i.e., when the particle hangs motionless); the weight of the particle is exactly balanced by the restoring force of the spring where d is the extension of the spring (Figure 7.1):

$$mg = kd. \tag{7.1}$$

Suppose the particle has been set into vertical vibration at the instant when the displacement of the particle is x; the extension of the spring is then $d + x$. The equation of motion of the particle is then given by, neglecting friction,

$$F = -k(d + x) + mg = m\ddot{x}$$

which, by using Equation 7.1, reduces to

$$m\ddot{x} = -kx$$

or

$$\ddot{x} + \omega_0^2 x = 0 \tag{7.2}$$

where $\omega_0 = (k/m)^{1/2}$, and $-kx$ is the restoring force of the spring, which is proportional to the displacement and directed toward the equilibrium. The potential function from which the force $F = -kx$ is derivable is

$$V(x) = -kx^2/2. \tag{7.3}$$

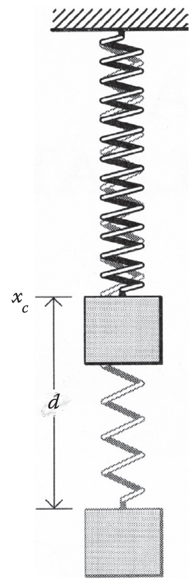


FIGURE 7.1 Oscillating masses.

7.1.2 The Bob of Simple Pendulum Swinging through a Small Arc

A simple pendulum consists of a small bob of mass m suspended by a light inextensible string of length b, free to swing in a vertical plane about the equilibrium position 0. A simple coordinate, θ or x, is required to describe completely the position of the bob at any time (Figure 7.2). Let us set up the equation of motion of the bob using the Lagrangian approach; the kinetic energy and the potential energy of the bob are, respectively,

$$T = \frac{m}{2}(\dot{x}^2 + \dot{y}^2) = \frac{m}{2}b^2\dot{\theta}^2, \; V = mgb(1 - \cos\theta) \tag{7.4}$$

where we have taken $V = 0$ at the point 0. The Lagrange L of the particle is

$$L = T - V = \frac{m}{2}b^2\dot{\theta}^2 - mgb(1 - \cos\theta). \tag{7.5}$$

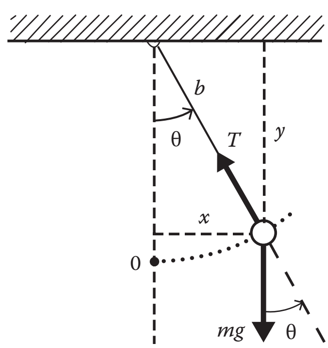


FIGURE 7.2 Simple pendulum.

The Lagrange's equation for coordinate θ gives

$$\ddot{\theta} + \omega_0^2 \sin\theta = 0$$

which reduces to, for small θ,

$$\ddot{\theta} + \omega_0^2 \theta = 0. \tag{7.6}$$

Again $\omega_0^2 = g/b$. It is evident that Equations 7.2 and 7.6 are mathematically identical in form.

The motion of a particle in a potential valley can often be approximately treated in terms of simple harmonic motion if the displacements of the particle from the minimum potential energy at $x = x_e$ are small. To see why, let us approximate the potential function $V(x)$ by a few terms of a series expansion about $x = x_e$ (Figure 7.3):

$$\begin{aligned} V(x) &= V(x_e) + \left.\frac{dV(x)}{dx}\right|_{x=x_e} (x - x_e) + \frac{1}{2}k(x - x_e)^2 + \dots \\ &= V(x_e) + \frac{1}{2}k(x - x_e)^2 + \dots \end{aligned} \tag{7.7}$$

where

$$\left.\frac{dV}{dx}\right|_{x=x_e} = 0, \quad k = \left.\frac{d^2V}{dx^2}\right|_{x=x_e} > 0.$$

The constant term $V(x_e)$ has no consequences for the physical motion, so we drop it. If we let $x' = x - x_e$, then $V(x') = kx'^2/2$, which shows that small oscillations in complicated systems can be approximately treated in terms of simple harmonic motion. As an example, we can easily find the linear harmonic oscillator approximation for the potential energy function $V(x) = ax^{-2} + bx^2$, where a and b are constants. We first find the equilibrium position x_e by setting $dV/dx = 0 = -2ax^{-3} + 2bx$, which gives $x_e = (a/b)^{1/4}$. Then we calculate $k = d^2V/dx^2\big|_{x=x_e} = 6ax_e^{-4} + 2b = 8b$. Therefore, $V_{LHO} = 2\,kx^2 = 4bx^2$.

7.1.3 Solution of Equation of Motion of SHM

We now proceed to solve Equation 7.2, which is a homogeneous, linear differential equation of the second order with constant coefficients. To solve it, we multiply both its sides by the integrating factor $2\dot{x}$:

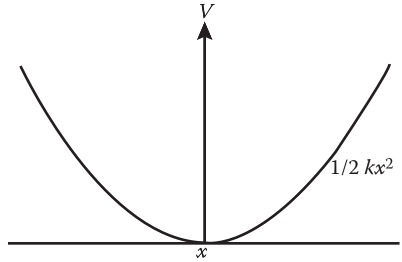


FIGURE 7.3 Parabolic potential function of a harmonic oscillator.

$$2\dot{x}\ddot{x} = -2\omega_0^2 x\dot{x}$$

or

$$d\dot{x}^2 = -\omega_0^2 dx^2.$$

Integration yields

$$\dot{x}^2 - \dot{x}_0^2 = -\omega_0^2\left(x^2 - x_0^2\right)$$

where x_0 and $\dot{x}_0$ are the initial position and speed, respectively. The last equation may be rewritten as

$$\dot{x}^2 = \omega_0^2(a^2 - x^2) \tag{7.8}$$

with

$$a^2 = \dot{x}_0^2 / \omega_0^2 + x_0^2.$$

From Equation 7.8, we obtain

$$\int_{x_0}^{x} \frac{dx}{\sqrt{a^2 - x^2}} = \int_0^t \omega_0 \, dt. \tag{7.9}$$

If we let $x = a \sin\theta$, then $dx = a \cos\theta d\theta$, and Equation 7.9 becomes

$$\int_{x_0}^{x} \frac{dx}{\sqrt{a^2 - x^2}} = \int_{\theta_0}^{\theta} \frac{a\cos\theta d\theta}{\sqrt{a^2(1-\sin^2\theta)}} = \int_{\theta_0}^{\theta} d\theta.$$

Combining this equation with Equation 7.9, we obtain

$$\omega t = \theta - \theta_0 = \sin^{-1}(x/a) - \theta_0 \quad \text{or} \quad x = a \sin(\omega_0 t + \theta_0) \tag{7.10}$$

where

$$\theta_0 = \sin^{-1}(x/a). \tag{7.11}$$

Equation 7.10 is the general solution of Equation 7.2 and contains two integration constants: a and θ_0. The motion is a sinusoidal oscillation of the displacement x. The quantity ω_0 is called the angular frequency (in rad/s) of the oscillations and is a fundamental characteristic of the oscillations. It is entirely determined by the properties of the mechanical systems [$\omega_0 = (k/m)^{1/2}$ or $(b/g)^{1/2}$]. But this property of the angular frequency ω_0 depends on the assumption that the oscillations are small and ceases to hold if the oscillations are not small. The constant a is called the amplitude of the motion and is the maximum value of the displacement x. The angle $\omega_0 t + \theta_0$ is called the phase at time t, and θ_0 is the initial phase (or phase constant).

The period τ of the motion is the time required to execute a complete oscillation. Thus, it is the time in which the argument of the sine increases by 2π:

$$\omega_0 t + \theta_0 + 2\pi = \omega_0(t + \tau) + \theta$$

or

$$\tau = 2\pi/\omega_0. \tag{7.12}$$

The period τ, like ω_0, is thus independent of the initial conditions. An oscillating system that exhibits this property is said to be isochronous. The frequency of oscillation ν is the number of oscillations per unit of time:

$$\nu = 1/\tau = \omega_0/2\pi. \tag{7.13}$$

It is often very convenient to use complex numbers in discussing periodic phenomena. We now expand Equation 7.10 as

$$x = a\cos\theta_0 \sin(\omega_0 t) + a\sin\theta_0 \cos(\omega_0 t)$$

which may be put into exponential form by using Euler's formula $e^{i\alpha} = \cos\alpha + i\sin\alpha$:

$$x = \mathrm{Re}[Ae^{i\omega_0 t}] \tag{7.14}$$

with $A = -ia\exp(i\theta_0)$, where the term Re means that we take the real part of the quantity inside the square brackets.

7.1.4 Kinetic, Potential, Total, and Average Energies of Harmonic Oscillator

The kinetic energy T is given by

$$T = \frac{1}{2}m\dot{x}^2 = \frac{1}{2}m\omega_0^2(a^2 - x^2). \tag{7.15}$$

The potential energy V is the work done in displacing the oscillator from $x = 0$ to x:

$$V = \int_0^x (-F)\,\mathrm{d}x = \int_0^x kx\,\mathrm{d}x = \frac{1}{2}kx^2. \tag{7.16}$$

Accordingly, the total energy E is

$$E = T + V = \frac{1}{2}m\omega_0^2(a^2 - x^2) + \frac{1}{2}kx^2 = \frac{1}{2}ka^2 \tag{7.17}$$

which shows that the total energy of a harmonic oscillator is constant and is proportional to the square of its amplitude. Note that this is true only when no damping mechanisms are present.

At $x = 0$, the potential energy is zero and the kinetic energy is maximum. At $x = a$, the potential energy is maximum and the kinetic energy is minimum. Thus, we see that the kinetic energy is maximum when the potential energy is minimum and vice versa. The total energy of the simple harmonic oscillator remains constant. The variations of kinetic energy, potential energy, and the total energy of a simple harmonic oscillator as a function of displacement from the mean position are shown in Figure 7.4.

We now prove an important characteristic of a harmonic oscillator related to the time average of the kinetic and potential energies. The time average of a quantity K over a time interval τ is

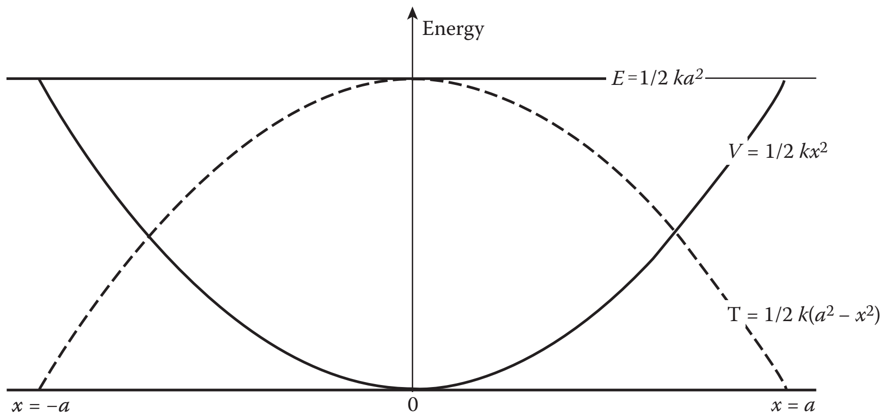


FIGURE 7.4 Kinetic, potential, and total energy variations of harmonic oscillator with displacement.

$$\langle K \rangle = \frac{1}{\tau}\int_0^{\tau} K(t)\,\mathrm{d}t.$$

Because the motion of an oscillator is repetitive, the time average over one period is the same as over many periods and is unique.

Now, for an oscillator whose motion obeys Equation 7.10, $x = a\sin(\omega_0 t + \theta_0)$, and the period $\tau = 2\pi/\omega_0$:

$$\langle T\rangle = \frac{1}{\tau}\int_0^{\tau}\frac{1}{2}m\dot{x}^2 = \frac{1}{2}m\omega_0^2 a^2\,\frac{\int_0^{2\pi/\omega_0}\cos^2(\omega_0 t+\theta_0)\,\mathrm{d}t}{2\pi/\omega_0}.$$

Because the integral is extended over a complete period, it does not matter what the value of the initial phase θ_0 is, and we may conveniently set $\theta_0 = 0$. Using the fact that

$$\frac{\omega_0}{2\pi}\int_0^{2\pi/\omega_0}\cos^2\omega_0 t\,\mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi}\cos^2 y = \frac{1}{2}$$

we find

$$\langle T\rangle = \frac{1}{\tau}\int_0^{\tau}\frac{1}{2}m\dot{x}^2 = \frac{1}{4}m\omega_0^2 a^2.$$

The potential energy is (again, $\theta_0 = 0$)

$$V = \frac{1}{2}kx^2 = \frac{1}{2}ka^2\sin^2\omega_0 t$$

and we find

$$\langle V\rangle = \frac{1}{4}m\omega_0^2 a^2.$$

Thus, $\langle V\rangle = \langle T\rangle$, and the total energy of the harmonic oscillator is

$$\langle E\rangle = \langle T\rangle + \langle V\rangle = \frac{1}{2}m\omega_0^2 a^2 = \frac{1}{2}ka^2.$$

Note that $\langle E\rangle = E$ because the total energy is a constant of the motion.

The equality of the time average of kinetic energy and potential energy is a special property of the harmonic oscillator. It is not true for anharmonic oscillators but is a good approximation for weekly damped oscillators as will be shown in Section 7.3.

Example 7.1

A cylindrical object of density d_c floats in a liquid of density d_l as shown in Figure 7.5. Determine the differential equation of motion and the natural frequency of oscillation when the cylinder is depressed by an external force and then released.

Solution:

First we select a coordinate system in which the undisturbed position of the cylinder corresponds to $y = 0$, which is shown on the left side of Figure 7.5. When the object is depressed, a buoyant force is established equal to the weight of the liquid displaced by the additional submersion beyond that required to maintain static equilibrium. For any y, the unbalanced buoyant force is

$$F_b = \pi R^2\, y d_\rho g \tag{7.18}$$

where the minus sign indicates that F_b acts to return the object to equilibrium. By using Newton's second law, we obtain

$$-\pi R^2 y d_\rho g = d_c \pi R^2 L\ddot{y}$$

or

$$\ddot{y} + \frac{g d_\rho}{L d_c} y = 0.$$

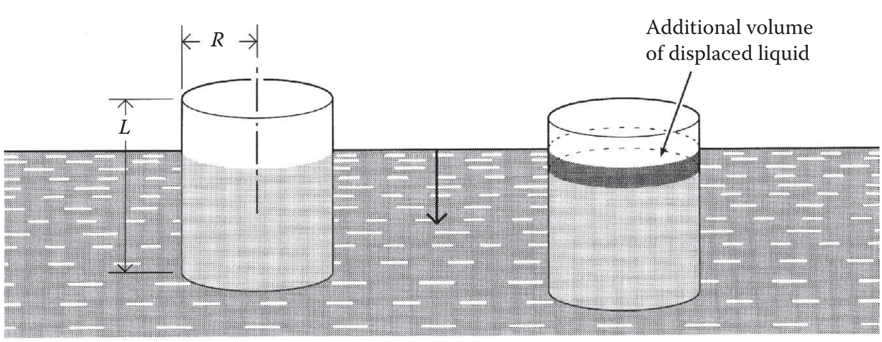


FIGURE 7.5 Cylindrical object floats in a liquid.

This differential equation should be easily recognized as the one governing the simple harmonic motion, and the natural frequency is

$$\omega_0 = \sqrt{g d_p / L d_c}\,.$$

7.2 ADIABATIC INVARIANTS AND QUANTUM CONDITION

An adiabatic invariant is a quantity that remains constant if the parameters that describe the mechanical system are changed extremely slowly (adiabatically, to use a thermodynamic term), even though they may ultimately be greatly changed. As early as 1902, Rayleigh studied a simple pendulum undergoing small oscillations. He found that, if the suspending string l is continuously and extremely slowly shortened, the ratio of energy E of the oscillatory motion to the frequency ν remains essentially constant, where E is measured during a period of time, during which the length of the string does not change appreciably. We now proceed to show this in detail. When we shorten the suspending string, we do work dW against gravity and the centrifugal force (Figure 7.6):

$$\mathrm{d}W = -\int_{l}^{l+\Delta l} (mg\cos\theta + ml\dot{\theta}^2)\,\mathrm{d}l. \tag{7.19}$$

If the shortening of the string takes place extremely slowly, so that the change in amplitude from one oscillation to the next is negligible, we can integrate over the motion on the supposition that the amplitude is constant. We then obtain

$$\mathrm{d}W = -\left(mg\langle\cos\theta\rangle + ml\langle\dot{\theta}^2\rangle\right)\Delta l \tag{7.20}$$

where $\langle A\rangle$ is the average value of the quantity A taken over the undisturbed motion. For smaller amplitudes, $\cos\theta \cong 1 - \theta^2/2$, and Equation 7.20 becomes

$$\mathrm{d}W \cong -mg\Delta l + m\left(\frac{1}{2}g\langle\theta^2\rangle - l\langle\dot{\theta}^2\rangle\right)\Delta l \cong -mg\Delta l + \Delta E. \tag{7.21}$$

The first term is the work done in raising the pendulum's position of equilibrium by Δl, which does not interest us, and ΔE represents the energy imparted to the oscillation. For the undisturbed simple harmonic motion, the mean kinetic energy and mean potential energy are equal and are equal to half of the total energy, that is,

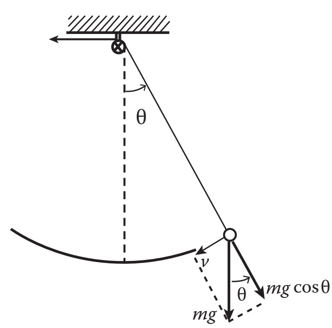


FIGURE 7.6 Simple pendulum of variable mass.

$$\frac{1}{2}ml^2\langle\dot{\theta}^2\rangle = \frac{1}{2}mgl\langle\theta^2\rangle = \frac{1}{2}E. \tag{7.22}$$

Hence, from Equations 7.21 and 7.22,

$$\Delta E \cong m\left(\frac{1}{2}g\langle\theta^2\rangle - l\langle\dot{\theta}^2\rangle\right)\Delta l \ \cong \frac{1}{l}\left(\frac{1}{2}E - E\right)\Delta l = -\frac{E}{2l}\Delta l$$

or

$$\frac{\Delta E}{E} \cong -\frac{1}{2}\frac{\Delta l}{l}. \tag{7.23}$$

On the other hand, because $\nu = 2\pi\sqrt{g/l}$, we have

$$\frac{\Delta\nu}{\nu} \cong -\frac{\Delta l}{2l}.$$

Combining this with Equation 7.23, we have

$$\frac{\Delta E}{E} \cong \frac{\Delta\nu}{\nu}.$$

This is the differential equation for E as a function of ν, and its solution is

$$\frac{E}{\nu} = \text{constant} = J. \tag{7.24}$$

Thus, during the extremely slow shortening of the pendulum's suspending string, the quantity $J = E/\nu$ remains constant. That is, the quantity J is an adiabatic invariant.

Adiabatic invariance played an important role in old quantum theory. According to Paul Ehrenfest, the adiabatically invariant quantities are quantizable (this is known as the principle of adiabatic invariance). We can therefore put J equal to an integral multiple of the fundamental unit h (the Planck constant):

$$J = \frac{E}{\nu} = nh$$

or

$$E = nh\nu,\ n = \text{integer}. \tag{7.25}$$

We thus obtain the energy levels of the harmonic oscillator in agreement with Planck's fundamental quantum assumption. As a second example, let us consider Bohr's quantization rule for orbiting electrons. If we imagine the strength of the attractive force between the electron and the nucleus to be slowly increased, the orbit will shrink, but because the force is still radial at all times, angular momentum $\vec{L}$ remains unchanged. Thus, $\vec{L}$ is an adiabatic invariant for a central orbit and can be quantized: $L = n\hbar$, where $\hbar = h/2\pi$. The adiabatic invariance also plays an important role in modem quantum theory; for example, a system that is stationary will stay stationary during adiabatic processes. Recently, there has been a renewed interest in adiabatic invariance because practical applications have been found in plasma physics, thermonuclear fusion, charged-particle accelerators, and even in galactic astronomy.

There is a geometrical interpretation of the invariant quantity $J = E/\nu$ of the simple pendulum. From Equation 7.7, we obtain the total energy E of the simple pendulum:

$$E = \frac{1}{2}m\dot{x}^2 + \frac{1}{2}\frac{mg}{b}x^2 = \frac{p^2}{2m} + \frac{1}{2}m\omega_0^2 x^2$$

or

$$\frac{p^2}{2mE} + \frac{x^2}{2E/\omega_0^2 m} = 1. \tag{7.26}$$

Equation 7.26 is an ellipse with a semi-major axis $a = (2E/m\omega_0^2)^{1/2}$ and a semi-minor axis $b = (2mE)^{1/2}$. A plot of p as a function of x is known as the particle's phase space diagram, several of which are shown in Figure 7.7. The arrows indicate the direction of the particle's trajectory. We can see from Equation 7.26 that as x decreases, the momentum p increases. When the initial condition E of the oscillation is changed, another ellipse results with different-sized major and minor axes. The ellipses never intersect.

Now, the area of the ellipse with semi-major and semi-minor axes a and b is equal to πab:

$$\oint p\,\mathrm{d}x = \pi ab$$

where the symbol ν means that we have to integrate over a whole period, that is, over the whole circumference of the ellipse. With $a = (2E/m\omega_0^2)^{1/2}$, $b = (2mE)^{1/2}$, and $\omega_0 = 2\pi\nu$, we have

$$\oint p\,\mathrm{d}x = \frac{E}{\nu} = J = nh. \tag{7.27}$$

Thus, the adiabatic invariant J is simply the area of the ellipse, and the phase space of a harmonic oscillator is quantized. Out of all the classically possible ellipses, only certain discrete ones are allowed. These discrete quantized ellipses represent the quantum states of the harmonic oscillator.

For the motions other than the harmonic oscillator, E/ν is no longer an adiabatic invariant, but the following integral is:

$$J = \oint p\,\mathrm{d}q. \tag{7.28}$$

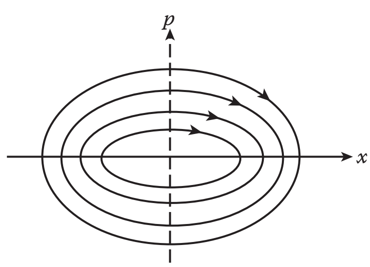


FIGURE 7.7 Phase path of a harmonic oscillator.

This integral (Equation 7.28) is the area in the phase space (p, q); J is known as an action variable in classical mechanics.

7.3 DAMPED HARMONIC OSCILLATOR

Free oscillation is an ideal physical case. In practice, dissipative forces are always present in any system. A simple example of the damped oscillator is the free motion of a particle of mass m attached to a (massless) spring in a resisting medium as shown in Figure 7.8. Let x be the displacement of m from position A at which point the spring is unstretched (of length a). The restoring force is $-kx$. We now limit ourselves to the case where the retarding force can be represented by $-bx$, and $b > 0$. The differential equation of motion of the particle is then given by

$$m\ddot{x} + b\dot{x} + kx = 0$$

or

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 = 0 \tag{7.29}$$

where $2\beta = b/m$, and $\omega_0^2 = k/m$. β is often called the damping parameter.

Equation 7.29 is a linear homogeneous differential equation of the second order with constant coefficients. A solution $x = Ae^{\alpha t}$ always exists, so we try

$$x = Ae^{\alpha t}. \tag{7.30}$$

Substituting this into Equation 7.29, we obtain

$$e^{\alpha t}\left(\alpha^2 + 2\alpha\beta + \omega_0^2\right) = 0.$$

Because $e^{\alpha t}$ cannot be zero (otherwise the solution is a trivial one), we are thus left with an algebraic equation of the second order in α, known as the auxiliary or the characteristic equation:

$$\alpha^2 + 2\alpha\beta + \omega_0^2 = 0 \tag{7.31}$$

which has two roots:

$$\alpha_1 = -\beta + \sqrt{\beta^2 - \omega_0^2}, \quad \alpha_2 = -\beta - \sqrt{\beta^2 - \omega_0^2}. \tag{7.32}$$

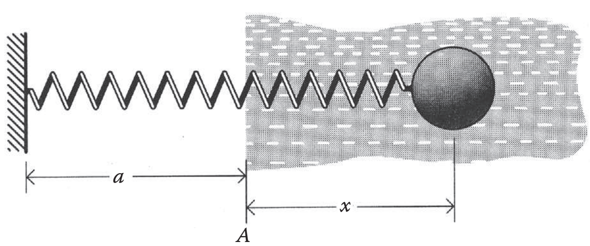


FIGURE 7.8 Damped oscillator.

For all $\beta \neq \omega_0$, this gives two independent solutions:

$$x_1 = A\exp(\alpha_1 t)$$

and

$$x_2 = B\exp(\alpha_2 t),$$

and so the general solution is

$$\begin{aligned} x(t) &= A_1 x_1 + A_2 x_2 \\ &= e^{-\beta t}\left[A_1 e^{\sqrt{\beta^2-\omega_0^2}\,t} + A_2 e^{-\sqrt{\beta^2-\omega_0^2}\,t}\right]. \end{aligned} \tag{7.33}$$

Constants A_1 and A_2 are dependent on the initial values of the position and velocity of the particle. m, β, and ω_0 are all positive numbers; there are three general cases of interest (Figure 7.9):

1. $\omega_0^2 > \beta^2$ underdamping (oscillatory motion)
2. $\omega_0^2 = \beta^2$ critical damping (not oscillatory)
3. $\omega_0^2 < \beta^2$ overdamping (not oscillatory)

We consider the underdamping case first.

1. Underdamping: $\omega_0^2 > \beta^2$. For convenience, we set

$$\omega_1^2 = \omega_0^2 - \beta^2. \tag{7.34}$$

The general solution (Equation 7.33) then becomes

$$x = e^{-\beta t}\left[A_1 e^{i\omega_1 t} + A_2 e^{-i\omega_2 t}\right].$$

Let

$$A_1 = \frac{1}{2}Ae^{i\theta}, \quad A_2 = \frac{1}{2}Ae^{-i\theta}$$

then we have

$$x = \frac{1}{2}e^{-\beta t}A\left[e^{i(\omega_1 t+\theta)} + e^{-i(\omega_1 t+\theta)}\right] = Ae^{-\beta t}\cos(\omega_1 t + \theta). \tag{7.35}$$

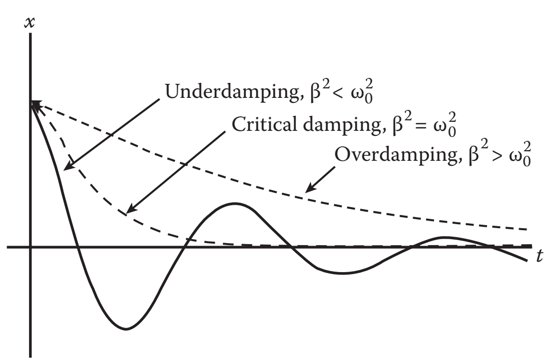


FIGURE 7.9 Three general cases of damped oscillations.

Because the constants A and θ are arbitrary, Equation 7.35 is the general solution. It is clear that the solution is oscillatory, but the frictional force has introduced a decaying exponential that can be called the damping factor. The quantity ω_1 is the natural frequency of the system (Remembering that ω_1 is related to ω_0 and the natural frequency of the undamped oscillator by Equation 7.34), we note that the natural frequency of the damped system is less than that for the undamped system.

Equation 7.35 states that displacements will lie between the two curves:

$$Ae^{-\beta t} \text{ and } -Ae^{-\beta t}$$

In Figure 7.10, the full solid line is the curve of Equation 7.35 with $\theta = 0$ or

$$x = Ae^{-\beta t}\cos\omega_1 t. \tag{7.36}$$

We see from Equation 7.36 that the solid curve will touch the envelope at points where $\cos\omega_1 t = \pm 1$ or, at times, $t_n = n\pi/\omega_1$, where n is an integer. Thus, we see that the points of contact are just a half period ½ τ_1 apart. The maxima and minima of Equation 7.35 do not occur quite at these points, but they are also separated by the same time interval. We may see this by differentiating Equation 7.36:

$$\dot{x} = -A\beta e^{-\beta t}\cos\omega_1 t - A\omega_1 e^{-\beta t}\sin\omega_1 t = 0$$

from which

$$\tan\omega_1 t = -\beta/\omega_1 \tag{7.37}$$

a function that is periodic with a period π/ω_1.

The ratio between the amplitudes of two successive maxima

$$\frac{Ae^{-\beta t_n}}{Ae^{-\beta(t_n+T)}} = e^{\beta T} \tag{7.38}$$

is called the decrement of the motion. The logarithm of this ratio, denoted by δ,

$$\delta = \ln e^{\beta T} = \beta T = \frac{2\pi\beta}{\omega_1} \tag{7.39}$$

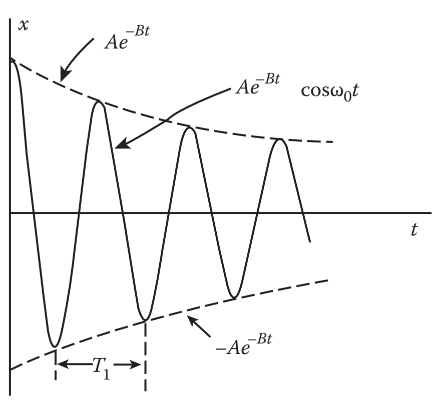


FIGURE 7.10 Underdamping oscillations.

is known as the logarithm decrement of the motion. If δ and ω_1 are measured, the damping constant β can be calculated from Equation 7.39.

The energy of the damped oscillator is not constant in time. Multiplying Equation 7.29 by $m\dot{x}$, we obtain

$$m\dot{x}\ddot{x} + b\dot{x}^2 + kx\dot{x} = 0.$$

This may be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{m\dot{x}^2}{2} + \frac{1}{2}m\omega_0^2 x^2\right) = -b\dot{x}^2 \le 0. \tag{7.40}$$

The left side is the time rate of change of the sum of the kinetic and potential energies; the right side is the rate at which energy is being dissipated by the damping force. The instantaneous loss of energy is therefore a function of time. It should be noted that Equation 7.40 is valid under all conditions: light, heavy, or critical damping.

It is of interest to consider the energy loss over a period or many periods of the motion. This can be obtained by comparing the total energy at two instants: t and $(t + T)$. At any time t, x and $\dot{x}$ are given by

$$x = Ae^{-\beta t}\cos(\omega_1 t + \theta)$$

$$\dot{x} = -Ae^{-\beta t}\omega_1 \sin(\omega_1 t + \theta) - Ae^{-\beta t}\cos(\omega_1 t + \theta).$$

The total energy is found to be

$$\begin{aligned} E(t) &= \frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2 \\ &= \frac{1}{2}mA^2 e^{-2\beta t}\left\{\left[\omega_1 \sin(\omega_1 t + \theta) + \beta\cos(\omega_1 t + \theta)\right]^2 + \omega_1^2\cos^2(\omega_1 t + \theta)\right\} \\ &= \frac{1}{2}mA^2 e^{-2\beta t} f(t) \end{aligned}$$

where $f(t)$ is the quantity in the braces, and it satisfies

$$f(t + T) = f(t).$$

Thus, we see that

$$E(t + T) = E(t)e^{-2\beta T}$$

or

$$E(T) = E(0)e^{-2\beta T}. \tag{7.41}$$

From Equation 7.41, it follows that the rate of energy loss is given at all times by

$$-\frac{1}{\langle E\rangle}\frac{\mathrm{d}\langle E\rangle}{\mathrm{d}t} = 2\beta. \tag{7.42}$$

Here, $\langle E \rangle$ is the average value of E over many periods of motion.

Equation 7.42 is a very useful formula for the damping parameter whose energy decay characteristics can be found. It shows that the energy decays exponentially in time: After a time interval T, the energy has a value that is $\exp(-2\beta T)$ times the energy at the beginning of the time interval. The decay can be characterized by the time τ required for the energy to drop to e^{-1}(= 0.368) of its initial value (see Figure 7.11):

$$E(\tau) = E(0)\exp(-2\beta\tau) = e^{-1}E(0)$$

from which we obtain

$$\tau = \frac{1}{2\beta} = \frac{m}{b}. \tag{7.43}$$

The time τ is often called the damping constant (or the time constant or characteristic time) of the system. In the limit of light damping, $\beta \to 0$ and $\tau \to \infty$, E is effectively constant, and the system thus behaves like an undamped oscillator.

The degree of damping of an oscillator is often specified by a dimensionless parameter Q, the quality factor, defined by

$$Q = \frac{\text{energy stored in the oscillator}}{\text{energy dissipated per radian}} \tag{7.44}$$

where "energy dissipated per radian" means the energy lost during the time it takes the system to oscillate through one radian. In the period $T = 2\pi/\omega_1$, the system oscillates through 2π radians. Thus, the time to oscillate through one radian is $T/2\pi = 1/\omega_1$.

According to Equation 7.42, the energy dissipated in a short time Δt is approximately given by

$$\Delta E \cong \left|\frac{\mathrm{d}E}{\mathrm{d}t}\right|\Delta t = 2\beta\, E\Delta t.$$

Now, one radian of oscillation requires time $\Delta t = 1/\omega_1$. Thus, the energy dissipated per radian is $2E\beta/\omega_1$, and accordingly, Equation 7.44 gives the quality factor Q:

$$Q = \frac{E}{2E\beta/\omega_1} = \frac{\omega_1}{2\beta} = \frac{\sqrt{\omega_0^2 - \beta^2}}{2\beta}. \tag{7.45}$$

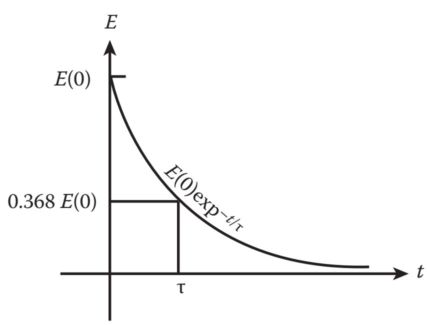


FIGURE 7.11 Damping time τ for underdamping motion.

A heavily damped system loses its energy rapidly, and its Q is low. For a lightly damped oscillator, $Q \cong \omega_0/2\beta$ and is much greater than 1. A tuning fork has a Q of a thousand or so; an undamped oscillator has infinite Q.

In summary, light damping slightly reduces the frequency of an oscillator and introduces a decay factor $\exp(-\beta t)$ in the amplitude. The average energy also decays exponentially, losing a fraction of its value per unit of time.

Example 7.2

A musician's tuning fork rings at *A* above middle *C*, 440 Hz. A sound-level meter indicates that the sound intensity decreases by a factor of 5 in 4 s. What is the quality factor Q of the tuning fork?

Solution:

The sound intensity from the tuning fork is proportional to the energy of oscillation. Because the energy of a damped oscillator decreases as $\exp(-2\beta t)$, we can find β by taking the ratio of the energy at $t = 0$ to that at $t = 4$ s:

$$5 = \frac{E(0)\exp(0)}{E(0)\exp[-4(2\beta)]} = \exp[4(2\beta)].$$

Hence,

$$4(2\beta) = \ln 5 = 1.6$$

or

$$2\beta = 0.4 \text{ s}^{-1}$$

and

$$Q \equiv \omega_0/2\beta = 2\pi(440)/0.4 \cong 7000.$$

The energy loss results primarily from the heating of the metal as it bends. Air friction and energy losses to the mounting point also contribute. The symmetrical design of a tuning fork minimizes loss to the mount.

Example 7.3

A paperweight suspended from a hefty rubber band has a period of 1.2 s, and the amplitude of oscillation decreased by a factor of 2 after three periods. What is the estimated Q of this system?

Solution:

From Equation 7.35, the amplitude is given by $A\exp(-\beta t)$. The ratio of the amplitude at $t = 0$ to that at $t = 3(1.2) = 3.6$ s is

$$2 = \frac{A\exp(0)}{A\exp(-3.6\beta)}.$$

Hence,

$$1.8(2\beta) = \ln 2 = 0.69$$

or

$$2\beta = 0.39\ \text{s}^{-1}$$

Therefore,

$$Q \equiv \frac{\omega_0}{2\beta} = \frac{2\pi/T}{0.39} = \frac{2\pi/1.2}{0.39} \cong 20.$$

A careful student may wonder whether it is justifiable to use the light damping result, $Q \cong \omega_0/2\beta$, when Q is so low. The approximation involved introduces errors of order $(2\beta/\omega_0)^2 = (1/Q)^2$. For $Q > 10$, the error is less than 1%.

It is interesting to note that the damping constants for the musician's tuning fork and for the rubber band are nearly the same. But the tuning fork has a much higher quality factor Q. The explanation for this is very simple: The tuning fork goes through many more cycles of oscillation in one damping time and correspondingly loses less of its energy per cycle. A rubber band exhibits a much lower quality factor Q than a tuning fork primarily because of the internal friction generated by the coiling of the long chain molecules. Several representative values of Q are given in Table 7.1.

2. $\omega_0 - \beta^2 = 0$: critically damped motion. Equation 7.33 now reduces to $x(t) = (A_1 + A_2)e^{-\beta t}$. This cannot be a general solution because it contains only one arbitrary constant. Let us return to the equation of motion, which now has the form

$$\ddot{x} + 2\beta\dot{x} + \beta^2 x = 0.$$

We can rewrite it in the form

$$\left(\frac{\mathrm{d}}{\mathrm{d}t} + \beta\right)\left(\frac{\mathrm{d}}{\mathrm{d}t} + \beta\right)x = 0. \tag{7.46}$$

Let

$$y = \left(\frac{\mathrm{d}}{\mathrm{d}t} + \beta\right)x. \tag{7.47}$$

TABLE 7.1
Several Typical Values of Q (Wide Variation May Be Expected)

2,501,400	Earth, for earthquake wave
10,000	Copper cavity microwave resonator
1000	Piano or violin string
10,000,000	Excited atom

Equation 7.46 now becomes

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+\beta\right)y=0$$

which has the solution

$$y = Ae^{-\beta t}.$$

Substituting this into Equation 7.47, we obtain

$$\left(\frac{\mathrm{d}}{\mathrm{d}t}+\beta\right)x = Ae^{-\beta t}.$$

Multiplying both sides by the integrating factor $e^{\beta t}$, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}(xe^{\beta t}) = A.$$

Integration gives

$$xe^{\beta t} = At + B$$

or

$$x = (At + B)e^{-\beta t}. \tag{7.48}$$

This is a general solution because it contains two arbitraries A and B.

Clearly, Equation 7.48 does not represent an oscillatory motion. The displacement approaches zero asymptotically at large values of time, and it has its maximum at $t = 0$ or at some positive values of t, depending on the ratio of B/A and $2m/b$ ($= 1/\beta$) as shown in Figure 7.12. These can be seen by differentiating Equation 7.48 with respect to time and setting the equation to zero. This will yield

$$t=-\frac{B}{A}+\frac{2m}{b}.$$

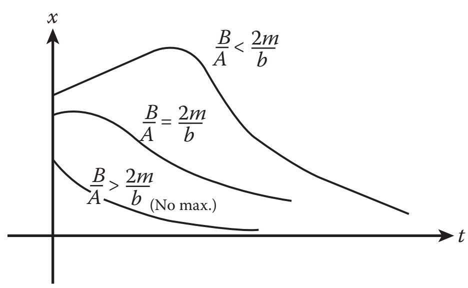


FIGURE 7.12 Motion of the critically damped oscillator.

3. $\omega_0^2 - \beta^2 < 0$: overdamped motion. For convenience, we set

$$\omega_2 = \sqrt{\beta^2 - \omega_0^2} > 0. \tag{7.49}$$

Equation 7.33 now reduces to

$$\begin{aligned} x(t) &= e^{-\beta t}[A_1 e^{\omega_2 t} + A_2 e^{\omega_2 t}] \\ &= A_1 e^{-(\beta-\omega_2)t} + A_2 e^{-(\beta+\omega_2)t} \end{aligned} \tag{7.50}$$

which does not represent an oscillatory motion. The system is prevented from undergoing oscillatory motion by the large damping force. The second term in Equation 7.50, with a larger decay constant in the exponent, will decay faster. Therefore, in the long run, the motion becomes a simple exponential decay:

$$x(t) = A_1 e^{-(\beta-\omega_2)t}.$$

Equation 7.50 also indicates that, depending on the initial conditions, there may be a change of sign of $x(t)$ before it approaches zero. If we take the initial time as zero, we obtain from Equation 7.50 the initial displacement x_0 and the initial velocity v_0:

$$x_0 = A_1 + A_2$$

and

$$v_0 = A_1(\beta - \omega_2) - A_2(\beta + \omega_2).$$

Solving the previous two equations for A_1 and A_2,

$$A_1 = \frac{v_0 + (\beta + \omega_2)x_0}{2\omega_2}$$

and

$$A_2 = -\frac{v_0 + (\beta - \omega_2)x_0}{2\omega_2}.$$

For positive values of x_0 and v_0, $A_1 > 0$, $A_2 < 0$, and A_1 is greater in magnitude than A_2. Thus, from Equation 7.50, we see that the displacement $x(t)$ will, at all times, remain positive, but it reaches a maximum before reaching zero. For positive x_0 and negative values of v_0 such that

$$v_0 < -(\beta + \omega_2)x_0$$

A_1 is negative while A_2 is positive. Thus, A_2 is larger in magnitude than A_1. Because the A_2 term decays more rapidly than the A_1 term, it follows that for some value of t and thereafter, the A_1 term will be the dominant term. In other words, the displacement will assume a negative value at some value of t and subsequently approach zero asymptotically as depicted in the curve in Figure 7.13.

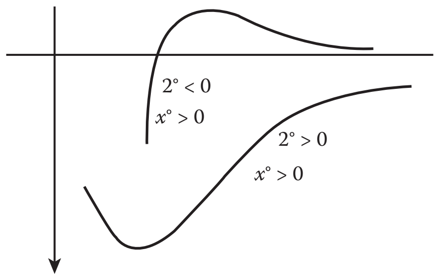


FIGURE 7.13 Motion of the overdamped oscillator.

Let us revisit Figure 7.9 in which we plot the displacement $x(t)$ versus t for the three cases under the same initial conditions $x(0) = x_0$ and $\dot{x}(0) = 0$. Figure 7.10 shows that the overdamped curve lies above the critically damped one and approaches zero at a slower rate. Because of the larger damping in the overdamped case, one might expect that the overdamped curve should lie below the critically damped one and approach zero at a faster rate. The explanation is simple: because of larger damping, the acceleration of an overdamped oscillator is smaller than that of the critically damped one at all times, and so is the velocity. Thus, a critically damped oscillator approaches equilibrium more rapidly than it does for either an overdamped or an underdamped oscillator. This is of great practical importance in the design of oscillator systems, for example, in a galvanometer.

7.4 PHASE DIAGRAM FOR DAMPED OSCILLATOR

A phase diagram is a very useful tool for the qualitative study of the behavior of a dynamic system in phase space. The phase diagram for the damped oscillator can be constructed as follows. For the case of underdamping, Equation 7.35 gives the general solution

$$x(t) = Ae^{-\beta t}\cos(\omega_1 t + \theta) \tag{7.35}$$

so

$$\dot{x}(t) = -Ae^{-\beta t}\left[\beta\cos(\omega_1 t + \theta) + \omega_1\sin(\omega_1 t + \theta)\right]. \tag{7.35a}$$

Equations 7.35 and 7.35a are the parametric equations of a family of spirals. Making the following transformations, we can see this more clearly:

$$u = \omega_1 x$$

and

$$w = \beta x + \dot{x}.$$

Then,

$$u = \omega_1 Ae^{-\beta t}\cos(\omega_1 t + \theta)$$

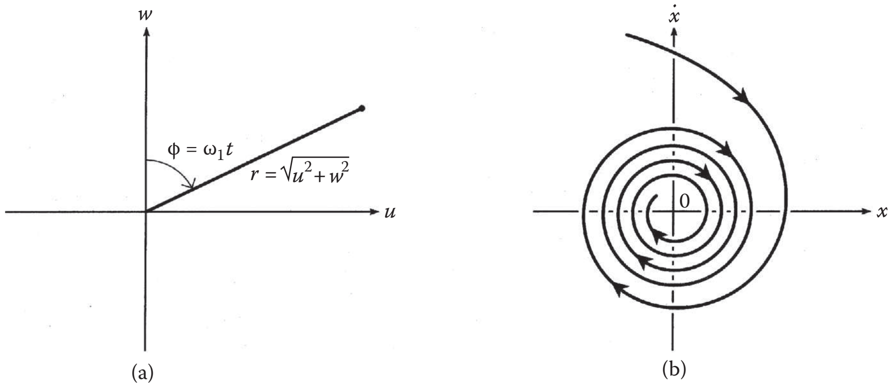


FIGURE 7.14 Phase diagram of an underdamped oscillator, (a) Polar coordinates r and ϕ and (b) phase path in the phase plan x-x.

and

$$w = -\omega_1 A e^{-\beta t} \sin(\omega_1 t + \theta).$$

Using polar coordinates r and ϕ (Figure 7.14a):

$$r = \sqrt{u^2 + w^2}, \quad \phi = \omega_1 t$$

we see that

$$r = \omega_1 A e^{-(\beta/\omega_1)\phi}$$

which is the equation of a logarithmic spiral. Now, the transformation from (x, y) to (u, w) is linear, so the form of the phase path in the phase plane $x - \dot{x}$ is the same as in the u–w plane. This is shown in Figure 7.14b, where the phase path is a spiral heading inward toward the asymptotically stable equilibrium state at the origin (0,0), which is called a point attractor or a focal point.

If $\beta > \omega_0$ (overdamping), the general solution is given by Equation 7.50, which can be put in the following form:

$$x = A \exp(-\beta t) \cosh(\omega_2 t + \alpha) \tag{7.50a}$$

and so

$$\dot{x} = A\exp(-\beta t)\left[\omega_2 \sinh(\omega_2 t + \alpha) - \beta \cosh(\omega_2 t + \alpha)\right]. \tag{7.50b}$$

A and α are constants. Equations 7.50a and 7.50b give the phase paths of the overdamped oscillator. Figure 7.15 shows phase paths for four alternative starts. The system moves back to its stable state of rest in a direct non-oscillatory fashion. It can be shown that the line $\dot{x} = -(\beta + \omega_2)x$ separated the phase plane into two parts. Again, we have a structurally stable point attractor at (0, 0).

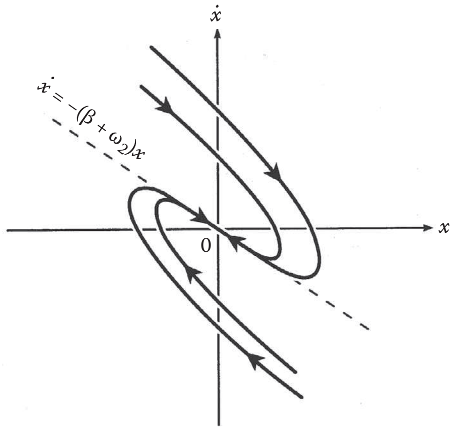


FIGURE 7.15 Phase diagram of an overdamped oscillator.

7.5 RELAXATION TIME PHENOMENA

A damped oscillator illustrates the concept of relaxation time when its mass is so small that the term $m\ddot{x}$ is negligible when compared with the restoring force kx and the dissipative force $b\dot{x}$. The equation of motion then reduces to

$$b\dot{x} + kx = 0$$

with the solution

$$x = x_0 e\ -kt/b$$

where the initial condition $x = x_0$ at $t = 0$ has been imposed. Obviously, x goes asymptotically to zero as t increases, and there is no oscillation. When $t = b/k$, $x = x_0 e^{-1}$, and it is said that the system has relaxed to the e^{-1}th of its initial displacement; the time $t = b/k$ is called the relaxation time. We can still speak of a relaxation time when the mass is not negligible for the overdamped or critically damped cases. The problem and the result are more complicated.

Every physical phenomenon in which there is a time lag in the attainment of the effect of a given cause will show a relaxation time. Hence, the notion of relaxation time is of considerable importance in physics. Consider another example, the process of charging a condenser of capacity C by means of a constant electromotive force (emf) E through a wire of resistance R. The relevant equation is

$$E = q/C + R\dot{q}$$

in which q is the charge on the condenser at instant t, and $\dot{q}$ is the instantaneous current. We assume that the inductance of the wire is negligible. The solution with $q = 0$ at $t = 0$ is

$$q(t) = EC[1 - e^{(-t/RC)}]. \tag{7.51}$$

The relaxation time here is $t = RC$, which is also referred to as the time constant of the RC circuit.

7.6 FORCED OSCILLATIONS WITHOUT DAMPING

We now consider oscillations of a system with an external force acted upon. These are called forced or driven oscillations. We shall consider our system to be a mass spring system and still consider small oscillations (this implies that the external force is weak).

The equation of motion now has the form

$$m\ddot{x} + kx = F(t)$$

or

$$\ddot{x} + \omega_0^2 x = F(t)/m \tag{7.52}$$

in which $F(t)$ is the external force acting on the system, and $\omega_0 = \sqrt{k/m}$. Equation 7.52 can also be obtained from Lagrange's equation with the following Lagrangian function L:

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 + xF(t).$$

7.6.1 Periodic Driving Force

It is especially interesting that the external force is a simple periodic function of t and ω: $F(t) = F_0 \cos\omega t$. Then, Equation 7.52 becomes

$$\ddot{x} + \omega_0^2 x = \frac{F_0}{m}\cos\omega t \tag{7.53}$$

which is an inhomogeneous linear differential equation with constant coefficients. The general solution for such equation consists of two parts:

$$x(t) = x_c + x_p$$

where the complementary solution x_c is the solution of the corresponding homogeneous equation, and x_p is the particular integral of the inhomogeneous equation.

Now, the complementary solution has the form

$$x_c = A\cos(\omega_0 t + \alpha).$$

Because the external force is a single harmonic driving force, we try a particular integral x_p in the form

$$x_p = B\cos\omega t. \tag{7.54}$$

Putting this back into Equation 7.53, we obtain

$$B = \frac{F_0}{m}\left(\omega_0^2 - \omega^2\right)^{-1}. \tag{7.55}$$

The dependence of B on ω is shown graphically in Figure 7.16.

The general solution of Equation 7.53 now takes the form

$$x(t) = A\cos(\omega_0 t + \alpha) + \frac{F_0/m}{\omega_0^2 - \omega^2}\cos\omega t. \tag{7.56}$$

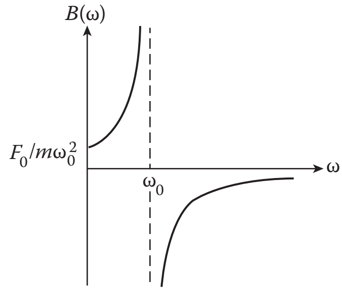


FIGURE 7.16 Amplitude of forced oscillations as a function of driving frequency ω.

The arbitrary constants A and α can be determined from the initial conditions x_0 and v_0.

Probably the most striking feature of a driven oscillator is the way in which a periodic force of fixed magnitude can produce very different results depending on its frequency. As we can see from Figure 7.16, if the driving frequency ω is made close to the natural frequency ω_0, then the amplitude of oscillations can be made very large by repeated applications of a quite small force. This is the phenomenon of resonance.

When ω is approaching ω_0, resonance occurs and the solution of Equation 7.56 is not valid anymore. In this case, we rewrite Equation 7.56 as

$$x(t) = a\cos(\omega_0 t + \alpha) + \frac{F_0/m}{\omega_0^2 - \omega^2}(\cos\omega t - \cos\omega_0 t)$$

where a has a value different from A. Now as $\omega \to \omega_0$, the second term is of the indeterminate form 0/0. But we can use the L'Hospital rule and obtain

$$x(t) = a\cos(\omega_0 t + \alpha) + \frac{F_0 t}{2m\omega}\sin\omega t.$$

Thus, the amplitude of oscillations in resonance increases linearly with time until the oscillations are no longer small, and the theory just cited becomes invalid.

Let us also ascertain the nature of small oscillations near resonance. For this purpose, we rewrite Equation 7.56 in the form

$$x(t) = a\cos\omega_0 t + b\sin\omega_0 t + \frac{F_0/m}{\omega_0^2 - \omega^2}\cos\omega t. \tag{7.57}$$

If $x = 0$ and $\dot{x} = 0$ at $t = 0$, then it can be shown that

$$a = -\frac{F_0/m}{\omega_0^2 - \omega^2}$$

and

$$b = 0$$

and Equation 7.57 becomes

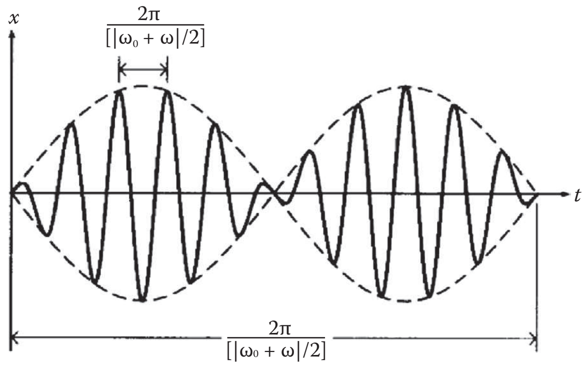


FIGURE 7.17 Beats.

$$\begin{aligned} x(t) &= \frac{F_0/m}{\omega_0^2-\omega^2}(\cos\omega t-\cos\omega_0 t) \\ &= \frac{F_0/m}{\omega_0^2-\omega^2}\sin\left(\frac{(\omega_0-\omega)t}{2}\right)\sin\left(\frac{(\omega_0+\omega)t}{2}\right). \end{aligned}$$

Thus, $(\omega_0 - \omega)$ is small near resonance, and $\sin[(\omega_0 + \omega)/2]$ will be a rapidly oscillating function compared with $\sin[(\omega_0 - \omega)/2]$. Figure 7.17 illustrates this motion. There is a periodic variation in the amplitude as well as a rapid periodic motion about the origin. This phenomenon is called "beats" and can be demonstrated with two tuning forks of nearly equal frequency. When they are sounded simultaneously, the periodic variation of the amplitude will be clear to the unaided ear.

7.6.2 Arbitrary Driving Forces

When the external force $F(t)$ is an arbitrary force, the equation of motion (Equation 7.52) can be integrated in a general form. Let us rewrite Equation 7.52 as

$$\frac{\mathrm{d}}{\mathrm{d}x}(\dot{x}+i\omega_0 x)-i\omega_0(\dot{x}+i\omega_0 x)=F(t)/m, \quad i=\sqrt{-1}$$

or

$$\frac{\mathrm{d}y}{\mathrm{d}t}-i\omega_0 y=F(t)/m \tag{7.58}$$

where

$$y=\dot{x}+i\omega_0 x. \tag{7.59}$$

The solution of Equation 7.58, when its right-hand side is set to zero, is

$$y=Ae^{i\omega_0 t}$$

where A is a constant. Thus, we try a solution of the inhomogeneous Equation 7.58 in the form

$$y=A(t)e^{i\omega_0 t}. \tag{7.60}$$

Putting this into Equation 7.58 yields

$$\dot{A}(t) = \frac{F(t)}{m} e^{-i\omega_0 t}. \tag{7.61}$$

Integration gives the solution of Equation 7.58:

$$y = e^{i\omega_0 t}\left[\int_0^t \frac{F(t')}{m} e^{-i\omega_0 t'}\,\mathrm{d}t' + y(0)\right] \tag{7.62}$$

where $y(0)$ is the constant of integration, the value of y at the time $t = 0$. Equation 7.62 is the required general solution of a first-order differential equation, Equation 7.58. From Equation 7.61, we see that the function $x(t)$ is given by the imaginary part of Equation 7.62, divided by ω_0:

$$x(t) = \frac{\operatorname{Im} y(t)}{\omega_0}. \tag{7.63}$$

Force $F(t)$ in Equation 7.62 or 7.63 must be written in real form.

The energy of the system cannot be conserved as the system gains energy from the external field. It is interesting to know the total energy transmitted to the system during the entire time. Let us determine this, assuming that the initial energy of the system is zero. The energy of the system is given by

$$E = \frac{1}{2} m\left(\dot{x}^2 + \omega_0^2 x^2\right) = \frac{1}{2} m\left|\dot{y}(t)\right|^2. \tag{7.64}$$

From Equation 7.62 with the lower limit of integration $-\infty$ instead of zero and with $y(-\infty) = 0$, we have

$$\left|y(\infty)\right|^2 = \frac{1}{m^2}\left|\int_{-\infty}^{\infty} F(t) e^{-i\omega_0 t}\,\mathrm{d}t\right|^2.$$

Putting this into Equation 7.64, we obtain the energy transferred:

$$E = \frac{1}{2m^2}\left|\int_{-\infty}^{\infty} F(t) e^{-i\omega_0 t}\,\mathrm{d}t\right|^2. \tag{7.65}$$

If $F(t)$ acts only for a short time in comparison with $1/\omega_0$, we can set $\exp(-i\omega_0 t) \cong 1$. Then, Equation 7.65 reduces to

$$E = \frac{1}{2m^2}\left(\int_{-\infty}^{\infty} F(t)\,\mathrm{d}t\right)^2. \tag{7.66}$$

This is a familiar result: a force of short duration gives the system an impulse $\int_{-\infty}^{\infty} F(t)\,\mathrm{d}t$ without producing appreciable displacement.

7.7 FORCED OSCILLATIONS WITH DAMPING

Forced oscillation without damping is an ideal situation. In any real oscillating system, energy dissipation of some kind always exists. Damping exists whenever there is energy dissipation. We now study forced oscillations with damping. The driving force is still a sinusoidal one, $F(t) = F_0 \cos\omega t$. Thus, a body executing forced oscillations under damping is now acted upon by three forces simultaneously: the restoring force $-kx$, the frictional force $-bx$, and the external force $F_0 \cos\omega t$; and the equation of motion of the body has the form

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = A\cos\omega t \tag{7.67}$$

where $A = F_0/m$, β, and ω are defined as before.

Equation 7.67, like Equation 7.63, is an inhomogeneous, linear differential equation with constant coefficients. Its general solution has two parts:

$$x(t) = x_c(t) + x_p(t).$$

The complementary solution $x_c(t)$ is the same as that given by Equation 7.33. We now seek a particular integral $x_p(t)$ in the form

$$x_p(t) = D\cos(\omega t - \delta) \tag{7.68}$$

with the same angular frequency ω of the driving force. It is straightforward to find the amplitude D and the phase constant δ by substituting Equation 7.68 into Equation 7.67. However, it is more convenient to use complex forms, and so we replace $\cos\omega t$ on the right-hand side of Equation 7.67 by $\exp(i\omega t)$, and remember that $\cos\omega t$ is the real part of this exponential function. Equation 7.67 now becomes

$$\ddot{x} + 2\beta\dot{x} + \omega_0^2 x = Ae^{i\omega t}. \tag{7.69}$$

For the particular solution, we now try

$$x = De^{i\omega t} \tag{7.70}$$

from which we have

$$\dot{x} = i\omega De^{i\omega t}$$

and

$$\ddot{x} = -D\omega^2 e^{i\omega t}.$$

Substituting these into Equation 7.69, we obtain

$$D = \frac{A}{\omega_0^2 - \omega^2 + 2i\beta\omega} = \frac{A\left(\omega_0^2 - \omega^2 - 2i\beta\omega\right)}{\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2}. \tag{7.71}$$

Writing

$$\frac{\omega_0^2 - \omega^2 - 2i\beta\omega}{\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2} = e^{-i\delta} = \cos\delta - i\sin\delta \tag{7.72}$$

from which we have

$$\cos\delta = \frac{\omega_0^2 - \omega^2}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2}} \quad \text{and} \quad \sin\delta = \frac{2\beta\omega}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2}} \tag{7.72a}$$

and

$$\delta = \tan^{-1}\left(\frac{2\beta\omega}{\omega_0^2 - \omega^2}\right). \tag{7.72b}$$

With the substitution of Equation 7.72, D in Equation 7.72 becomes

$$D = \frac{A}{\left[\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2\right]^{1/2}} e^{-i\delta} \tag{7.73}$$

and Equation 7.70 takes the following form:

$$x = \frac{A}{\left[\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2\right]^{1/2}} e^{i(\omega t - \delta)}.$$

Finally, taking the real part of the preceding equation, we find the particular integral of Equation 7.67:

$$x_p = \frac{F_0/m}{\left[\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2\right]^{1/2}} \cos(\omega t - \delta). \tag{7.74}$$

Adding this to the complementary solution x_c, we obtain the general solution of the equation of motion (Equation 7.66). However, x_c, given by Equation 7.33, dies away exponentially with time through the damping factor $\exp(-\beta t)$, and it, therefore, represents a transient solution. Thus, after a long time $t \gg 1/\beta$, the general solution reduces to x_p, and we can call the particular integral x_p the steady-state solution. Thus, no matter what initial conditions we choose, solely the driving force ultimately governs the resultant motion. It differs, however, in phase from the driving force, lagging behind the driving force by δ radians. That is, there is delay between the action of the driving force and the response of the system. We can see from Equation 7.72b that for a fixed ω_0, the response of the system is in phase ($\delta = 0$) with the driving force for $\omega = 0$, and the response is 90° out of phase ($\delta = \pi/2$) with the driving force at $\omega = \omega_0$. For a very large value of $\omega(\omega \to \infty)$, the motion of the system is 180° ($\delta = \pi$) out of phase with the driving force. For convenience, we denote the amplitude of the steady-state solution x_p by B:

$$B = \frac{F_0/m}{\left[\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2\right]^{1/2}}. \tag{7.75}$$

The behavior of B and δ as functions of ω depends markedly on the ratio $2\beta/\omega_0$ as the sketches in Figure 7.18 show.

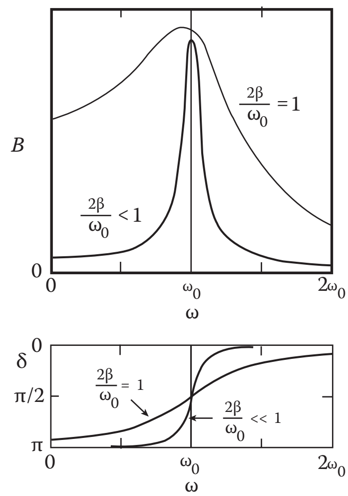


FIGURE 7.18 Behavior of B and δ as a function of ω.

7.7.1 Resonance

We can see from Equation 7.75 that the amplitude B of the oscillation is proportional to that of the driving force F_0 and depends on the relationship between the frequency ω of the driving force and the natural frequency ω_0 of the system. At some particular frequency, say, $\omega = \omega_R$, amplitude B of the oscillation reaches its maximum value (the resonance). To find the frequency ω_R at which amplitude B is maximum, we set

$$\frac{dB}{d\omega} = \frac{d}{d\omega}\left[\frac{F_0/m}{\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2}}\right]_{\omega=\omega_R} = 0$$

from which we find

$$\omega_R = \sqrt{\omega_0^2 - 2\beta^2}\ . \tag{7.76}$$

Thus, ω_R is reduced as the damping coefficient β is increased. There is no resonance if $2\beta^2 > \omega_0^2$ for, in this case, ω_R is imaginary.

The steady-state amplitude at the resonance frequency ω_R, which we shall denote by B_R, can be obtained by setting $\omega = \omega_R$ and $\omega_0^2 - \omega_R^2 = 2\beta^2$ in the amplitude of Equation 7.75:

$$B_R = \frac{F_0/m}{\sqrt{4\beta^4 + 4\beta^2\left(\omega_0^2 - 2\beta^2\right)}} = \frac{F_0/m}{2\beta\sqrt{\omega_0^2 - \beta^2}}\ . \tag{7.77}$$

In the case of weak damping, $\omega_0 \gg \beta$, β^2 can be ignored, and we have $\omega_R \cong \omega_0$ and

$$B_R \cong \frac{F_0}{2m\beta\omega_0} = \frac{F_0}{b\omega_0}\ . \tag{7.78}$$

Thus, the maximum value of amplitude B_R is inversely proportional to the damping coefficient b; for this reason, the damping in the system cannot be neglected at resonance even if it is slight.

The origin of the amplification of the oscillations by resonance can be understood by considering the relationship between the phases of the external force $F(t)$ and the velocity $\dot{x}$. When $\omega \neq \omega_0$, there is a difference in phase ($\delta \neq 0$), and therefore, force $F(t)$ is in the opposite direction from the velocity during a certain fraction of each period, and there is then a tendency for the motion to be retarded instead of accelerated. At resonance, however, the phases of force $F(t)$ and velocity are the same ($\omega_R = \sqrt{\omega_0^2 - 2\beta^2} \approx \omega_0$ for light damping, and thus, $\delta = 0$). The force, therefore, always acts in the direction of the motion and continually pushes it.

The sharpness of the resonance peak is frequently of interest. Let us consider the case of weak damping $\beta \ll \omega_0$ and $\omega_R \cong \omega_0$. The expression for the steady-state amplitude, Equation 7.76, can then be simplified. Near resonance ($\omega = \omega_0$), we can make the following approximations:

1. $\omega_0^2 - \omega^2 = (\omega + \omega_0)(\omega - \omega_0) \cong 2\omega_0(\omega - \omega_0)$
2. $4\beta^2\omega^2 \cong 4\beta^2\omega_0^2$

Together with Equation 7.78, these approximations allow us to rewrite Equation 7.96 in the following approximate form:

$$B = \frac{F_0/m}{\left[\left(\omega_0^2 - \omega^2\right)^2 + 4\beta^2\omega^2\right]^{1/2}}$$
$$\cong \frac{F_0/m}{\left[4\omega_0^2(\omega_0 - \omega)^2 + 4\beta^2\omega_0^2\right]^{1/2}} = \frac{F_0/m}{2\omega_0\sqrt{(\omega_0 - \omega)^2 + \beta^2}} \tag{7.79}$$

from which we can readily see that when $|\omega_0 - \omega| = \beta$ or, equivalently, when $\omega = \omega_0 \pm \beta$.

Equation 7.79 then gives $B = B_R/\sqrt{2}$. That is, the amplitude falls off to $1/\sqrt{2}$ of its peak value when $\omega = \omega_0 \pm \beta$ (this occurs when the two terms in the denominator of Equation 7.79 become comparable in magnitude). For this reason, β is called the half-width of the resonance. Now, the height of the peak (for given F_0) is inversely proportional to β, and accordingly, it is also inversely proportional to the width. Thus, the smaller the damping, the narrower the resonance curve and the higher its peak.

Energy considerations provided useful information about the unforced damped oscillator, and these considerations will be useful to us in the present case—the forced oscillator. For the steady-state motion, the amplitude is constant in time as indicated clearly in Equation 7.74, and from this equation, we have

$$T = \frac{1}{2}m\dot{x}^2 = \frac{1}{2}m\omega^2 B^2 \sin^2(\omega t - \delta)$$

and

$$V = \frac{1}{2}kx^2 = \frac{1}{2}kB^2\cos^2(\omega t - \delta)$$

so

$$E = T + V$$
$$= \frac{1}{2}B^2\left[m\omega^2\sin^2(\omega t - \delta) + k\cos^2(\omega t - \delta)\right].$$

Energies are time-dependent, and our analysis is simplified if we focus on time-average values. Now,

$$\frac{1}{2\pi/\omega}\int_0^{2\pi/\omega}\sin^2(\omega t-\delta)\,\mathrm{d}t=\frac{1}{2\pi/\omega}\int_0^{2\pi/\omega}\cos^2(\omega t-\delta)\,\mathrm{d}t=\frac{1}{2}.$$

Thus, for an average over one period of oscillation, we have

$$\langle T\rangle=\frac{mB^2\omega^2}{4},\quad \langle V\rangle=\frac{kB^2}{4}$$

and

$$\langle E\rangle=\frac{B^2(m\omega^2+k)}{4}=\frac{mB^2\left(\omega^2-\omega_0^2\right)}{4}.$$

Let us now find the frequency ω_k at which $\langle T\rangle$ is a maximum:

$$\left.\frac{\mathrm{d}\langle T\rangle}{\mathrm{d}\omega}\right|_{\omega=\omega_k}=0 \tag{7.80}$$

from which we obtain

$$\omega_k=\omega_0. \tag{7.81}$$

Thus, the kinetic energy resonance occurs at the natural frequency ω_0 of the system for an undamped oscillator. Because the potential energy is proportional to the square of the amplitude, we expect that the potential energy reaches its maximum at frequency ω_0. We next consider how $\langle E\rangle$ varies as a function of ω. Using Equation 7.75 for B, we have

$$\langle E\rangle=\frac{F_0^2}{4m}\frac{\omega_0^2+\omega^2}{\left(\omega_0^2-\omega^2\right)^2+4\beta^2\omega^2}. \tag{7.82}$$

This relationship is exact but awkward. It can be simplified for the case of weak damping. Figure 7.19 gives the sketch of $\langle E\rangle$ for $2\beta/\omega_0 = 0.1$ and $2\beta/\omega_0 = 0.4$, which shows clearly that

1. $\langle E\rangle$ is maximum (the resonance) when $\omega = \omega_0$.
2. For β that is sufficiently small, $\langle E\rangle$ is effectively zero except near resonance.

Because of the second fact, there is not much error introduced by replacing ω by ω_0. However, caution must be exercised in the term $\left(\omega_0^2-\omega^2\right)^2$ in the denominator of the expression for $\langle E\rangle$. We simplify this term as

$$\omega_0^2-\omega^2=(\omega_0+\omega)(\omega_0-\omega)\cong 2\omega_0(\omega_0-\omega).$$

With this expression, $\langle E\rangle$ takes the simple form

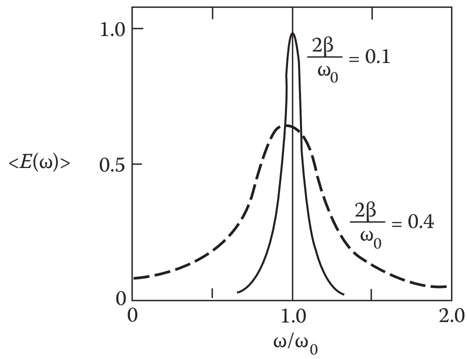


FIGURE 7.19 $<E>$ as a function of ω.

$$\langle E(\omega)\rangle = \frac{1}{4}\frac{F_0^2}{m}\frac{2\omega_0^2}{4\omega_0^2(\omega-\omega_0)^2+4\omega_0^2\beta^2} = \frac{1}{8}\frac{F_0^2}{m}\frac{1}{(\omega-\omega_0)^2+\beta^2}$$

which shows clearly that resonance occurs at ω_0.

Another way of designating the sharpness of the resonance peak is in terms of Q, the quality factor. Recall that Q is defined as the ratio of energy stored in the oscillator to energy lost per radian of oscillation, and it has, for a weak damped oscillator without driving force, the value $\omega_0/2\beta$. Here, ω_0 is the resonance frequency, and 2β is the frequency width of the resonance curve. The same oscillator, when driven by an external force, has a resonance curve with the frequency width $\Delta\omega = 2\beta$ and the resonance frequency ω_0. Thus, the ratio of resonance frequency to resonance width, $\omega_0/\Delta\omega$, is $\omega_0/2\beta = Q$. If we had applied the definition of Q in terms of energy, the result would be the same. Thus, Q is often defined by

$$Q = (\text{resonance frequency})/(\text{resonance width of the resonance curve}).$$

We now indicate how the behavior of the resonance system changes as the Q of the system is changed. First, we replace the damping parameter 2β in Equations 7.72b and 7.75 for δ and B by Q, $2\beta = \omega_0/Q$, giving us

$$B(\omega) = \frac{F_0/m}{\left[\left(\omega_0^2-\omega^2\right)^2+(\omega\omega_0/Q)^2\right]^{1/2}} = \frac{F_0\omega_0/\omega}{\left[(\omega_0/\omega-\omega/\omega_0)^2+1/Q^2\right]^{1/2}} \tag{7.83}$$

$$\tan\delta = \frac{\omega\omega_0/Q}{\omega_0^2-\omega^2} = \frac{1/Q}{\omega_0/\omega-\omega/\omega_0}. \tag{7.84}$$

Figure 7.20 shows the resonance and phase curves for different values of Q. Clearly the frequency ω_0 is an important property of the resonance system, even though it is not (except for zero damping) the frequency with which the system would oscillate when left to itself. Amplitude B passes through a maximum for all values of Q except the most heavily damped systems. This maximum amplitude B occurs at a frequency ω_R that is less than ω_0. If we denote by B_0 the amplitude F_0/k obtained for $\omega \to 0$, then we can show that

$$\omega_R = \omega_0\left(1-\frac{1}{2Q^2}\right)^{1/2}$$

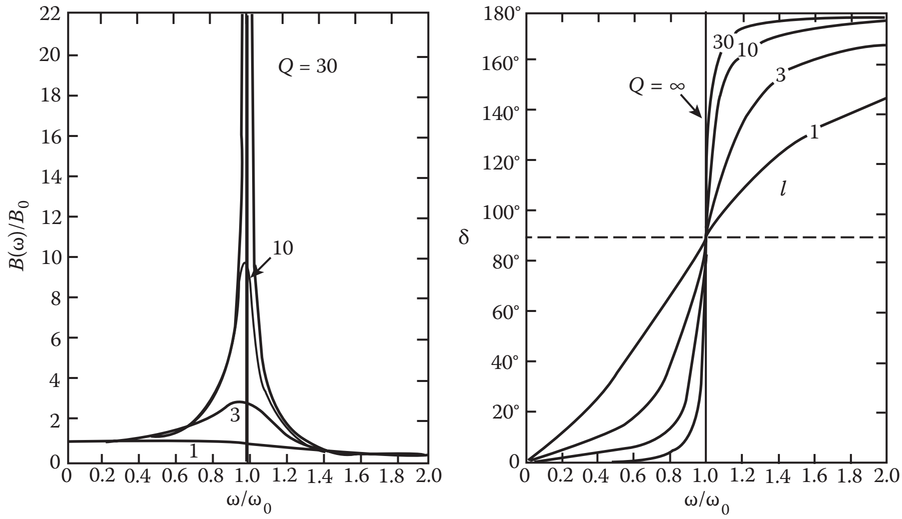


FIGURE 7.20 Resonance and phase curves for different values of Q. (Courtesy of Prof. A.P. French.)

and

$$B_R = \frac{B_0 Q}{(1-1/4Q^2)^{1/2}}.$$

The larger the value of Q, the sharper the resonance curve. The sharpness of the resonance curve means that the system will not respond unless driven very near its resonance frequency. Certain atomic systems can have a Q greater than 10^8. The response frequency is determined by atomic constants, and the frequency of oscillation is essentially independent of external influences. Frequencies from such atomic clocks are so accurate that they have superseded astronomical time standards.

7.7.2 Power Absorption

It is of interest and also often a matter of importance to know at what rate energy must be fed into a driven oscillator to maintain its oscillations at a fixed amplitude. The instantaneous power input P can be calculated as the driving force times the velocity as in any other dynamic situation:

$$P = \frac{\mathrm{d}W}{\mathrm{d}t} = \frac{F\mathrm{d}x}{\mathrm{d}t} = F\dot{x}. \quad (7.85)$$

To see its connection with the system's energy dissipation, we multiply Equation 7.67 by $\dot{x}$

$$m\dot{x}\ddot{x} + b\dot{x}^2 + kx\dot{x} = \dot{x}F_0 \cos \omega t$$

or

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{1}{2}m\dot{x}^2 + \frac{1}{2}kx^2\right) + b\dot{x}^2 = \dot{x}F_0 \cos \omega t = \dot{x}F.$$

The right-hand side of the last equation is the instantaneous power input P, the rate at which energy is being supplied by the driving force $F_0\cos\omega t$. The first term on the left-hand side of this equation is the time rate of change of the sum of the kinetic and potential energies, and the second term is the rate at which energy is being dissipated by the damping force. Now, from Equations 7.74 and 7.75, we obtain

$$\dot{x} = -B\omega\sin(\omega t - \delta).$$

Substituting this into Equation 7.85 gives

$$P = \dot{x}F_0\cos\omega t = -B\omega F_0\cos\omega t\sin(\omega t - \delta)$$
$$= BF_0\omega(\cos^2\omega t\sin\delta - \cos\omega t\sin\omega t\cos\delta).$$

The second term within the parentheses may, at times, outweigh the first term, and at these times, the driving force is extracting energy from the system. Thus, the driving force is alternately introducing and withdrawing energy to and from the system. We are interested in $\langle P\rangle$, the average power over many cycles, and because

$$\langle\cos^2\omega t\rangle = \frac{1}{\tau}\int_0^\tau \cos^2\omega t\,\mathrm{d}t = \frac{1}{2}, \quad \langle\sin\omega t\cos\omega t\rangle = 0$$

so that the average power input is given by

$$\langle P\rangle = \frac{1}{2}BF_0\omega\sin\delta.$$

Taking the value of B from Equation 7.75, this becomes

$$\langle P(\omega)\rangle = \frac{1}{2kQ}\frac{F_0^2\omega_0}{(\omega_0/\omega - \omega/\omega_0)^2 + 1/Q^2}. \tag{7.86}$$

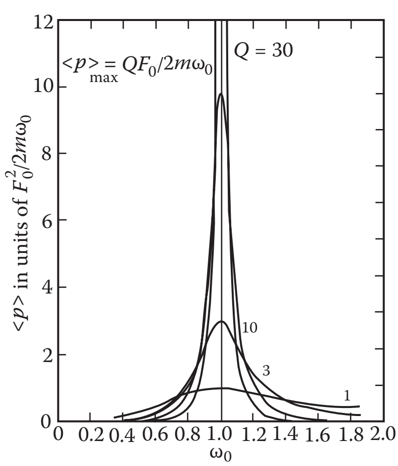


FIGURE 7.21 $<P>$ as a function of ω for various Q. (Courtesy of Prof. A.P. French.)

We see that $\langle P\rangle$, like velocity $\dot{x}$, passes through a maximum at $\omega = \omega_0$ with a maximum value given by

$$\langle P\rangle = \frac{\omega_0 Q F_0^2}{2k} = \frac{Q F_0^2}{2m\omega_0}.$$

Figure 7.21 shows the dependence of $\langle P\rangle$ on ω for various Q.

As an illustration of the generality of the Lagrangian dynamics, we considered its application to an LC-circuit in Chapter 4. We now consider the following example.

Example 7.4: A Driven Electric Oscillator

Consider an electric circuit containing an inductance L, resistance R, and capacitance C, connected in series with an applied emf $V = V_0\cos\omega t$. Show that the equation of the circuit corresponds to the equation of motion of a forced oscillator with damping, Equation 7.69.

Solution:

Call the current at any instant I and $\pm q$ the charges on the plates of the capacitor. The potential difference across L, R, and C are $L(\mathrm{d}i/\mathrm{d}t)$, iR, and q/C, respectively. Because the applied emf must equal the sum of the potential differences that it maintains, we write the equation of the circuit:

$$L\frac{\mathrm{d}i}{\mathrm{d}t} + iR + \frac{q}{C} = V_0\cos\omega t. \qquad (7.87)$$

Because $i = \mathrm{d}q/\mathrm{d}t$, we can rewrite Equation 7.87 in the form

$$\frac{\mathrm{d}^2 q}{\mathrm{d}t^2} + \frac{R}{L}\frac{\mathrm{d}q}{\mathrm{d}t} + \frac{1}{LC}q = \frac{V_0}{L}\cos\omega t. \qquad (7.88)$$

Equation 7.88 is exactly the same type as Equation 7.67. If we set $R/L = 2\beta$, $1/LC = \omega_0^2$, and $A = V_0/L$, then Equation 7.88 becomes identical to Equation 7.67, although the symbols and the quantifiers they stand for are different. The solution will again consist of two exponential transient terms plus a harmonic term as in Equation 7.67. We shall not consider these fully here.

There are forced oscillations with damping of all kinds, many of which have little in common with a mass attached to a spring or with the electric series circuit just discussed. A simple example is the Helmholtz resonator, which is one of the acoustical systems that behave as forced oscillators with damping. The resonator consists of a cavity that communicates with the external medium through a small opening (Figure 7.22). The volume of, say, air in the cavity can be compressed and expanded by the flow of air in the opening. This cavity volume thus provides the stiffness element of the system. If a certain volume of air, X, is displaced in the opening and enters the cavity, the pressure in the cavity will increase by an amount that we shall call p_1. We see that the volume displacement X is analogous to the linear displacement x of the spring or charge q of the electric circuit, and p_1 corresponds to the restoring force F of the spring. In the electric case, we saw that q/V is the electric capacitance of a capacitor, and we can define X/p_1 (= C_a) as the acoustical capacitance.

The plug of air that moves back and forth in the opening provides the inertia element. We have seen that mass is a measure of inertia, and $m = F/\ddot{x}$ (Newton's second law). For the acoustic resonator, we can define p_2/X (= M) as the acoustical inertance, where p_2 is the driving pressure necessary to give the volume acceleration X.

What is the damping mechanism? Acoustic energy can be dissipated in the opening of the Helmholtz resonator in two ways. First, in the small, tube-like opening, filled with cotton or a

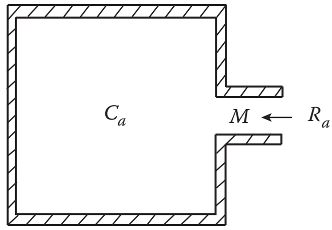


FIGURE 7.22 Helmholtz resonator.

similar porous material, energy is dissipated because of the viscosity of the air. Second, the motion of the air at the opening causes the sound to radiate into the medium outside, thus dissipating acoustical energy. Similar to electric resistance $R = V/i = V/\dot{q}$ (Ohm's law), we define the acoustical resistance R_a as $R_a = p_3/\dot{X}$, where p_3 is the pressure needed to derive the volume current through the dissipative element.

If the sound waves of a given frequency approach the opening of the resonator at a steady state, we have an alternating applied pressure difference that can be represented by $p_0\cos\omega t$. Equalizing the applied force to the sum of the kinetic reaction, resistance force, and restoring force, we obtain

$$M\ddot{X} + R_a\dot{X} + \frac{1}{C_a}X = p_0 \cos\omega t.$$

This equation is of the same form as Equation 7.67.

Another interesting example is the classical absorption of energy by an electron in an atom when it is irradiated by a plane electromagnetic wave. H.A. Lorentz addressed this question in the early part of the last century. An atom consists of a nucleus of positive charge and orbiting electrons that carry negative charges. The instantaneous positions of the electrons are very uncertain, but when looked at over a sufficiently long time ($>10^{-16}$ s), the electronic structure appears as a quasi-stationary, smeared-out charged cloud surrounding the nucleus. The cloud is approximately spherically symmetric, dense in the center, and decreasing in density at the edge. In this picture, the atom does not have an intrinsic dipole moment of its own. However, when the atom is acted upon by an electric field, the electron cloud becomes displaced relative to the nucleus, and a dipole moment is generated. The electron behaves approximately as if it were bound to its equilibrium position by a quasi-elastic, spring-like restoring force $-kx$. If we let the x-axis lie along the oscillation direction of the field, the equation of motion of the electron is given by

$$m\ddot{x} = -kx - m\gamma\dot{x} + qE_0 \cos\omega t \tag{7.89}$$

where m is the mass of the electron, E_0 is the field amplitude, and $-m\gamma\dot{x}$ represents radiative damping. The energy dissipates through y in continual re-radiation of some of the absorbed energy. According to classical electrodynamics, the damping coefficient γ is given by

$$\gamma = 2q^2\omega^2/3mc^2. \tag{7.90}$$

The term $-kx$ describes a certain elastic force by which an electron is pulled back toward its equilibrium position after having been displaced from it. When compared with Equation 7.74, the steady-state solution is seen to be

$$x = \frac{qE_0}{m\sqrt{\left(\omega_0^2 - \omega^2\right)^2 + \omega^2\gamma^2}} \cos(\omega t - \delta). \tag{7.91}$$

From Equation 7.87, the average rate of absorption of energy is

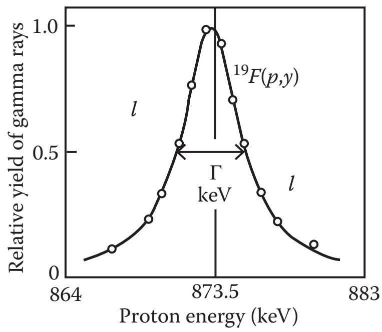


FIGURE 7.23 Nuclear resonance curve.

$$\langle P(\omega)\rangle = \langle \dot{x}qE_0\cos\omega t\rangle = \frac{q^2E_0^2\omega\gamma}{2m\left[\left(\omega_0^2-\omega^2\right)^2+\omega^2\gamma^2\right]} \tag{7.92}$$

and the half-width of the resonance is $\Delta = \beta = y/2$.

The response of an atom to an applied electric field is a subject that would require us to use the quantum theory of atomic structure. It is interesting to note that the absorption profile calculated from Lorentz's model agrees very well with the profile calculated by quantum theory. In a certain sense, the electron of quantum theory can be considered to be bound by a linear restoring force. Thus, Lorentz's model can yield quite adequate results. Very often, a simplified yet quite adequate classical model can provide some insight into full quantum theory.

So far, our discussion has been limited to the behavior of a simple physical system, and we found that the quantities that characterize resonance are frequency, amplitude, and absorption power. When a driving force acts upon the system, one of these parameters, frequency, is varied. The response of the system, as measured by its amplitude and phase or by the power absorption, undergoes rapid changes as the frequency passes through a certain value. Can we carry over these ideas to the resonance behavior of other physical systems? We shall find that the quantities that characterize resonance are not always frequency, amplitude, and absorption power. As an example, consider nuclear resonance. Figure 7.23 shows a graph of the relative yield of gamma rays as a target of fluorine is bombarded with protons of different energies approximately 885 keV:

$${}_1^1H + {}_9^{19}F \to {}_{10}^{20}Ne^* \to {}_{10}^{20}Ne + \gamma.$$

The resonance system is the compound nucleus ${}_{10}^{20}Ne^*$ (${}_{10}^{20}Ne$ in an excited state). This compound nucleus is unstable and decays through the emission of gamma rays. The controllable parameter is the energy of the bombarding proton, not a frequency, and the response of the system is measured in terms of the probability that an incident proton will cause a gamma ray to be produced, not in terms of amplitude or absorbed power.

7.8 OSCILLATOR UNDER ARBITRARY PERIODIC FORCE

The solution we obtained in the previous section can be generalized to the case where the external driving force $F(t)$ is a periodic but non-sinusoidal function. In this general case, it is convenient to use the principle of superposition and the Fourier series method. The superposition principle is applicable to any system governed by a linear differential equation. In our case, the principle states that if the external driving force $F(t)$ acting on a damped oscillator is given by a superposition of n force functions

$$F(t) = \sum_{j=1}^{n} F_j(t) \tag{7.93}$$

such that the differential equations

$$m\ddot{x}_j + b\dot{x}_j + kx_j = F_j(t), \quad j = 1,2,\ldots,n \tag{7.94}$$

are individually satisfied by the functions $x_j(t)$, then the solution of the differential equation of motion

$$m\ddot{x} + b\dot{x} + kx = F(t) \tag{7.95}$$

is given by the superposition

$$x(t) = \sum_{j=1}^{n} A_j x_j(t). \tag{7.96}$$

This is an extremely useful result. If the driving force is periodic with a period τ, $F(t + \tau) = F(t)$, then the force function can be expressed as a superposition of the harmonic terms according to Fourier's theorem.

7.8.1 Fourier's Series Solution

We now digress to review Fourier's method for those who may need a review.

Let $f(t)$ be any single-valued function that is defined in the interval $(-L, L)$ and has a period $2L$, $f(t + 2L) = f(t)$. The Fourier series or Fourier expansion corresponding to $f(t)$ is given by

$$f(t) = \frac{1}{2}a_0 + \sum_{j=1}^{n}\left\{a_j \cos\frac{j\pi t}{L} + b_j \sin\frac{j\pi t}{L}\right\} \tag{7.97a}$$

where the coefficients a_j and b_j are

$$a_j = \frac{1}{L}\int_{-L}^{L} f(t')\cos\frac{j\pi t'}{L}\,\mathrm{d}t' \quad b_j = \frac{1}{L}\int_{-L}^{L} f(t')\sin\frac{j\pi t'}{L}\,\mathrm{d}t'. \tag{7.97b}$$

Alternatively, we can express $f(t)$ as a sum of the complex exponential

$$f(t) = \sum_{k=1}^{n} c_k e^{ik\pi t/L} \quad i = \sqrt{-1} \tag{7.98}$$

with

$$c_k = \frac{1}{2L}\int_{-L}^{L} f(t')e^{-ik\pi t'/L}\,\mathrm{d}t'. \tag{7.99}$$

When the function $f(t)$ is periodic of angular frequency ω with period $2\pi/\omega$, then $L = \pi/\omega$.

Example 7.5

As an example of the Fourier representation of periodic functions, we consider the function $f(t)$ having the sawtooth form shown in Figure 7.24 and given by

$$f(t) = A\frac{t}{\tau} = \frac{\omega A}{2\pi}t \quad -\frac{\pi}{\omega} < t < \frac{\pi}{\omega}.$$

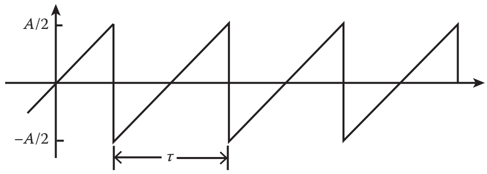


FIGURE 7.24 "Sawtooth" driving force.

Because $f(t)$ is an odd function, the coefficients a_n all vanish identically. The coefficients b_n are given by

$$b_n = \frac{\omega^2 A}{2\pi^2}\int_{-\pi/\omega}^{\pi/\omega} t\sin(n\omega t)\,\mathrm{d}t = \frac{\omega^2 A}{2\pi^2}\left[-\frac{t\cos n\omega t}{n\omega} + \frac{\sin n\omega t}{n^2\omega^2}\right]_{-\pi/\omega}^{\pi/\omega} = (-1)^{n+1}\frac{A}{n\pi}$$

where the term $(-1)^{n+1}$ accounts for the fact that

$$\cos n\pi = \begin{cases} -1 & n \text{ odd} \\ +1 & n \text{ even} \end{cases}.$$

The Fourier representation of $f(t)$ is given by

$$F(t) = \frac{A}{\pi}\left(\sin\omega t - \frac{1}{2}\sin 2\omega t + \frac{1}{3}\sin 3\omega t - \frac{1}{4}\sin 4\omega t + \ldots\right).$$

The solid curve in Figure 7.25 that shows the Fourier expansion does not converge toward the exact sawtooth function very rapidly. This is because a sawtooth function is a highly discontinuous function. For a function that is smoother than the sawtooth function, the first few terms of the Fourier expansion will usually give a fairly close approximation to the exact function. It is easy to see that the expansion overshoots the exact function in the region immediately adjacent to the points of discontinuity. This phenomenon, known as Gibb's overshoot or Gibb's phenomenon, amounts to approximately 8.9% on each side of the discontinuity in the limit of an infinite series.

We now return to our subject and arrive at the important conclusion. When the external driving force $F(t)$ in Equation 7.95 is a general periodic function, we first try to decompose it into a Fourier series, Equation 7.97a. Thus, each of the individual force functions $F_j(t)$ in Equation 7.93 has a simple harmonic dependence on time t, such as $\cos(j\omega t)$ or $\sin(j\omega t)$. Equation 7.94 is now of exactly the same form as Equation 7.67; the corresponding solution $x_j(t)$ is of the form given by Equation 7.74. In particular, if $F(t)$ is an even function with the period $2\pi/\omega$, then we can see from Equation 7.97b that the coefficients b_n all vanish identically, and $F(t)$ has the form

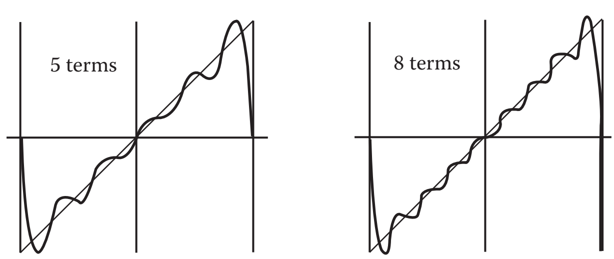


FIGURE 7.25 Fourier expansion of a sawtooth function.

$$F(t) = \sum_{j=1}^{n} a_j \cos(j\omega t - \alpha_j).$$

The corresponding steady-state solution is

$$x(t) = \frac{1}{m}\sum_{j=1}^{n} \frac{a_j}{\left[\left(\omega_0^2 - j^2\omega^2\right) + 4j^2\omega^2\beta^2\right]^{1/2}} \cos(j\omega t - \alpha_j - \delta_j)$$

where

$$\delta_j = \tan^{-1}\frac{2j\omega\beta}{\omega_0^2 - \omega^2}.$$

On the other hand, if $F(t)$ is an odd function, then the coefficients a_n all vanish identically, and $F(t)$ is thus represented by a series of terms, $\sin(j\omega t - \alpha_j)$. We can obviously write down a similar solution for this case.

It should be noted that the essence of the Fourier series solution to the forced oscillator is the decomposition of the arbitrary driving force into a spectrum of sinusoidal varying components. Thus, the method has the decided advantage of emphasizing the frequency response of the system. We now consider a simple illustrative example.

Example 7.6

A damped linear oscillator, originally at rest in its equilibrium, is subject to an external force consisting of a succession of rectangular pulses:

$$F_{ext}(t) = \begin{cases} F0 & NT - \Delta T/2 \le t \le NT + \Delta T/2 \\ 0 & \text{otherwise} \end{cases}$$

where $N = 0, \pm1, \pm2, \ldots$, T is the time from one pulse to the next, and ΔT is the width of each pulse as shown in Figure 7.26. Find the response function (i.e., the steady-state solution to the equation of motion of the system).

Solution:

$F_{ext}(t)$ is an even function of t, so it can be expressed as a Fourier cosine series. Equation 7.97b gives the coefficients

$$a_n = \frac{2}{T}\int_{-\Delta T/2}^{\Delta T/2} F_0 \cos(n\omega t)\mathrm{d}t = \frac{2}{T}F_0 \left.\frac{\sin(n\omega t)}{n\omega}\right|_{-\Delta T/2}^{\Delta T/2} = F_0\frac{2\sin(n\pi\Delta T/T)}{n\pi}$$

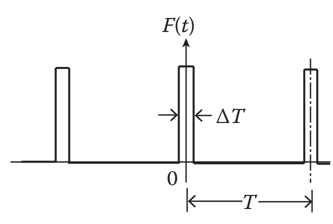


FIGURE 7.26 Rectangular-pulse driving force.

where, in the last step, we have used the fact that $\omega = 2\pi/T$. For $n = 0$ we see that

$$a_0 = F_0\,(2\Delta T/T).$$

Then Equation 7.97a gives our periodic pulse force:

$$F_{ext}(t)/F_0 = \frac{\Delta T}{T} + \frac{2}{\pi}\sin\left(\pi\frac{\Delta T}{T}\right) + \frac{2}{2\pi}\sin\left(2\pi\frac{\Delta T}{T}\right)\cos(2\omega t)$$
$$+\frac{2}{3\pi}\sin\left(3\pi\frac{\Delta T}{T}\right)\cos(3\omega t) + \ldots.$$

The first term is just the average value of the external force, and the second term is the Fourier component at the fundamental frequency. The remaining terms are harmonics of the fundamental: 2ω, 3ω, and so on.

Equation 7.97a gives the response function $x(t)$, which describes the motion of our pulse-driven damped oscillator:

$$x(t) = \sum_n x_n(t) = \sum_n A_n\cos(n\omega t - \delta_n)$$

in which the respective amplitudes A_n are given by

$$A_n = \frac{a_n/m}{D_n(\omega)} = \frac{(F_0/m)(2/n\pi)\sin(n\pi\Delta T/T)}{\sqrt{\left(\omega_0^2 - n^2\omega^2\right)^2 + 4\beta^2 n^2\omega^2}}$$

and the phase angles δ_n are given by

$$\delta_n = \tan^{-1}\left(\frac{2\beta\, n\omega}{\omega_0^2 - n^2\omega^2}\right).$$

Here, m is the mass, β is the damping constant, and ω_0 is the frequency of the free oscillation without damping.

7.9 VIBRATION ISOLATION

There are many instances where it is desirable or important to isolate an apparatus from the harsh effects of its vibration environment. For example, how can delicate instrumentation, such as an optical table, sit on a vibrating floor in such a way that the vibration is minimized? One arrangement to effect such isolation of vibration is shown in Figure 7.27. A heavy base is supported on the vibrating floor by a spring system with a spring constant k and viscous resistance b (represented by a dashpot). The floor is assumed to experience vertical vibration $Y = A\cos\omega t$ about its equilibrium; y is the corresponding vertical displacement of the base about its rest position. The function of the insulator is to keep the ratio y/A to a minimum. The equation of motion of the base is

$$-b(\dot{y} - \dot{Y}) - k(y - Y) = m\ddot{y} \qquad (7.100)$$

which can be rewritten as

$$m\ddot{y} + b\dot{y} + ky = b\dot{Y} + kY$$
$$= -bA\omega\sin\omega t + kA\cos\omega t.$$

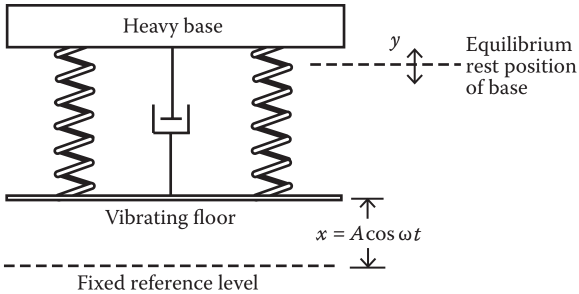


FIGURE 7.27 Vibration isolation.

Dividing by m, we obtain

$$\ddot{y} + 2\beta\dot{y} + \omega_0^2 y = \omega_0^2 A\cos\omega t - 2\beta A\omega\sin\omega t$$
$$= A\sqrt{\omega_0^4 + (2\beta\omega)^2}\,\cos(\omega t + \alpha) \quad (7.101)$$

where $\delta = \tan^{-1}\left(2\beta\omega/\omega_0^2\right)$. Because Equation 7.101 is of the same form as Equation 7.67 with $A\sqrt{\omega_0^4 + (2\beta\omega)^2}$ substituted for F_0/m, the solutions are identical in form. The amplitude of the steady-state response is, therefore,

$$y = \frac{A\sqrt{\omega_0^4 + (2\beta\omega)^2}}{\left(\omega_0^2 - \omega^2\right)^2 + (2\beta\omega)^2}$$

or

$$y/A = \frac{\sqrt{1 + (\omega/\omega_0)^2(2\beta/\omega_0)^2}}{\sqrt{\left(1 - (\omega/\omega_0)^2\right)^2 + (2\beta/\omega_0)^2(\omega/\omega_0)^2}}$$

where y/A is the ratio of the amplitude of the response to that of the input vibration. Thus, y/A serves as an indication of the effectiveness of isolation. Figure 7.28 is a plot of y/A as a function of ω/ω_0 for various values of $2\beta/\omega_0$. We seek values of $y/A < 1$. These occur at $\omega > \sqrt{2}\omega_0$ for all values of $2\beta/\omega_0$. The curves indicate that damping is detrimental in this region—damping has the effect of increasing the amplitude of vibration of the mass. Even though damping increases the steady-state

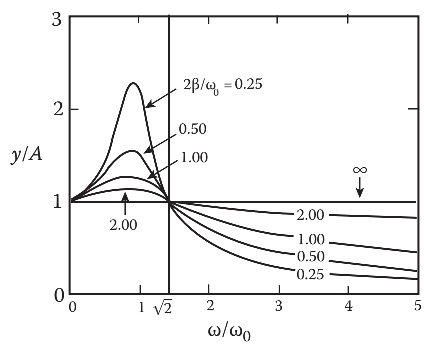


FIGURE 7.28 y/A as a function of ω/ω_0.

ratio y/A, it serves to limit the amplitude during the transient stage. Thus, $y/A > 1$, if $\omega > \sqrt{2}\omega_0$. Therefore, ω_0 should be as low as possible to protect against a given frequency.

7.10 PARAMETRIC EXCITATION

Large oscillatory motions can be increased not only by a periodic external force but also by periodically changing a parameter of the system in step with its oscillations, which is why the phenomenon is called parametric excitation. Parametric systems are very common in plasma and solid-state physics. Parametric devices also abound in electronics. For example, by varying the capacitance of a *LC* circuit, we can increase its current and voltage oscillations. Parametric amplifiers are used at radio and microwave frequencies where extremely low noise operation is required. And when a semiconductor junction functions as the voltage variable capacitance, it is placed in a circuit in a way that it can receive two voltages, one at the signal frequency ω_0 and the other at the pump frequency $2\omega_0$. A thousandfold gain in intensity of the input signal is typical for these instruments. More familiar still is the process of pumping a swing where the swinger rhythmically stands up and sits down (thereby periodically changing the position of the system's center of gravity). A simple analogue of this is a pendulum whose length can be varied by pulling and releasing the string on a pulley (Figure 7.29). Suppose we change the length l of the pendulum periodically, increasing it by a small amount a when the pendulum is at its extreme positions and decreasing it by the same amount a when the pendulum is at the equilibrium (vertical) position. Because the string is lengthened when the pendulum is deflected, it will descend a distance ($a \cos \theta$), where θ is the angular amplitude of the oscillations of the pendulum. Each time the string is raised and lowered, the external force acting on the string does an amount of work $Mga(1 - \cos \theta)$ against gravity. The external force also does work against the centrifugal force that tightens the string. This work is zero at the extreme positions, and it is Mv_0^2a/l at the lowest positions, where v_0 is the velocity of the pendulum at its lowest position. Thus, the total work done by the external force in one period of oscillation is

$$W = 2\left[Mga(1-\cos\theta) + \frac{Mv_0^2 a}{l}\right]$$

where the factor 2 is a result of the fact that, during each period, the pendulum is lengthened and shortened twice. Because θ is assumed to be small, $\cos \theta \cong 1 - \theta^2/2$, and $v_0 = l \sin \theta\, \omega \cong \theta\omega l$, where $\omega = (g/l)^{1/2}$ is the frequency of the pendulum's oscillations. With these substitutions, W becomes

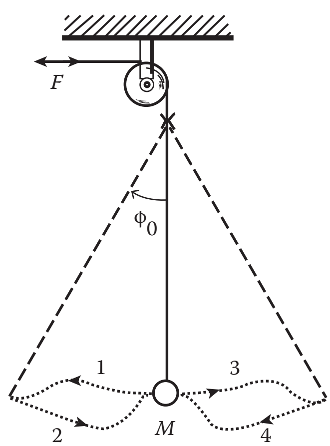


FIGURE 7.29 Simple parametric pendulum.

$$W = 2\left[\frac{Mga\theta^2}{2} + \frac{Mv_0^2 a}{l}\right] = 6(a/l)\frac{Mv_0^2}{l}.$$

Thus, the work done on the pendulum by the external force is positive and is proportional to the pendulum's energy. The energy of the pendulum will, therefore, increase steadily by a small amount in each period proportional to the energy and to the quantity (a/l). This is the mechanism of parametric resonance. Because the rate of increase of the mean energy is proportional to the energy, it can be written as $dE/dt = 2\alpha E$, where α is a small constant that can be called the amplification coefficient. Thus, if the amplification coefficient α exceeds the damping coefficient, the energy and, hence, the amplitude of the oscillations increase exponentially with time. Probably the most important aspect of a parametric oscillator, therefore, is that it becomes unstable at certain resonant frequencies.

To examine the general features of a parametric oscillator and its instability, we need to know its equation of motion. A parametric oscillator may be thought of as a harmonic oscillator whose parameters have a time-dependent piece, and its motion obeys an equation of the form known as Hill's equation:

$$\ddot{x} + \left[\omega_0^2 + k(t)\right]x = 0 \tag{7.102}$$

where $k(t)$ is a real, periodic function of t. If $k(t)$, when expanded harmonically in a Fourier series, contains a frequency close to ω_0, the equation looks somewhat like that of a harmonic oscillator driven by a force that has a component close to the natural frequency of the oscillator. The driving term causes x to grow, and the increase in x leads in turn to increased driving, which may cause the system to become unstable.

In many physical systems, the function $k(t)$ is harmonic to some approximation and has the form $\sim\cos \omega t$. The equation of motion now can be put into the form

$$\ddot{x} + \left(\omega_0^2 + h\cos\omega t\right)x = 0 \tag{7.103}$$

which is known as Mathieu's equation, where $h \ll 1$, and we shall suppose h to be positive as can always be done by suitably choosing the origin of time. As we shall see, parametric resonance is strongest if the frequency ω is nearly twice ω_0. Hence, we write $\omega = 2\omega_0 + \varepsilon$, where $\varepsilon \ll 1$. Equation 7.103 then becomes

$$\ddot{x} + \omega_0^2[1 + h\cos(2\omega_0 + \varepsilon)\, t]x = 0. \tag{7.104}$$

In this approximation, its solution can be written in the form

$$x(t) = a(t)\cos(\omega_0 + \varepsilon/2)t + b(t)\sin(\omega_0 + \varepsilon/2)t. \tag{7.105}$$

Substituting this into Equation 7.104 and retaining only first-order terms in ε, and assuming that $\dot{a} \sim \varepsilon a$ and $\dot{b} \sim \varepsilon b$, we find

$$\begin{aligned} &-\left(2\dot{a} + b\varepsilon + \frac{1}{2}hb\omega_0\right)\omega_0\sin(\omega_0 + \varepsilon/2)t \\ &+\left(2\dot{b} - a\varepsilon + \frac{1}{2}ha\omega_0\right)\omega_0\cos(\omega_0 + \varepsilon/2)t = 0. \end{aligned} \tag{7.106}$$

In obtaining Equation 7.106, we replaced the products of trigonometric functions by sums:

$$\cos\left(\omega_0+\frac{\varepsilon}{2}\right)t\cos\left(2\omega_0+\frac{\varepsilon}{2}\right)t=\frac{1}{2}\cos 3\left(\omega_0+\frac{\varepsilon}{2}\right)t+\frac{1}{2}\cos\left(\omega_0+\frac{\varepsilon}{2}\right)t$$

and omitted terms with frequency $3(\omega_0 + \varepsilon/2)$.

If Equation 7.106 is true, the coefficients of the sine and cosine must both be zero. This gives two linear differential equations for the functions $a(t)$ and $b(t)$:

$$2\dot{a}+b\varepsilon+\frac{1}{2}hb\omega_0=0,\quad 2\dot{b}-a\varepsilon+\frac{1}{2}ha\omega_0=0. \tag{7.107}$$

If

$$a \sim e^{\alpha t}$$

and

$$b \sim e^{\alpha t}$$

then the two preceding differential equations for $a(t)$ and $b(t)$ become

$$2\alpha\, a + b(\varepsilon + h\omega_0/2) = 0$$

or

$$\frac{a}{b}=-\frac{(\varepsilon+h\omega_0/2)}{2\alpha}$$

$$2\alpha\, b - b(\varepsilon - h\omega_0/2) = 0$$

or

$$\frac{a}{b}=-\frac{2\alpha}{(\varepsilon-h\omega_0/2)}.$$

From these equations, we find

$$\alpha^2=\frac{1}{4}\left[\left(\frac{1}{2}h\omega_0\right)^2-\varepsilon^2\right].$$

The condition for parametric resonance is that α is real, so $\alpha^2 > 0$, that is,

$$\alpha^2=\frac{1}{4}\left[\left(\frac{1}{2}h\omega_0\right)^2-\varepsilon^2\right]>0. \tag{7.108}$$

Thus, the parametric resonance occurs in the range

$$-\frac{1}{2}h\omega_0<\varepsilon<\frac{1}{2}h\omega_0 \tag{7.109}$$

on either side of the frequency $2\omega_0$.

PROBLEMS

1. Find the linear harmonic oscillator approximation for the potential energy given by

$$V(x) = ax^{-2} + bx^2, \quad a,b > 0.$$

2. A hydrometer floats in a liquid of density d. It is slightly depressed from its normal position of equilibrium and periodic motion about the equilibrium results. Show that the period of oscillations is given by

$$\tau = 2\pi\sqrt{\frac{W}{gAd}}$$

where g is the local gravitational acceleration, W is the weight of the hydrometer, and A is the cross-sectional area (Figure 7.30).

3. A straight, uniform stick having a length $2b$ (meter) and a mass M (kg) is freely pivoted at one end. Find the frequency of oscillation about the pivot, assuming the angle θ is small.
4. A frictionless vertical cylinder of cross-sectional area A contains a gas, which is trapped by a piston that fits the cylinder perfectly. The piston has a mass m, and the atmospheric pressure above the piston is p_0 and is assumed to be constant. The piston is slightly displaced and, when released, oscillates about its equilibrium position. Show that the oscillations are approximately simple harmonic.
5. A particle of mass m is placed on a frictionless horizontal table and attached by two identical massless spiral springs of natural length b and stiffness k to two fixed points A and B on the table. A and B are $2a$ apart, where $2a > 2b$, so that the two springs are stretched. The particle is then given a displacement along the line of the springs and released. Show that the motion that ensues is simple harmonic and find its period.
6. A particle of mass m is suspended from a point 0 by means of a massless elastic string of natural length b and modulus λ, and the system is set in motion by giving the particle a small vertical displacement. Find the equation of motion of the particle and the natural frequency of oscillation.
7. A mass m is attached to a massless rigid rod, and the rod is connected by means of a linear spring k to a wall. Find the equation of motion and the natural frequency of oscillations of the system (Figure 7.31).
8. A particle of mass m moves under the influence of the potential energy $V(x) = (1 - \alpha x)\, e^{-\alpha x}$ in the region $x \geq 0$, and $\alpha > 0$. Determine (a) the location of the equilibrium points, (b) the nature of the equilibrium, and (c) the natural frequency of oscillations for motion restricted to small displacement in the neighborhood of the equilibrium. Show that for a large displacement of approximately $x = 2/\alpha$, the motion is anharmonic.
9. An electric dipole, a pair of oppositely charged particles attached to a massless rod of length b, is placed in a uniform electric field E. Find the natural frequency of oscillation

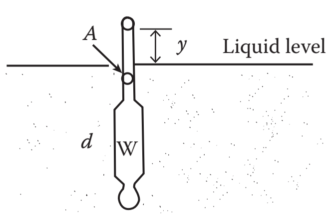


FIGURE 7.30 Hydrometer floats in a liquid.

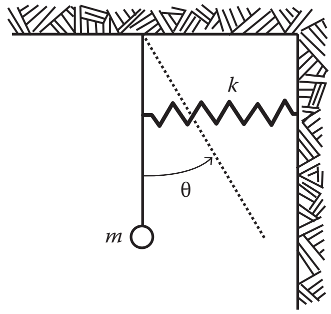


FIGURE 7.31 Mass–spring system.

if the dipole, shown in Figure 7.32, is slightly rotated from its equilibrium position and released. If a damping torque proportional to angular velocity of the dipole is present, determine the frequency of oscillations.

10. An electron in an atom that is freely radiating power behaves as a damped simple harmonic oscillator. If the radiated power is given by $\langle P \rangle = q^2\omega^4 x_0^2/12\pi\varepsilon_0 c^3$ watts at a wavelength of 0.6 μm (600 Å), show that the Q-value of the atom is approximately 10^{-8} and its free radiation lifetime is approximately 10^{-8} s (the time for its energy to decay to e of its original value).
11. A block of mass m is attached to a spring of stiffness k and moves on a horizontal surface with a coefficient of friction μ. If the block is initially at a displacement x from the equilibrium position and released from rest, find the position where the block finally comes to rest. Assume that the coefficients of static friction and sliding friction are equal.
12. An oscillator is subject to an external driving force $F(t)$, which is an exponentially decreasing function of time t:

$$F(t) = \begin{cases} F_0 e^{-\alpha t}, & t \geq 0 \\ 0, & t < 0 \end{cases}.$$

Find the position of the oscillator at any time t.

13. A bead of mass is restricted to slide on a frictionless wire in the shape of a cycloid (Figure 7.33) whose parametric equations are

$$x = a\,(\phi - \sin\phi) \quad y = a\,(1 - \cos\phi)$$

which lie in a vertical plane. The bead is released from reset at point 0.

(a) Find its speed at the bottom of the path.

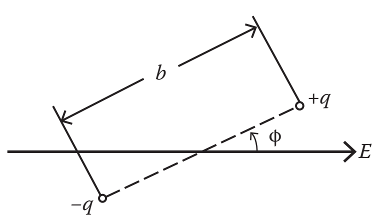


FIGURE 7.32 Electric dipole.

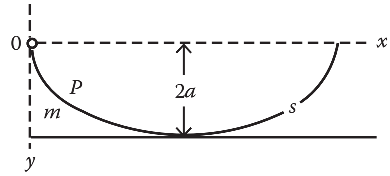


FIGURE 7.33 A bead slides on a cycloid wire.

14. Two masses, m and M, are connected by a spring whose force constant is k and slide freely on a horizontal frictionless surface. If the masses have an initial separation not equal to the natural length b, find the frequency of oscillatory motion for this system.
15. A particle of mass m is suspended from the end of a vertical spring whose force constant is k. A constant downward force F is applied continuously to the mass for a time t. Show that after the force is removed, the displacement of the mass from its equilibrium position x is

$$x = F \cos \omega(t - t_0) - \cos \omega t, \quad \omega^2 = k/m.$$

16. A damped linear harmonic oscillator is subjected simultaneously to two driving forces $A_1 \cos \omega_1 t$ and $A_2 \cos(\omega_2 t + \phi)$. Show that the resultant displacement in the steady state is the sum of the displacements resulting from the driving forces acting separately. Show also that the rate at which work is being done by the driving forces is the sum of the rates obtained if the forces were acting separately.
17. Suppose that there is a negative frictional force proportional to the velocity on a harmonic oscillator so that the equation of motion is

$$m\ddot{x} - b\dot{x} + kx = 0.$$

    Without actually solving the equation of motion, discuss what the qualitative nature of the motion will be.
18. A mass m is connected to a wall by means of a linear spring whose force constant is k. The mass is at rest on a frictionless surface, and the spring is undeformed. The wall suddenly receives a velocity shock and instantaneously gains a constant velocity u to the right. Determine the response of the system and the spring force as a function of time (Figure 7.34).

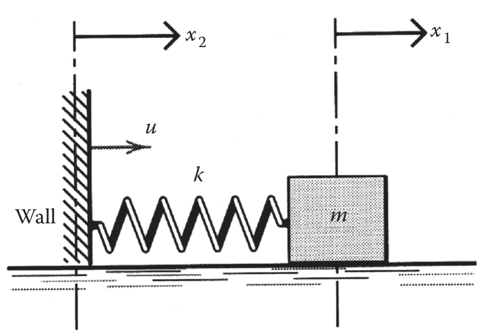


FIGURE 7.34 Mass–spring system connected to a moving wall.

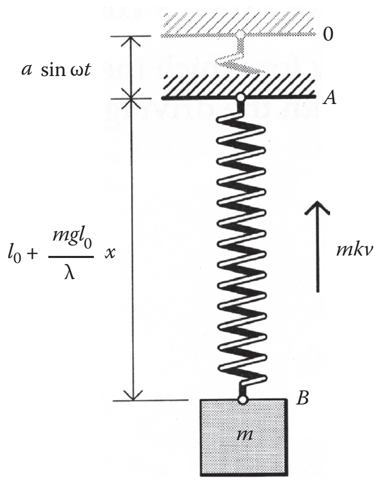


FIGURE 7.35 Vertical elastic spring.

19. A particle of mass m is attached to the lower end, B, of a vertical elastic spring AB whose modulus is λ and natural length is b. The upper end, A, is forced to undergo a vertical oscillation, $a \sin \omega t$. The mass m is subject to a resistance equal in magnitude to mk times its speed. If the spring's length at time t is $b + mgb/\lambda + x$, show that x satisfies

$$\ddot{x} + k\dot{x} + (\lambda/mb)x = a\omega\left[\omega \sin(\omega t) - k\cos(\omega t)\right].$$

Find the amplitude of the forced oscillation of the mass when $k = \omega$ and $\lambda/mb = \omega^2$, and find the phase angle of this oscillation relative to A's position (Figure 7.35).

20. A block of mass m is attached to a fixed wall by means of a spring whose force constant is k. The block slides on an inclined plane of angle θ. The kinetic friction coefficient between the block and the inclined surface is μ. Show that (a) the period of oscillations is equal to that of a simple harmonic oscillator, and (b) the amplitude is decreased by the same amount each cycle.
21. Use the Fourier series method to obtain the response of a damped oscillator subject to a periodic force that has a rectangular shape of amplitude unity and period 4 (Figure 7.36).
22. An electric circuit contains a resistance R, capacitance C, and inductance L, where R, C, and L are constants, and $4L > CR$. Find the current I when there is zero initial current and charge and when the impressed voltage is given as $V(t) = V_0 t$ for $t > 0$ (Figure 7.37).
23. Describe the motion with friction of an oscillator that is initially at rest and that is acted upon by a force $F(t) = F\cos \omega t$.

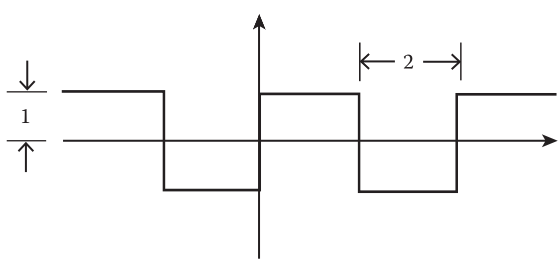


FIGURE 7.36 Rectangular shaped periodic force.

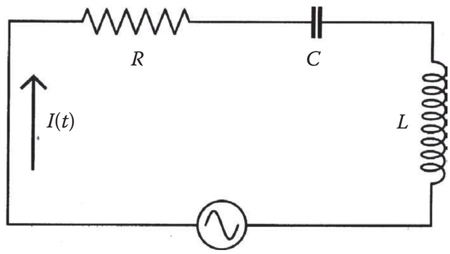


FIGURE 7.37 RLC electric circuit.

24. Find the driven frequency for which the velocity of a forced damped oscillator is exactly in phase with the driving force.
25. Given the Lagrangian L of the damped oscillation

$$L = \frac{1}{2}e^{bt/m}m\dot{x}^2 - \frac{1}{2}kx^2,$$

show that the Lagrange's equation of motion is given by Equation 7.29.
Show that Equation 7.29 can also be obtained from the modified Lagrange equation

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L}{\partial \dot{x}}\right) - \frac{\partial L}{\partial x} + \frac{\partial R}{\partial \dot{x}} = 0$$

where

$$L = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2$$

and

$$R = \frac{1}{2}k\dot{x}^2.$$

8 Coupled Oscillations and Normal Coordinates

We have learned in some detail how a single vibrating system will behave. However, oscillators rarely exist in complete isolation. We now consider systems of interacting oscillators. This subject is of general interest because numerous physical systems are well approximated by coupled harmonic oscillators. Coupled oscillators, in general, are able to transmit their energy to each other because two oscillators share a common component, capacitance or stiffness, inductance or mass, or resistance. Resistance coupling inevitably brings energy loss and a rapid decay in the vibration, but nonresistance coupling consumes no power, and continuous energy transfer over many oscillators is possible.

We shall investigate first a mechanical example of stiffness coupling between two pendulums. Two atoms set in a crystal lattice experience a mutual coupling force and would be amenable to a similar treatment. Motion of this type can be quite complex if it is described in ordinary coordinates that describe the geometrical configuration of the system. Fortunately, as we shall see, it is always possible to describe the motion of any oscillatory system in terms of normal coordinates that are constructed from the original position coordinates in such a way that there is no coupling among the oscillators. Thus, each normal coordinate oscillates with a single, well-defined frequency. Before we take up the general analytic approach, let us illustrate the concepts of normal coordinates and normal frequencies with a very simple example: the coupled pendulum.

8.1 COUPLED PENDULUM

Consider a pair of identical pendulums of mass m suspended on a massless rigid rod of length b. The masses are connected by a massless spring whose spring constant is k and whose natural length equals the distance between the masses when either is displaced from equilibrium. The displacements from rest of the two pendulum bobs are measured in x and y, respectively, as indicated in Figure 8.1. The kinetic and potential energies for the coupled system are given by, for small displacement,

$$T = \frac{1}{2}m(\dot{x}^2 + \dot{y}^2) \tag{8.1a}$$

$$V = \frac{1}{2}K(x^2 + y^2) + \frac{1}{2}k(x - y)^2 \tag{8.1b}$$

where $K = mg/b$. It is clear that in this small displacement approximation the coupled pendulum system of Figure 8.1 is equivalent to the mass-spring system of Figure 8.2, with a spring constant K for the outer spring and k for the inner spring. Now, from the Lagrangian $L = T - V$ and Lagrange's equations $\mathrm{d}(\partial L/\partial x)/\mathrm{d}t - \partial L/\partial x = 0$, $\mathrm{d}(\partial L/\partial y)/\mathrm{d}t - \partial L/\partial y = 0$, we obtain the differential equations of motion:

$$m\ddot{x} = -Kx - k(x - y) \tag{8.2a}$$

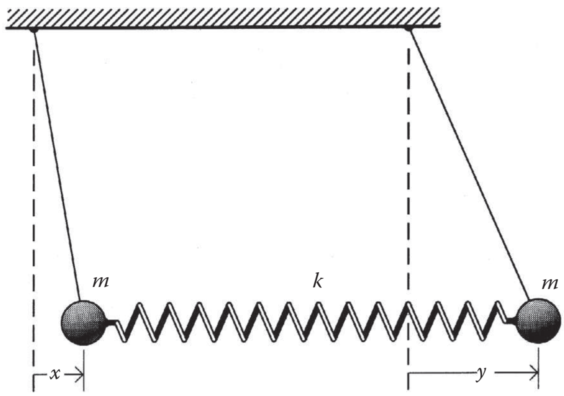


FIGURE 8.1 Two simple pendulums coupled by a spring.

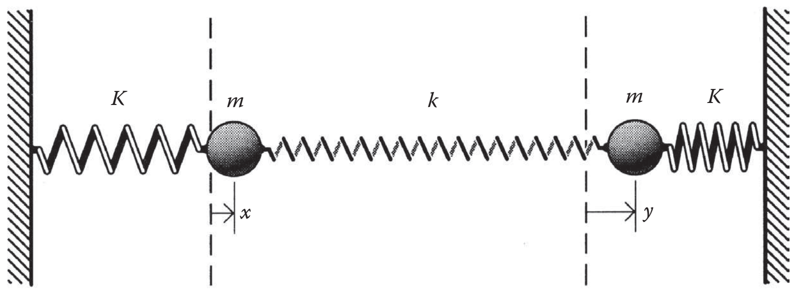


FIGURE 8.2 Equivalent mass–spring system for the coupled pendulum system.

$$m\ddot{y} = -Ky + k(x - y). \tag{8.2b}$$

It is obvious that Equation 8.2a contains a term in y, and Equation 8.2b contains a term in x. These two differential equations cannot be solved independently and must be solved simultaneously. A motion given to one bob affects the other. As a simple approach to the solution of these simultaneous differential equations, we try an oscillatory state in which both pendulums oscillate with the same frequency ω:

$$x(t) = Ae^{i\omega t} \text{ and } y(t) = Be^{i\omega t} \tag{8.3}$$

where A and B are constants. Because damping is not present, ω will be a real quantity. Substitution of Equation 8.3 into Equations 8.2a and 8.2b gives a pair of simultaneous linear algebraic equations in the undetermined amplitude A and B:

$$(-m\omega^2 + K + k)A - kB = 0 \tag{8.4a}$$

$$kA + (m\omega^2 - K - k)B = 0. \tag{8.4b}$$

Such is a system of two linear homogeneous equations whose solution is not zero only if the determinant of its coefficients vanishes

$$\begin{vmatrix} -m\omega^2 + K + k & -k \\ k & m\omega^2 - K - k \end{vmatrix} = 0. \tag{8.5}$$

Upon expanding this secular determinant, we have

$$(-m\omega^2 + K + k)^2 - k^2 = 0, \tag{8.6}$$

an equation for the determination of ω, which may be rewritten as

$$(m\omega^2 - K)(m\omega^2 - K - 2k) = 0$$

from which we obtain the characteristic frequencies (often called eigenfrequencies) for the system, either

$$\omega^2 = K/m = g/b; \quad \omega_{1,2} = \pm\sqrt{g/b} \tag{8.7a}$$

or

$$\omega^2 = g/b + 2k/m; \quad \omega_{3,4} = \pm\sqrt{g/b + 2k/m}. \tag{8.7b}$$

One of the characteristic frequencies is of the free pendulum $(g/b)^{1/2}$ exactly; the two pendulums move in phase. The other frequency reactivates both the pendulums and the springs (with a factor of 2); in this state, the pendulums move in opposite directions. The most general motion of the system is a superposition of these two modes of oscillation:

$$x(t) = A_1 e^{i\omega_1 t} + A_2 e^{i\omega_2 t} + A_3 e^{i\omega_3 t} + A_4 e^{i\omega_4 t}$$

$$y(t) = B_1 e^{i\omega_1 t} + B_2 e^{i\omega_2 t} + B_3 e^{i\omega_3 t} + B_4 e^{i\omega_4 t}$$

where, as usual, only the real or only the imaginary part is to be taken. Only four of the eight arbitrary constants present in the preceding equations are independent. This follows because the differential equations are of the second order and are two in number. The ratio between A and B can be determined by inserting the allowed value of ω into Equation 8.4a or 8.4b:

$$\text{at } \omega = \omega_{1,2} \qquad A_{1,2} = +B_{1,2}$$
$$\text{at } \omega = \omega_{3,4} \qquad A_{3,4} = -B_{3,4}$$

The complete solutions then become

$$x(t) = A_1 e^{i\sqrt{g/b}t} + A_2 e^{-i\sqrt{g/b}t} + A_3 e^{i\sqrt{g/b+2k/m}t} + A_4 e^{-i\sqrt{g/b+2k/m}t} \tag{8.8a}$$

$$y(t) = A_1 e^{i\sqrt{g/b}t} + A_2 e^{-i\sqrt{g/b}t} - A_3 e^{i\sqrt{g/b+2k/m}t} - A_4 e^{-i\sqrt{g/b+2k/m}t}. \tag{8.8b}$$

We now have only four arbitrary constants, A_1, A_2, A_3, and A_4, present in the general solution of the two second-order differential equations.

8.1.1 Normal Coordinates

It is very clear, by inspection of Equations 8.8a and 8.8b or Equations 8.4a and 8.4b that it is possible to make linear combinations of x and y such that a combination involves a single frequency. These linear combinations are, in our case, merely the sum and difference of x and y:

$$X(t) = x(t) + y(t) = 2\left[A_1 e^{i\sqrt{g/bt}} + A_2 e^{-i\sqrt{g/bt}}\right] \quad (8.9a)$$

$$Y(t) = x(t) - y(t) = 2\left[A_3 e^{i\sqrt{g/b+2k/mt}} + A_4 e^{-i\sqrt{g/b+2k/mt}}\right] \quad (8.9b)$$

where X describes one mode of oscillation with $\omega_0 = (g/b)^{1/2}$, while Y describes the other with $\omega = (g/b + 2k/m)^{1/2}$. The terms X and Y are called normal coordinates, and the two modes of oscillation, $(g/b)^{1/2}$ and $(g/b + 2k/m)^{1/2}$, are the corresponding normal modes. We see that a normal mode vibration involves only one dependent variable X (or Y) and has its own normal frequency.

Coordinates X and Y have no immediate geometric meaning. In order to learn something about the geometrical configuration of the system, we would have to transform back from (X, Y) to (x, y):

$$x = (X + Y)/2 \text{ and } y = (X - Y)/2. \quad (8.10)$$

However, the normal coordinates X and Y have some interesting and useful properties from the physical viewpoint:

(1) The equations of motion, when expressed in terms of normal coordinates, are linear equations with constant coefficients, and each contains but one dependent variable. That is, each of which has an equation of motion that is simple harmonic:

$$\ddot{X} = -(g/b)X, \quad \ddot{Y} = -(g/b + 2k/m)Y. \quad (8.11)$$

Thus, the normal coordinates are independent of each other in the sense that each normal oscillation may be excited while the other is at rest. If the boundary conditions are such that only one normal coordinate is excited and the other is zero initially, the latter will remain zero for all time.

(2) The total energy of the system, which in ordinary position coordinates is a mixed function of both x and y and their time derivatives, assumes in normal coordinates the following form:

$$E_{tot} = \frac{m\left(\dot{X}^2 + \dot{Y}^2\right)}{4} + \frac{1}{4}\left[\left(\frac{mg}{b}\right)X^2 + \left(\frac{mg}{b} + k\right)Y^2\right]. \quad (8.12)$$

The potential energy is now expressed as a sum of squares of the normal coordinates, multiplied by constant coefficients. There are no longer any cross terms present in E_{tot}. Thus, when one normal coordinate is excited while the other remains zero, there is no tendency for energy to pass from one normal coordinate to another, and the coordinate that is zero initially remains zero for all time.

From the preceding discussion, it is evident that transformation from ordinary position coordinates to normal coordinates leads to replacement of the actual system of interacting oscillators by a system of as many independent oscillators. It is very clear, by inspection of Equations 8.8a and 8.8b, that the motion of the pendulums is, in general, a superposition of the two normal modes of oscillation. To excite a single frequency (i.e., one normal mode), either A_1 and A_2 or A_3 and A_4 must be zero. These two possibilities correspond to the pendulum bobs swinging synchronously (in phase) with frequency $(g/b)^{1/2}$ or swinging anti-synchronously (out of phase) with frequency $(g/b + 2k/m)^{1/2}$ as depicted in Figure 8.3. Notice that the synchronous mode has the lower frequency and is a general result. In a complex system of linearly coupled oscillators, the mode having the highest degree of symmetry has the lowest frequency.

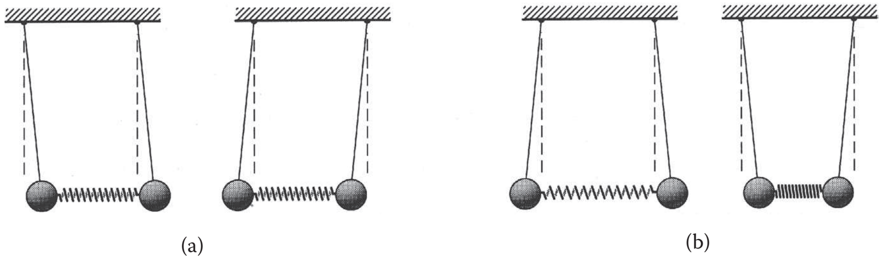


FIGURE 8.3 (a) The two pendulum swing in phase and (b) the two pendulum swing out of phase.

The general motion of the system is a linear combination of the synchronous and anti-synchronous modes (see Equations 8.8a and 8.8b). At the moment of excitation, $t = 0$, if $x = \dot{x} = 0$, $y = C$, and $\dot{y} = 0$ (i.e., we displace pendulum 2 a distance C and release it from rest). Equations 8.8a and 8.8b then give

$$A_1 = A_2 = C/4 \text{ and } A_3 = A_4 = C/4.$$

Substituting these results into Equations 8.8a and 8.8b, we find

$$x(t) = \frac{1}{2}C(\cos\omega_1 t - \cos\omega_3 t) \text{ and } y(t) = \frac{1}{2}C(\cos\omega_1 t + \cos\omega_3 t)$$

where ω_1 and ω_2 are given in Equations 8.7a and 8.7b. These equations may be put in the form

$$x(t) = C\sin\left[\frac{1}{2}(\omega_3 - \omega_1)t\right]\sin\left[\frac{1}{2}(\omega_3 + \omega_1)t\right] \tag{8.13a}$$

$$y(t) = C\cos\left[\frac{1}{2}(\omega_3 - \omega_1)t\right]\cos\left[\frac{1}{2}(\omega_3 + \omega_1)t\right]. \tag{8.13b}$$

Now, in the case of weak coupling,

$$\begin{aligned}\omega_3 - \omega_1 &= \sqrt{\omega_1^2 + 2k/m} - \omega_1 \\ &= \omega_1\sqrt{1 + (2k/m)\omega_1} - \omega_1 \\ &= \frac{k}{m}\omega_1 \ll 1.\end{aligned}$$

Thus, the first factors of the right members of Equations 8.13a and 8.13b vary slowly with time. These slowly varying factors constitute an envelope for the rapidly oscillating sinusoidal factors of argument as depicted in Figure 8.4. As the amplitude of $x(t)$ becomes large, that of y becomes smaller, and vice versa. Thus, there is a transfer of energy back and forth. At $t = \pi/(\omega_3 - \omega_1)$, pendulum 2 has come to rest, and all the energy has been transferred through the coupling to the first pendulum. It is this circumstance that gives rise to the phenomenon known as beats. The beat frequency is $(\omega_3 - \omega_1)/2$, and the frequency of the envelope of the amplitude is $(\omega_3 - \omega_1)$.

The preceding procedure can be applied to a system of oscillators with different masses; also, there is no limit to the number of oscillators constituting the system. For a system of n coupled oscillators, we will find n normal coordinates, $Q_1, Q_2, \ldots, Q_n$, each representing one normal mode of

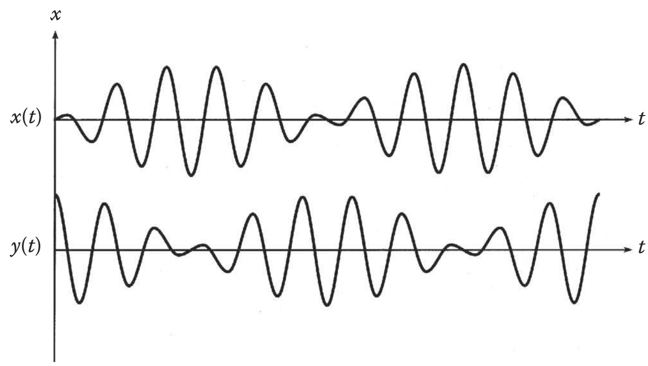


FIGURE 8.4 Oscillograph of two weakly coupled pendulums.

oscillation or one independent oscillator. We examine this general problem of n coupled oscillators in the next section.

8.2 COUPLED OSCILLATORS AND NORMAL MODES: GENERAL ANALYTIC APPROACH

For an oscillatory system with n degrees of freedom, the maximum number of normal frequencies is n. When some of the normal frequencies are identical, the system is said to be degenerate. We shall limit our discussion to the nondegenerate case.

In the simple example given earlier, the normal coordinates were introduced from symmetry consideration. Suppose it was not easy to discover the normal coordinates from symmetry consideration. How then could we plow through to a solution? Here we give a general analytic approach to this problem.

8.2.1 The Equation of Motion of a Coupled System

Consider a conservative system that has n degrees of freedom and a position of stable equilibrium where the potential energy of the system assumes a minimum value. We measure the generalized coordinates $q_1, q_2, \ldots, q_n$ of the system from this equilibrium position. If the motion of the system is of small amplitude and takes place about the point of equilibrium, then the potential energy V of the system can be expanded into a Taylor's series about the equilibrium configuration:

$$V(q_1, q_2, \ldots, q_n) = V_0 + \sum_i \left(\frac{\partial V}{\partial q_i}\right)_0 q_i + \frac{1}{2}\sum_i \sum_j \left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0 q_i q_j + \ldots. \qquad (8.14)$$

Terms of higher orders of approximation have been dropped. The second sum vanishes because all the generalized forces, $(\partial V/\partial q_i)_0$, vanish at the equilibrium configuration. Also, no generality will be lost by taking $V_0 = 0$. Equation 8.14 now reduces to the quadratic form

$$V = \frac{1}{2}\sum_i \sum_j K_{ij} q_i q_j \qquad (8.15)$$

where

$$K_{ij} = \left(\frac{\partial^2 V}{\partial q_i \partial q_j}\right)_0. \qquad (8.16)$$

Because V is measured from its minimum value, and this minimum is taken to be zero, we must have, in general, $V > 0$. That is, Equation 8.15 is a positive-definite quadratic form. The K_{ij} are constants and symmetrical ($K_{ij} = K_{ji}$) because the second derivatives are evaluated at the equilibrium position and the order of differentiation is immaterial under our assumption. The diagonal components of K_{ij} represent the force constants of the restoring force acting on a single particle when that particle alone is displaced.

If the constraints are time independent, kinetic energy T can be written as a homogeneous quadratic form in the velocities

$$T = \frac{1}{2}\sum_i \sum_j m_{ij}\dot{q}_i\dot{q}_j \tag{8.17}$$

where the m_{ij} are, in general, functions of the generalized coordinates and contain the masses, or other inertial parameters, of the system. We may expand the quantity m_{ij} into a Taylor's series about the equilibrium values of the q_i in a manner similar to that shown in Equation 8.14. As the q_i are assumed to be small, we shall take the constant values of m_{ij} at the equilibrium position as an approximation and neglect all the higher-order terms in the expansion. Call these constants A_{ij}; then Equation 8.17 becomes

$$T = \frac{1}{2}\sum_i \sum_j A_{ij}\dot{q}_i\dot{q}_j \tag{8.18}$$

where $A_{ij} = A_{ji}$. It is clear that T is a positive-definite quadratic form. The Lagrangian of the system is

$$L = T - V = \frac{1}{2}\sum_i \sum_j (A_{ij}\dot{q}_i\dot{q}_j - K_{ij}q_iq_j). \tag{8.19}$$

The equations of motion follow from the Lagrange's equation $\frac{d}{dt}\frac{\partial L}{\partial q_i} - \frac{\partial L}{\partial q_i} = 0$

$$\sum_i (A_{jk}\ddot{q}_k + K_{jk}q_k) = 0,\ j = 1,2,\ldots,n \tag{8.20}$$

where the terms $A_{jk}\ddot{q}_k$ and $K_{jk}q_k$ are called dynamical and static coupling terms, respectively.

Equation 8.20 is a system of n second-order linear homogeneous differential equations with constant coefficients, and each of these equations involves all n coordinates. So it is desirable to separate the variables and obtain n equations, each only involving a single unknown. This means that we introduce normal coordinates.

8.2.2 Normal Modes of Oscillation

Our first step is to find a set of constants C_j. If we multiply the first of Equation 8.20 by C_1, the second by C_2, and so on, and then add the resulting equations, we obtain a new set of equations of the form

$$\ddot{\eta}_j + \omega^2\eta_j = 0,\ j = 1,2,\ldots,n \tag{8.21}$$

where η_j are linear combinations of q_j:

$$\eta_j = \sum_k h_{jk} q_k. \tag{8.22}$$

The various constants are related by the sets of equations

$$\sum_j C_j A_{jk} = \omega^2 \sum_j C_j K_{jk} = h_{jk}. \tag{8.23}$$

From the last equation, we have

$$\sum_k (\omega^2 A_{jk} - K_{jk}) C_k = 0,\ j = 1,2,\ldots,n \tag{8.24}$$

which is a system of n linear homogeneous algebraic equations with the unknowns C_1, C_2,…, C_n. They have a trivial solution $C_1 = C_2 = \ldots = C_n$ and a nontrivial solution given by, as before, setting the determinant of its coefficients equal to zero:

$$\begin{bmatrix} K_{11} - \omega^2 A_{11} & K_{12} - \omega^2 A_{12} & . & . & . & K_{1n} - \omega^2 A_{1n} \\ K_{21} - \omega^2 A_{21} & K_{22} - \omega^2 A_{22} & . & . & . & K_{2n} - \omega^2 A_{2n} \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & . & . \\ K_{n1} - \omega^2 A_{n1} & K_{n2} - \omega^2 A_{n2} & . & . & . & K_{nn} - \omega^2 A_{nn} \end{bmatrix} = 0. \tag{8.25}$$

It is evident that if we assume that the solution of the equation of motion, Equation 8.20, is (compared with Equation 8.3)

$$q_j(t) = C_j e^{i(\omega t - \delta)}$$

then, by substitution of q_j into Equation 8.20, the resulting equations will take the same form as Equation 8.24.

Now the determinant of Equation 8.25 is an nth-degree equation in ω that can be solved for the n roots: ω_1^2, ω_2^2,…, ω_n^2. It can be proved that these roots are all real. To accomplish this, we multiply through the jth equation of Equation 8.24 by C_j and sum over j, which gives

$$\omega^2 \sum_j \sum_k A_{jk} C_j C_k - \sum_j \sum_k K_{jk} C_j C_k = 0. \tag{8.26}$$

We notice that the double sum in the first term is the function $2T$ of Equation 8.18 with $\dot{q}_j$ and $\dot{q}_k$ replaced by C_j and C_k. The double sum in the second term is the function $2U$ of Equation 8.15, with C_j and C_k replacing q_1 and q_k. And Equation 8.26 may thus be written in the form

$$\omega^2 T(\mathrm{C}) - V(C) = 0$$

or

$$\omega^2 = T(\mathrm{C})/V(C). \tag{8.27}$$

Here, T and V are positive for all sets of values of the variables, and accordingly, ω^2 is real and positive.

For the nondegenerate case, Equation 8.25 has n different real positive roots ω_j^2, $j = 1, 2,..., n$. This means that there are n different modes of oscillation with frequencies given by ω_j. ω_j are the natural frequencies of the system and are called the characteristic frequencies or eigenfrequencies of the system. Each value of ω can then be used, in turn, in Equation 8.24 to calculate a set of C_j. For a given value of ω_j, $(n - 1)$ of the C_j can be determined in terms of the nth one from Equation 8.24. The value for C_n must then be determined arbitrarily. Each set of C_j can be considered to define the components of an n-dimensional vector, called an eigenvector, of the system. We designate C_k as the eigenvector associated with the eigenfrequency ω_k and C_{kj} the kth component of the jth eigenvector. Accordingly, the factor C_i in Equations 8.23 and 8.24 should now be replaced by C_{ir}:

$$\sum_i C_{ir}A_{ij} = \omega_r^{-2}\sum_i C_{ir}K_{ij} = h_{jr} \tag{8.23a}$$

and

$$\sum_k (K_{jk} - \omega_r^2 A_{jk})C_{kr} = 0,\ j = 1,2,\ldots,n. \tag{8.24b}$$

The general motion of the system is a linear combination of the normal modes:

$$q_j(t) = \sum_k C_{jk}e^{i(\omega_k t - \delta_k)},\ j = 1,2,\ldots,n \tag{8.28a}$$

or, passing over to the real part of this expression,

$$q_j(t) = \sum_k C_{jk}\cos(\omega_k t - \delta_k),\ j = 1,2,\ldots,n. \tag{8.28b}$$

Thus, in general, a given coordinate will depend on C_{jk} and the frequencies of all modes of oscillation at the same time. In certain special circumstances, the system may only oscillate at one of the characteristic frequencies alone. When this happens, we speak of a normal mode of oscillation.

8.2.3 Orthogonality of Eigenvectors

There is a fundamental property of the eigenvector $\vec{C}_j$ that is important in determining the values of its components C_{jk} from the initial conditions. To investigate this fundamental property of the eigenvector $\vec{C}_j$, let us return to Equation 8.24, which gives, for the sth root ω_s,

$$\omega_s^2\sum_k A_{jk}C_{ks} = \sum_k K_{jk}C_{ks}. \tag{8.29a}$$

Interchanging j and k and replacing s by r, we obtain a comparable equation for the rth root of ω_r:

$$\omega_r^2\sum_k A_{jk}C_{jr} = \sum_k K_{jk}C_{jr} \tag{8.29b}$$

where use has been made of the symmetry of A_{jk} and C_{jk}.

First, multiplying Equation 8.29a by C_{jr} and summing over j, multiplying Equation 8.29b by C_{ks} and summing over k, and then subtracting the resulting equations, we obtain

$$(\omega_r^2 - \omega_s^2)\sum_{j,k} A_{jk} C_{jr} C_{ks} = 0. \tag{8.30}$$

For the case $\omega_r \neq \omega_s$, the sum must be zero identically:

$$\sum_{j,k} A_{jk} C_{jr} C_{ks} = 0. \tag{8.31}$$

The preceding condition is generally called the "orthogonality condition." When $\omega_r = \omega_s$, the factor $(\omega_r^2 - \omega_s^2)$ in Equation 8.30 vanishes, and the sum is indeterminate. But it can be shown that the sum is, in general, positive. To show this, we return to Equation 8.18 for the kinetic energy of the system and Equation 8.28 for q_j:

$$\begin{aligned} T &= \frac{1}{2}\sum_{j,k} A_{jk}\dot{q}_j\dot{q}_k \\ &= \frac{1}{2}\sum_{j,k} A_{jk}\left[\sum_r \omega_r C_{jr}\cos(\omega_r t - \delta_r)\right]\left[\sum_s \omega_s C_{ks}\cos(\omega_s t - \delta_s)\right] \end{aligned}$$

or

$$T = \frac{1}{2}\sum_{r,s}\omega_r\omega_s\cos(\omega_r t - \delta_r)\cos(\omega_s t - \delta_s)\sum_{j,k} A_{jk} C_{jr} C_{ks}.$$

For $\omega_r = \omega_s$, this expression becomes

$$T = \frac{1}{2}\sum_r \omega_r^2\cos^2(\omega_r t - \delta_r)\sum_{j,k} A_{jk} C_{jr} C_{kr}. \tag{8.32}$$

Now T is positive and can become zero only if all of the velocities vanish, and we note that

$$\omega_r^2\cos^2(\omega_r t - \delta_r) \geq 0.$$

Thus, in general, we have

$$\sum_{j,k} A_{jk} C_{jr} C_{kr} \geq 0 \tag{8.33}$$

where the equal sign applies when $T = 0$.

As remarked earlier, $(n - 1)$ of the C_{jr} can be obtained in terms of the nth one from Equation 8.24 when the ω_r are known. The value for the nth one must then be determined arbitrarily. We now remove this indeterminacy by choosing C_{jr} so that the eigenvector $\vec{C}_r$ has unit length:

$$\sum_{j,k} A_{jk} C_{jr} C_{kr} = 1. \tag{8.34}$$

Equation 8.34 is generally called the normalization condition. The orthogonality condition 8.31 and normalization condition 8.34 are often combined into one statement by the use of the Kronecker delta symbol δ_{ij}:

$$\sum_{j,k} A_{jk} C_{jr} C_{ks} = \delta_{rs}. \tag{8.35}$$

The vector $\vec{C}_r$ defined in this way constitutes an orthonormal set: the set is orthogonal according to Equation 8.31 and has been normalized according to Equation 8.34. The orthogonality and normalization conditions of Equation 8.35 are of more than academic interest; as we shall see, they prove useful in establishing the constants resulting from the initial conditions.

8.2.4 Normal Coordinates

The η_j given by Equation 8.22 are the desired normal coordinates. In terms of these coordinates, the small oscillations of the system are described by a set of harmonic oscillators of frequencies ω_j. Each oscillation can be excited independently. Once we know C_j, h_{jk} are readily obtained from Equation 8.23a. Finally, η_j follows from Equation 8.22. In terms of normal coordinates, the equations of motion are completely separated, as shown by Equation 8.21.

The normal coordinates are often introduced through the use of Equation 8.28 instead of Equation 8.20. However, because we have normalized C_{jr} according to Equation 8.34, we need to introduce a constant scale factor α that will depend on the initial conditions of the problem to account for the loss of generality that has been introduced by the arbitrary normalization. Thus,

$$q_j(t) = \sum_k \alpha C_{jk} e^{i(\omega_k t - \delta_k)} \tag{8.36}$$

or

$$q_j(t) = \sum_k \beta_k C_{jk} e^{i\omega_k t} \tag{8.36a}$$

where $\beta_k = \alpha e^{-i\delta_k}$, the new scale factors. The normal coordinates η_k are defined by the following relationships:

$$\eta_k(t) = \beta_k e^{i\omega_k t}. \tag{8.37}$$

It is evident that η_k undergoes oscillation at only one frequency. In terms of these new quantities, Equation 8.36a becomes

$$q_j(t) = \sum_k C_{jk} \eta_k(t). \tag{8.38}$$

Equation 8.38 can be considered as the inverse transformation of Equation 8.22.

The kinetic and potential energies assume the simple form

$$T = \frac{1}{2}\sum_j \dot{\eta}_j^2, \quad V = \frac{1}{2}\sum_j \omega_j^2\, \eta_j^2. \tag{8.39}$$

To show these, we first note that $\dot{q}_j(t) = \sum_k C_{jk}\dot{\eta}_k(t)$. Equation 8.18 becomes

$$T = \frac{1}{2}\sum_{j,k} A_{jk}\dot{q}_j\dot{q}_k = \frac{1}{2}\sum_{j,k} A_{jk}\left(\sum_r C_{jr}\dot{\eta}_r\right)\left(\sum_s C_{ks}\dot{\eta}_s\right)$$

from which we obtain, by rearrangement of summations,

$$T = \frac{1}{2}\sum_{r,s}\left(\sum_j\sum_k A_{jk}C_{jr}C_{ks}\right)\dot{\eta}_r\dot{\eta}_s$$

$$= \frac{1}{2}\sum_{r,s}(\delta_{rs})\dot{\eta}_r\dot{\eta}_s \quad \text{(with the aid of Equation 8.35)}$$

$$= \frac{1}{2}\sum_r \dot{\eta}_r^2, \qquad \text{Q.E.D.}$$

In terms of η_r, Equation 8.15 becomes

$$V = \frac{1}{2}\sum_{j,k} K_{jk}q_jq_k = \frac{1}{2}\sum_{r,s}\left(\sum_{j,k} K_{jk}C_{jr}C_{ks}\right)\eta_r\eta_s.$$

Now, multiplying Equation 8.29a by C_{jr} and summing over j, we obtain

$$\omega_s^2\sum_{j,k} A_{jk}C_{ks}C_{jr} = \sum_{j,k} K_{jk}C_{ks}C_{jr}.$$

Substituting this into the expression for U, we obtain

$$V = \frac{1}{2}\sum_{r,s}\left(\sum_{j,k} K_{jk}C_{jr}C_{ks}\right)\eta_r\eta_s = \frac{1}{2}\sum_{r,s}\left(\omega_s^2\sum_{j,k} A_{jk}C_{jr}C_{ks}\right)\eta_r\eta_s$$

$$= \frac{1}{2}\sum_{r,s}\left(\omega_s^2\delta_{rs}\right)\eta_r\eta_s = \frac{1}{2}\sum_r \omega_r^2\eta_r^2.$$

The Lagrangian L takes the simple form

$$L = T - V = \frac{1}{2}\sum_r\left(\dot{\eta}_r^2 - \omega_r^2\eta_r\right) \tag{8.40}$$

and the Lagrange's equation gives Equation 8.21:

$$\ddot{\eta}_r + \omega_r\eta_r = 0.$$

We now apply the general formulation just developed to the coupled pendulum.

Example 8.1: Coupled Pendulum Revisited

1. Eigenfrequencies
   The eigenfrequencies for the two-coupled pendulums are given by Equation 8.25. Thus, we must first calculate the quantities K_{ij} and A_{ij}. The potential energy of the system is given by Equation 8.15:

$$V = \frac{1}{2}K(x_1^2 + x_2^2) + \frac{1}{2}k(x_1 - x_2)^2 = \frac{1}{2}(K+k)(x_1^2 + x_2^2) - kx_1x_2 \tag{8.41}$$

   where we have replaced the original coordinates x and y by x_1 and x_2. From Equation 8.16,

$$K_{ij} = \left.\frac{\partial^2 U}{\partial x_1 \partial x_2}\right|_0, \quad i, j = 1, 2$$

   we obtain

$$\left.\begin{aligned} K_{11} &= \left.\frac{\partial^2 V}{\partial x_1^2}\right|_0 = K + k \\ K_{12} &= \left.\frac{\partial^2 V}{\partial x_1 \partial x_2}\right|_0 = -k = K_{21} \\ K_{22} &= \left.\frac{\partial^2 V}{\partial x_2^2}\right|_0 = K + k. \end{aligned}\right\}. \tag{8.42}$$

   The kinetic energy of the system is given by Equation 8.1a:

$$T = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{x}_2^2\right). \tag{8.43}$$

   According to the general formalism, we have

$$T = \frac{1}{2}\sum_{i,j} A_{ij}\dot{x}_i\dot{x}_j. \tag{8.44}$$

   Comparing these two expressions for T, we find

$$A_9 = A_{22} = m \text{ and } A_{12} = A_{21} = 0. \tag{8.45}$$

   Substituting the results of Equations 8.42 and 8.45 into Equation 8.25, we obtain

$$\begin{vmatrix} K + k - m\omega^2 & -k \\ -k & K + k - m\omega^2 \end{vmatrix} = 0. \tag{8.46}$$

   This is exactly Equation 8.5, and the solutions are the same as those given by Equations 8.7a and 8.7b.
2. Eigenvectors
   We return to Equation 8.24a to determine the eigenvector components C_{jr}:

$$\sum_k (K_{jk} - \omega_r^2 A_{jk})C_{kr} = 0. \tag{8.24a}$$

For $r = 1$ and $j = 1$, we have

$$(K_{11} - \omega_1^2 A_{11})C_{11} + (K_{12} - \omega_1^2 A_{12})C_{21} = 0. \tag{8.47}$$

Substituting $\omega_1^2 = K/m$ (from Equation 8.7a) and A_9 and K_9, and so forth, from Equations 8.42 and 8.45, we find

$$C_9 = -C_{21}. \tag{8.48}$$

The orthonormality condition can be used to determine C_9 and C_{21}. For the coupled pendulum, the orthonormality condition

$$\sum_{j,k} A_{jk} C_{jr} C_{ks} = \delta_{rs}$$

reduces to

$$\sum_{j,k} m\delta_{jk} C_{jr} C_{ks} = m \sum_j C_{jr} C_{js} = \delta_{rs}. \tag{8.49}$$

For $r = s = 1$, we have

$$C_{11}^2 + C_{21}^2 = 1/m. \tag{8.50}$$

Solving Equations 8.48 and 8.50, we find

$$C_{11} = -C_{21} = \sqrt{1/2m}. \tag{8.51}$$

For $r \neq s$, Equation 8.49 becomes

$$\sum_j mC_{jr} C_{js} = 0 = \sqrt{2m}(C_{12} - C_{22}) \tag{8.52}$$

from which we find

$$C_{12} = C_{22}. \tag{8.53}$$

For $r = s$, Equation 8.49 gives

$$C_{12}^2 + C_{22}^2 = 1/m. \tag{8.54}$$

Solving Equation 8.53 and 8.54, we obtain

$$C_{12} = C_{22} = \sqrt{1/2m}. \tag{8.55}$$

3. Normal coordinates
   We first calculate the quantities $h_{jk}\left(h_{jk} = \sum_i C_{ik} A_{ij}\right)$ and find

$$h_{11} = C_{11}A_{11} + C_{21}A_{21} = \sqrt{m/2}$$
$$h_{21} = C_{11}A_{12} + C_{21}A_{22} = -\sqrt{m/2}$$
$$h_{12} = C_{12}A_{11} + C_{22}A_{21} = \sqrt{m/2}$$
$$h_{22} = C_{12}A_{12} + C_{22}A_{22} = \sqrt{m/2}.$$

The normal coordinates η_1 and η_2 are given by

$$\eta_1 = h_{11}x_1 + h_{12}x_2 = \sqrt{m/2}(x_1 + x_2). \tag{8.56}$$

$$\eta_2 = h_{21}x_1 + h_{22}x_2 = \sqrt{m/2}(-x_1 + x_2). \tag{8.57}$$

They differ from Equations 8.9a and 8.9b by a constant factor.

Example 8.2: Longitudinal Vibrations of a CO_2 Molecule

Atoms in polyatomic molecules behave as the masses of our pendulum; the method of normal coordinates may be applied to the study of molecular vibrations—often with rich rewards. The carbon dioxide molecule that has the chemical structure O–C–O provides a simple example. We can regard this system as equivalent to a set of three particles joined by elastic springs (Figure 8.5). Clearly, the system will vibrate in some manner when subjected to an external force. For simplicity, we shall consider only longitudinal vibrations, and the interaction of the oxygen molecules with one another will be neglected (so we consider only the nearest neighbor interactions). Our coordinates are x_1, x_2, and x_3. The Lagrangian function L for the system is clearly

$$L = \frac{1}{2}m(\dot{x}_1^2 + x_3^2) + \frac{1}{2}M\dot{x}_2^2 - \frac{1}{2}k(x_2 - x_1)^2 - \frac{1}{2}k(x_3 - x_2)^2. \tag{8.58}$$

The equations of motion are found to be

$$\begin{aligned} m\ddot{x}_1 - k(x_2 - x_1) &= 0 \\ M\ddot{x}_2 + k(x_2 - x_1) - k(x_3 - x_2) &= 0 \\ m\ddot{x}_3 + k(x_3 - x_2) &= 0. \end{aligned} \tag{8.59}$$

Let

$$x_j = A_j \cos(\omega t - \alpha),\ j = 1, 2, 3. \tag{8.60}$$

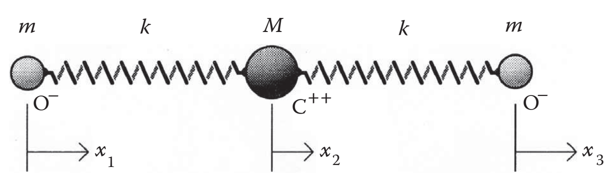


FIGURE 8.5 Linear symmetrical CO_2 molecule.

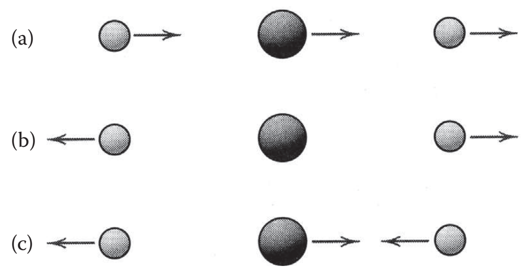


FIGURE 8.6 Longitudinal vibrations of a CO_2 molecule, (a) A pure translation of the system as a whole, (b) with the center mass at rest, the two end masses vibrate in opposite direction, and (c) the two end masses vibrate in unison, the center mass vibrates oppositely.

Substituting Equation 8.60 into Equation 8.59, we obtain

$$\begin{aligned}(-m\omega^2+k)A_1-kA_2&=0\\ -kA_1+(-M\omega^2+2k)A_2-kA_3&=0\\ -kA_2-(-m\omega^2+k)A_3&=0.\end{aligned}\tag{8.61}$$

The secular equation is

$$\begin{vmatrix}-m\omega^2+k & -k & 0\\ -k & -M\omega^2+2k & -k\\ 0 & -k & -m\omega^2+k\end{vmatrix}=0\tag{8.62}$$

or

$$\omega^2(-m\omega^2+k)(-mM\omega^2+kM+2km)=0\tag{8.62a}$$

from which we find the normal frequencies of the system

$$\omega_1=0,\quad \omega_2=\sqrt{k/m},\quad \omega_3=\sqrt{k/m+2k/M}.\tag{8.63}$$

1. Setting $\omega = 0$ in Equation 8.61, we find that $A_1 = A_2 = A_3$. Thus, this mode is no oscillation at all but is a pure translation of the system as a whole (Figure 8.6a).
2. Setting $\omega=\sqrt{k/m}$ in Equation 8.61, we find $A_2 = 0$ and $A_1 = -A_3$. Thus, the center mass M is at rest, while the two end masses vibrate in opposite directions with the same amplitude (Figure 8.6b).
3. Setting $\omega = (k/m + 2k/M)^{1/2}$ in Equation 8.61, we find $A_1 = A_3$ and $A_2 = -2A_1(m/M) = -2A_3(m/M)$. Thus, in this mode, the two end masses vibrate in unison, and the center mass vibrates oppositely with a different amplitude (Figure 8.6c).

8.3 FORCED OSCILLATIONS OF COUPLED OSCILLATORS

We shall find the transformation from the q-coordinates to the η-coordinates also useful in discussing the forced vibrations of coupled systems. Suppose that, in addition to the restoring forces that we have already considered in earlier sections, the component masses of the system are also acted upon by generalized forces $Q_1, Q_2,\ldots, Q_n$. In order to handle the resulting motion in terms of the normal coordinates, we must first find the generalized forces $R_1, R_2,..., R_n$ in the normal coordinate system during the displacements $\eta_1, \eta_2,\ldots, \eta_n$. We do this by setting up the work done in both coordinate systems. The work done, Q_iq_i by $Q_1, Q_2,\ldots, Q_n$ in displacements $q_1, q_2,..., q_n$ must be the same as the work done by a corresponding set of generalized forces $R_1, R_2, \ldots, R_n$ in the normal coordinate system during the displacements $\eta_1, \eta_2, \ldots, \eta_n$:

$$\sum_{i=1}^{n} Q_i q_i = \sum_{j=1}^{n} R_j \eta_j. \quad (8.64)$$

Substituting its value in terms of C_{ij} and η_j for each q_i from Equation 8.38 and equating terms in η_j, we obtain

$$R_j = \sum_{i=1}^{n} C_{ij} Q_i, \; j = 1, 2, \ldots, n \quad (8.65)$$

for the jth mode of oscillation. We notice from Equation 8.65 that a force applied to a particle cannot excite a mode in which that particle does not oscillate but will be very effective in exciting a mode in which it oscillates strongly. In terms of R_j, the Lagrange's equation of motion gives

$$\ddot{\eta}_j + \omega_j^2 \eta_j = R_j. \quad (8.66)$$

We can use all the methods learned in Chapter 7 in discussing the motion of each of the normal coordinates. As an example, consider that the forces Q_i all depend on time t through a factor $\exp(i\omega t)$, where ω is arbitrary. Then R_j will likewise vary according to this exponential. To solve Equation 8.66, we assume that η_j varies in the same way, and we find, as in Chapter 7,

$$\eta_j = \frac{R_j}{\omega_j^2 - \omega^2} \quad (8.67)$$

and

$$q_i = \sum_j \frac{C_{ij} R_j}{\omega_j^2 - \omega^2}. \quad (8.68)$$

Thus, the phenomenon of resonance occurs in the forced motion of coupled systems just as it does in the case of a single forced oscillator. The essential difference is that there are now n separate resonance points, corresponding to the n frequencies ω_j of the normal modes.

The forces Q_i could be damping forces. For example, as a result of friction, there will be a force $-c_i\dot{q}_i$ acting on the component mass m_i, where c_i is a constant. Then

$$R_j = -\sum_i C_{ij} c_i \dot{q}_i = -\sum_i C_{ij} c_i \sum_k C_{ik} \dot{\eta}_k = -\sum_k \gamma_{jk} \dot{\eta}_k \quad (8.69)$$

where

$$\gamma_{jk} = \sum_i c_i C_{ij} C_{ik}. \quad (8.70)$$

The equations of motion now take the form

$$\ddot{\eta}_j + \omega_j^2 \eta_j = -\sum_k \gamma_{jk} \dot{\eta}_k. \tag{8.71}$$

These equations cannot be separated if all the overtones are excited, and in this case, we cannot have a general solution. A simple solution exists only if one overtone is excited, say, the *j*th, for then all the η's except η_j are zero, and Equation 8.71 becomes

$$\ddot{\eta}_j + \gamma_{jj} \dot{\eta}_j + \omega_j^2 \eta_j = 0 \tag{8.72}$$

which is of the same form as Equation 8.29 for a damped oscillator.

It is evident that for forced oscillations of coupled damped oscillators, we have the following equations of motion:

$$\ddot{\eta}_j + \gamma_{jj} \dot{\eta}_j + \omega_j^2 \eta_j = R_j - \sum_{k \neq j} \gamma_{jk} \dot{\eta}_k. \tag{8.73}$$

We shall not go on to treat this problem in detail.

8.4 COUPLED ELECTRIC CIRCUITS

The examples we have been taking up so far have related to coupled mechanical oscillators. We shall see that methods developed earlier apply to other coupled systems, such as electric circuits. When a *LC* circuit is brought close to another similar electric circuit, oscillations in the one can be induced in the other as shown in Figure 8.7a. Figure 8.7b and c shows two different coupling schemes. For the purposes of illustration, we shall analyze the circuit in Figure 8.7a. Kirchhoff's circuit rules require that, in traversing each circuit completely, the total voltage drop be zero:

$$L\dot{I}_1 + q_1/C + M\dot{I}_2 = 0 \tag{8.74a}$$

$$L\dot{I}_2 + q_2/C + M\dot{I}_1 = 0. \tag{8.74b}$$

Differentiating with respect to time, we obtain

$$L\ddot{I}_1 + I_1/C + M\ddot{I}_2 = 0 \tag{8.75a}$$

$$L\ddot{I}_2 + I_2/C + M\ddot{I}_1 = 0. \tag{8.75b}$$

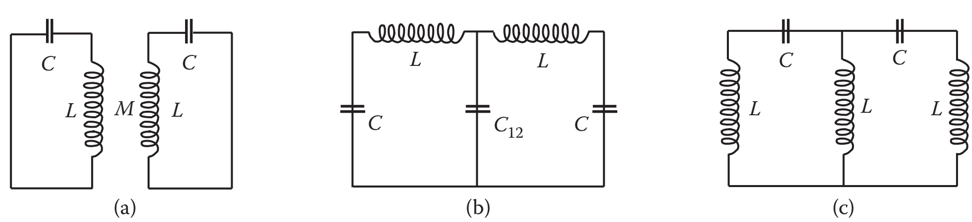


FIGURE 8.7 Coupled electric oscillator, (a) Inductive coupling, via mutual inductance *M*, (b) Capacitive coupling, via C_{12}, and (c) Inductive coupling, via L_{12}.

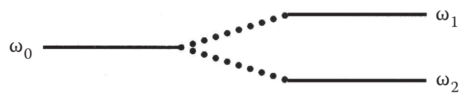


FIGURE 8.8 Splitting the common frequency by coupling.

Multiplying Equation 8.75a by L and Equation 8.75b by M and then subtracting the resulting equations, we obtain

$$(L^2 - M^2)\ddot{I}_1 = -(L/C)I_1 + (M/C)I_2. \tag{8.76a}$$

Now, multiplying Equation 8.75a by M and Equation 8.75b by L and then subtracting the resulting equations, we have

$$(L^2 - M^2)\ddot{I}_2 = -(L/C)I_2 + (M/C)I_1. \tag{8.76b}$$

These equations are just Equations 8.2a and 8.2b with different coefficients. They can be solved in the same way, and the characteristic frequencies of the two normal modes can be found to be

$$\omega_1 = 1/\sqrt{LC - MC}, \quad \omega_2 = 1/\sqrt{LC + MC}. \tag{8.77}$$

The effect of coupling is to split the common frequency for the uncoupled motion $\omega_0 = 1/(LC)^{1/2}$ into two frequencies ω_1 and ω_2 with one being lower than ω_0 and the other higher than ω_0 as shown in Figure 8.8. The same phenomenon was observed for the two-coupled pendulums.

PROBLEMS

1. A particle of mass is free to move along a line and is attached to a spring whose other end is fixed at a point, say, A (Figure 8.9). Point A is at a distance l from the line. If a force F is required to extend the spring to length Δl, determine the frequency of small oscillations of the mass.

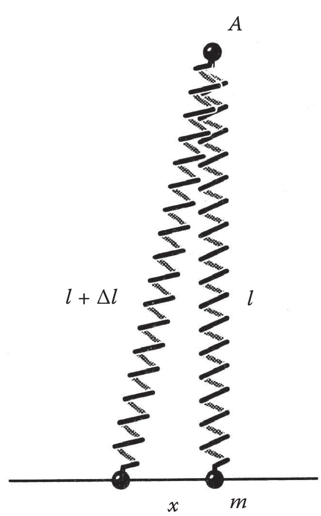


FIGURE 8.9 Mass–spring system; mass can move along a horizontal line.

2. When the masses of the coupled pendulum of Figure 8.1 are no longer equal, the equations of motion become

$$m_1\ddot{x} = -m_1\left(\frac{g}{b}\right)x - k(x-y). \tag{8.78}$$

$$m_2\ddot{y} = -m_2\left(\frac{g}{b}\right)y + k(x-y). \tag{8.79}$$

Show that we can choose the normal coordinates

$$X = \frac{m_1 x + m_2 y}{m_1 + m_2}$$

with a normal mode frequency $\omega_1^2 = g/b$, and $Y = x - y$ with a normal mode frequency $\omega_2^2 = g/b + k(1/m_1 + 1/m_2)$. Note that X is the coordinate of the center of mass of the system while the effective mass in the Y mode is the reduced mass μ of the system ($1/\mu = 1/m_1 + 1/m_2$).

3. Let the system of Problem 2 be set in motion with the initial conditions $x = A$, $y = 0$, and $\dot{x} = \dot{y} = 0$ at $t = 0$. Show that the normal amplitudes are $X_0 = (m_1/M)A$ and $Y_0 = A$ to yield

$$x = \frac{A}{M}(m_1\cos\omega_1 t + m_2\cos\omega_2 t)$$

$$y = A\frac{m_1}{M}(\cos\omega_1 t - \cos\omega_2 t)$$

where $M = m_1 + m_2$.

4. Determine the oscillations of a system with two degrees of freedom whose Lagrangian is

$$L = \frac{1}{2}(\dot{x}^2 + \dot{y}^2) - \frac{1}{2}\omega_0^2(x^2 + y^2) + \alpha\, xy.$$

5. Consider the mass–spring system of Figure 8.2 in the event that the two masses are different, m and M. Find the frequencies of the normal modes of vibration.
6. A double pendulum, both parts of which are of equal mass and length, vibrate in a vertical plane. Find the Lagrangian of the system and obtain the equations of motion for the case where the oscillations are assumed to be small. Find the normal modes and normal frequencies.
7. Three equal masses m are connected by four identical springs (Figure 8.10). Find the normal modes for motion along the line joining the masses.
8. A water molecule is a triatomic molecule, consisting of two hydrogen atoms and one oxygen atom constrained by chemical bonding to take up the planar triangle configuration. When the three atoms are at their equilibrium positions, the OH distance is 0.958×10^{-8} cm and the HOH angle 2α is 104.5°. Find the normal coordinates and the modes of vibration associated with each, considering only nearest neighbor interaction (Figure 8.11).
9. Consider two identical springs with two equal masses, suspended as in Figure 8.12.

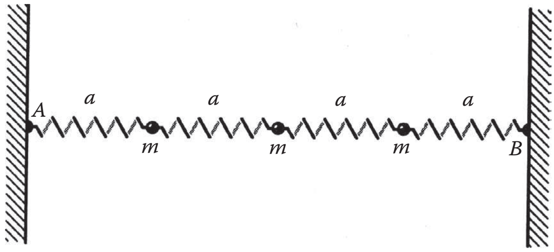


FIGURE 8.10 Three equal masses connected by four identical springs.

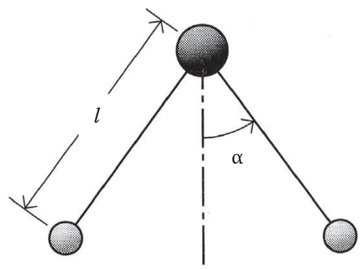


FIGURE 8.11 Water molecule.

Gravity acts vertically downward. Find the frequencies of the system and the normal coordinates, assuming that the motion is only in the vertical direction.

10. A bead of mass m slides on a smooth uniform circular ring of mass M and radius R. If the ring oscillates vertically under gravity about a fixed point on its circumference as shown in Figure 8.13, find the frequencies and the normal coordinates.
11. Show that the equations of motion for longitudinal vibrations of a loaded string are of exactly the same form as the equations for transverse motion.
12. Find the characteristic frequencies of the coupled electric circuits of Figure 8.7c.
13. A simple pendulum of mass m and length b is attached to a block of mass M that is constrained to slide on a smooth track. Find the frequencies of the small oscillations of the system and the modes of vibration associated with them (Figure 8.14).

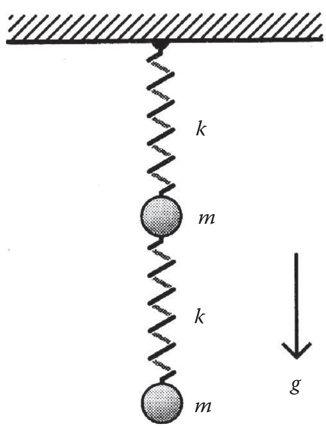


FIGURE 8.12 Two identical springs and two equal mass systems.

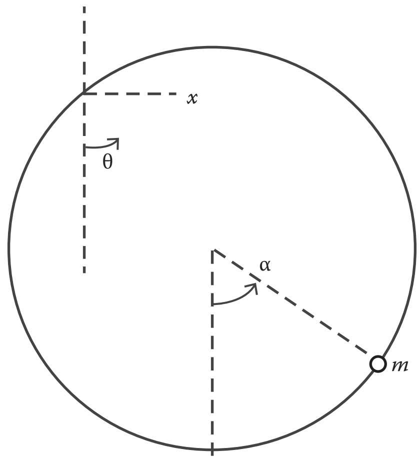


FIGURE 8.13 Bead–circular ring system.

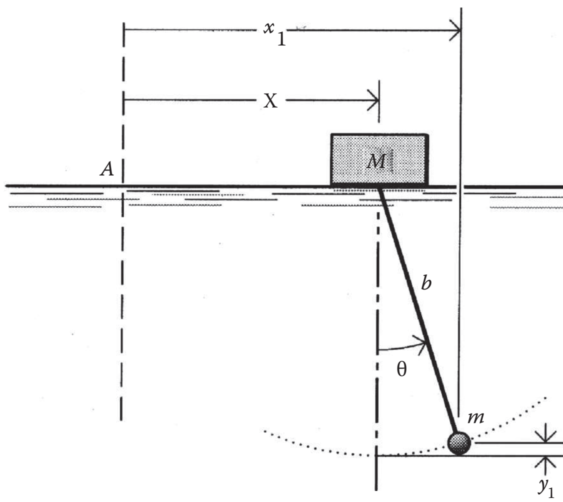


FIGURE 8.14 Simple pendulum attached to a moving mass.

14. Consider the coupled pendulums of Figure 8.1. A force $F\cos \omega t$ is applied to each particle in the same direction, and damping forces are assumed to be proportional to the velocities of the particles. Find the normal coordinates, and show that only one normal coordinate is excited.
15. Crystalline solids are characterized by a regular and symmetric arrangement of the atoms in the solid. A structure of this type is called a crystal lattice, for example, the sodium chloride crystal NaCl. The atoms are bound together by chemical bonding forces. We can regard the system as a set of particles joined by elastic springs, a very oversimplified model for a crystal lattice but one that often yields interesting results. For simplicity, consider the vibration of a linear crystal, consisting of a chain of two different types of atoms occupying alternative positions. In equilibrium, the particles are spaced at equal intervals along the chain. Investigate the longitudinal vibrations of this chain.
16. Three equal masses are so connected by springs that they move along a circle (Figure 8.15). Point A is fixed. Find the normal coordinates and eigenfrequencies of the system.

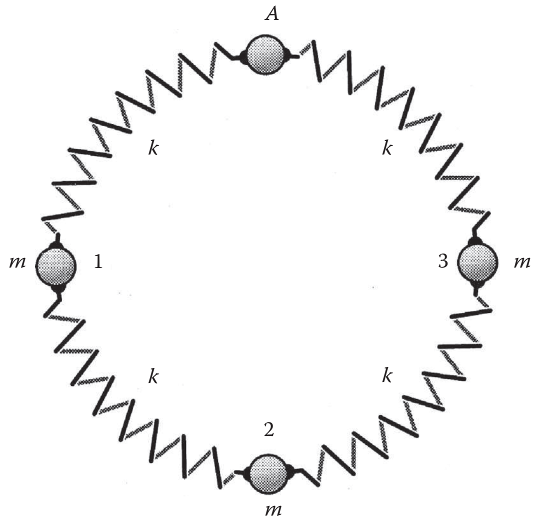


FIGURE 8.15 Special mass–spring system.

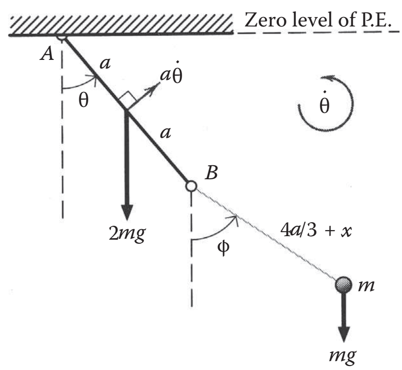


FIGURE 8.16 Uniform rod-elastic spring–mass system.

17. A uniform rod AB, of mass $2m$ and length $2a$, swings about a horizontal axis through the end A. One end of a light elastic string is attached to the end B of the rod and carries a particle of mass m at its other end (Figure 8.16). When the system is in stable equilibrium, the string is of length $4a/3$, its extension being ε. If the system performs small oscillations in the vertical plane through AB, the rod and the string making angle θ and φ, respectively, with the downward vertical and the length of the string being $(x + 4a/3)$, show that

$$x,\ \varphi + 2\theta,\ 2\varphi - 3\theta$$

are normal coordinates for the system, and the length of the corresponding simple pendulums are ε, $8a/3$, and $a/3$.

9 Nonlinear Oscillations

Most oscillation systems are really nonlinear, and linear oscillations are the exceptions. If the restoring force deviates from a linear dependence in the displacement or if the damping force is nonlinear in the velocity, the motion about a point of stable equilibrium is no longer linear and the motion is anharmonic or nonlinear. The general equation of motion for such a system is, for the one-dimensional case, a nonlinear differential equation of the following form:

$$m\ddot{x} + g(\dot{x}) + f(x) = 0. \tag{9.1}$$

Here, $f(x)$, the restoring force, is not proportional to x; $g(\dot{x})$, the damping term, is not proportional to $\dot{x}$. The principle of superposition does not apply to Equation 9.1. Thus, if two particular solutions are known, this does not immediately lead to a general solution. In general, every problem must be treated as a special case, and an approximate method must be employed to find solutions of the equation of motion. Hence, the subject of anharmonic or nonlinear effects has long been one of the more interesting and difficult problem areas in physics and engineering. Nonlinear effects are responsible for many important physical properties—thermal expansion and lattice thermal conductivity in solids, just to name a few. Radio and television communication would not exist in the way we know them without nonlinear elements. The way we see and hear is also based on nonlinearities in the eye and ear.

The simple pendulum with a large angular displacement is a good example of anharmonic oscillations. The exact equation of motion for an idealized pendulum is

$$\ddot{\theta} + (g/l)\sin\theta = 0 \tag{9.2}$$

which is a nonlinear differential equation because of the function $\sin\theta$. We know that

$$\sin\theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \frac{\theta^7}{7!} + \ldots .$$

It is only in the limit of small oscillations that $\sin\theta \cong \theta$ and Equation 9.2 represents simple harmonic motion. As a second example, consider an intrinsically nonlinear spring system. In Figure 9.1, a particle of mass m is held at rest by two identical springs, each of natural length b_0 with a spring constant of k, and the whole system is on a smooth table. In equilibrium, the springs are extended to a new length b; the particle is pulled to the side and released. If the horizontal displacement is x, each spring is stretched by a small amount $\left(\sqrt{x^2 + b_0^2} - b_0\right)$. For practice, let us use the Lagrange method to obtain the equation of motion. The potential energy of the system is

$$V(x) = 2 \cdot \frac{1}{2} k \left(\sqrt{x^2 + b_0^2} - b_0\right)^2.$$

The Lagrangian L of the system is

$$L = T - V = \frac{1}{2} m\dot{x}^2 - k\left(\sqrt{x^2 + b_0^2} - b_0\right)^2$$

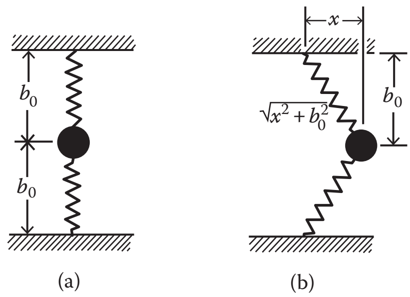


FIGURE 9.1 (a) Mass tethered symmetrically by two springs and (b) mass is pulled to the side.

and the equation of motion is

$$m\ddot{x} + 2kx\left[1 - \left(1 + \frac{x^2}{b_0^2}\right)^{-1/2}\right] = 0. \tag{9.3}$$

If x/b_0 is small, then the expansion of the radical gives

$$2kx\left[1 - \left(1 - \frac{1}{2}\left(\frac{x}{b_0}\right)^2 + \ldots\right)\right] = \frac{kx^3}{b_0^2}$$

and Equation 9.3 reduces to

$$m\ddot{x} + 2\frac{k}{b_0^2}x^3 = 0 \tag{9.4}$$

which indicates that the system is intrinsically nonlinear.

A large number of qualitative and quantitative methods for analyzing nonlinear problems have been developed. We shall discuss several of these useful methods.

9.1 QUALITATIVE ANALYSIS: ENERGY AND PHASE DIAGRAMS

Dynamic systems whose equations of motion do not explicitly contain the time variable are called autonomous systems. For a nonlinear autonomous system of the general form of Equation 9.1, we can extract some information about the properties of the system without actually solving the equation of motion; qualitative considerations can often be expressed graphically in terms of energy diagrams or phase diagrams. We begin with energy diagrams.

In Chapter 3, we saw that qualitative information about the motion of a particle subject to a conservative force can be deduced from a graph of the potential energy. If there is no damping present, the energy equation is, for the one-dimensional case,

$$\frac{1}{2}m\dot{x}^2 + V(x) = E$$

or

$$\dot{x} = \sqrt{\frac{2}{m}[E - V(x)]}. \tag{9.5}$$

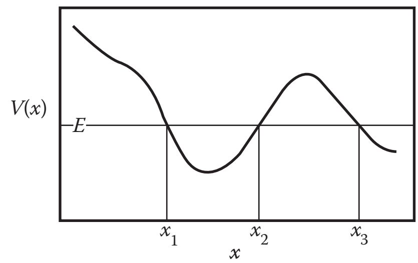


FIGURE 9.2 Graph of a one-dimensional potential $V(x)$.

If the potential energy has the general form shown in Figure 9.2 and the particle has a total energy E, then we can distinguish several possible cases for the motion:

(a) $x < x_1$. The region is excluded because $E < V(x)$.
(b) $x_1 < x < x_2$. The particle is permitted to move within the limits $x = x_1$ and $x = x_2$, and these two points are the turning points of the motion. The period of motion is given by

$$\tau(E) = \int_{x_1(E)}^{x_2(E)} \frac{\mathrm{d}x}{\sqrt{[E - V(x)]/2m}} \tag{9.6}$$

where the turning points x_1 and x_2 are given for $\dot{x} = 0$ by the roots of Equation 9.5. Note that the integrand becomes infinite at x_1 and x_2, but on physical grounds, the integral must exist because the motion lies entirely within the potential well.
(c) $x_2 < x < x_3$. The particle is excluded from this region because $E < V(x)$.
(d) $x > x_3$. In this region $E > V(x)$, it is a permissible region. If the particle approaches $x = x_3$ from larger x, it will encounter the potential barrier and be reflected back.

If the potential energy is of the form $V(x) = 2kx^2$, then the motion is just a simple harmonic motion, and the restoring force is $-kx$. If $V(x)$ deviates from $-2kx^2$, say, $V(x) = 2kx^2 - 3\varepsilon x^4$, where ε is a small quantity, then, except for the region in the vicinity of $x = 0$, where $V(x) \sim 2kx^2$, the system is a nonlinear one. A nonlinear system is said to be soft if the restoring force is less than the linear approximation; otherwise, it is said to be hard. For example, the restoring force derived from the potential energy just given is $f(x) = -kx + \varepsilon x^3$, so if $\varepsilon > 0$, the system is hard, and if $\varepsilon < 0$, the system is soft (Figure 9.3).

Phase diagrams are very useful tools for a qualitative study of nonlinear problems. In fact, almost every description of nonlinear dynamic systems is rooted in phase space. It is relatively easy to visualize, in phase space, the behavior of a particle that is under the influence of an arbitrary potential.

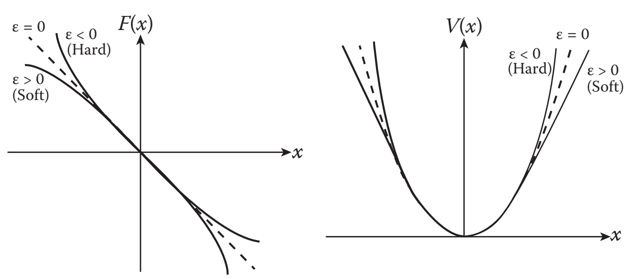


FIGURE 9.3 Form of force and potential energy for both hard and soft systems.

Returning now to Equation 9.1, let us treat the velocity $\dot{x}$ of the system as an independent variable and write it as y. Equation 9.1 then can be written as a pair of first-order differential equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = y, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = -g(y) - f(x) \tag{9.7}$$

where we have absorbed the constant factor $1/m$ into the functions $g(y)$ and $f(x)$. Next, we use the chain rule to eliminate t:

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}y/\mathrm{d}t}{\mathrm{d}x/\mathrm{d}t} = -\frac{g(y)+f(x)}{y}, \quad y \neq 0. \tag{9.8}$$

Equation 9.8 defines a definite curve in the phase plane with coordinates (x,y). The curve defined by Equation 9.8 is thus called the phase trajectory or phase path of the system. Equation 9.7 is a special case of the more general system

$$\frac{\mathrm{d}x}{\mathrm{d}t} = P(x,y), \quad \frac{\mathrm{d}y}{\mathrm{d}t} = Q(x,y) \tag{9.9}$$

where P and Q are functions of x and y. The phase trajectory of the system defined by Equation 9.9 is given by the differential equation

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{Q(x,y)}{P(x,y)}, \quad P(x,y) \neq 0. \tag{9.10}$$

Now, we can see the similarity between this and the phase concepts introduced in Chapter 6, that is, the canonical variables q and p are now replaced by the variables x and y, respectively. A point (x, y) of the phase plane for which the functions $P(x, y)$ and $Q(x, y)$ vanish simultaneously is called a singular point. Clearly, at a singular point, we have $\dot{x} = 0$ and $\dot{y} = 0$. This implies that a singular point is a stationary point of flow.

For the purpose of illustration, let us look at the following simple example.

Example 9.1

Consider the equation of motion of a particle being repulsed from the origin by a force proportional to its distance from the origin:

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = a^2x \tag{9.11}$$

where $a = \sqrt{k/m}$, and k is the force constant.

Solution:

Because

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\mathrm{d}x}{\mathrm{d}t}\right) = \frac{\mathrm{d}y}{\mathrm{d}t},$$

the equation of motion can be rewritten as

$$\frac{dy}{dt} = a^2 x. \tag{9.12}$$

Thus, the phase trajectories are given by the equation

$$\frac{dy}{dx} = \frac{dy/dt}{dx/dt} = \frac{a^2 x}{y}$$

or

$$y dy - a^2 x dx = 0.$$

Integration gives

$$y^2 - a^2 x^2 = c \tag{9.13}$$

with c being an integration constant. Equation 9.13 is a hyperbola in the phase plane. For various values of c, we get a family of hyperbolas whose two asymptotes are $y = \pm ax$ (Figure 9.4). The origin O is a singular point of a special type called a saddle point.

Sometimes the phase path can be easily sketched for the motion of a particle in a potential well. For example, Figure 9.5a shows the variation of some arbitrary potential energy $V(x)$ as a function of x. The potential energy represents a system that is soft for $x > 0$ and hard for $x < 0$. In the absence of damping, $\dot{x}$ is proportional to $\sqrt{E - V(x)}$, and the phase diagram is of the form shown in Figure 9.5b. If the system is under damping, the oscillating particle will spiral down the potential well and eventually come to rest at $x = 0$, which is a position of stable equilibrium.

Figure 9.6 shows another example: the potential energy curve with two symmetrically placed minima. The trajectories in phase space must be of the form shown in Figure 9.6b. When $E < V(0)$, the motion is confined in a single potential well; if the displacement is small, the restoring force is approximately linear, which means that the trajectory in phase space is an ellipse. When $E = V(0)$ and $\dot{x} = 0$, the intersecting points are the turning points of the motion along the x-axis. Note that the particle approaches the turning point $x = 0$ indefinitely, while the turnaround time at the turning point $x = 0$ is finite. We can understand this difference in behavior by examining the force that slows down the particle. The restoring force gets progressively weaker as the particle approaches the origin $x = 0$, where it vanishes. The three trajectories, two lobes and the central point, are distinct noninteracting trajectories for all finite times.

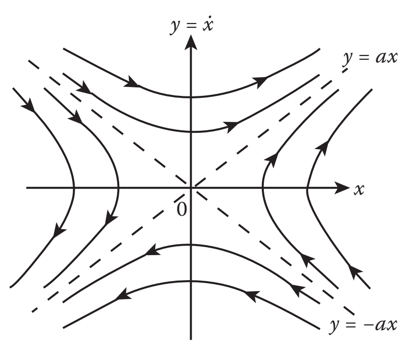


FIGURE 9.4 Phase trajectories of a periodic motion.

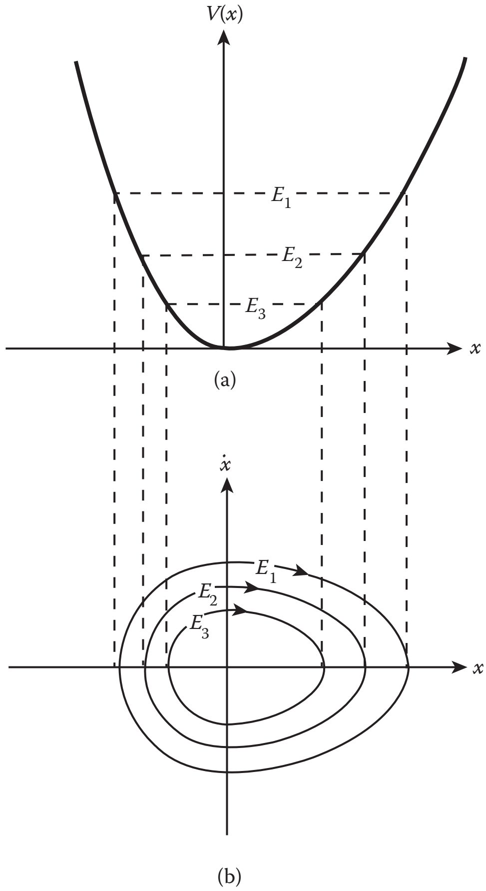


FIGURE 9.5 (a) Potential curve in configuration space and (b) qualitative distinct trajectories in phase plane $x - \dot{x}$.

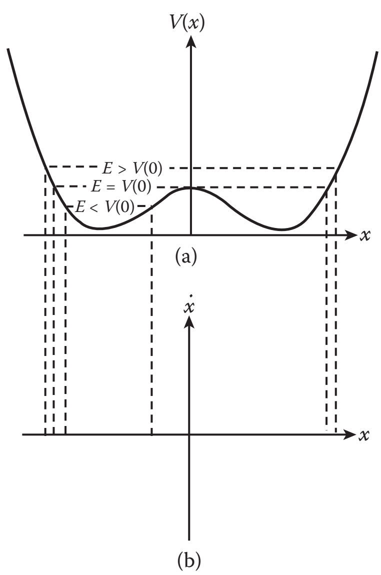


FIGURE 9.6 (a) Potential curve in configuration space and (b) qualitative distinct trajectories in phase plan $\dot{x} - x$.

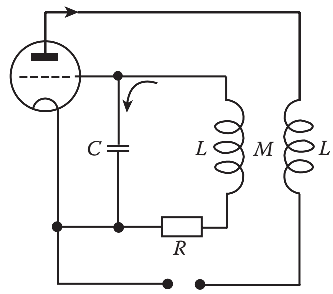


FIGURE 9.7 Electric circuit containing a triode valve.

An important feature of the trajectory in phase space is offered by van der Pol's equation. This equation was originally connected with a certain electric circuit containing a triode valve (Figure 9.7), and it is this valve that leads to the nonlinearity of the equation.

The non-dimensionalized electric current x is governed by the following equation of motion, the van der Pol's equation:

$$\ddot{x} + \varepsilon(x^2 - 1)\dot{x} + x = 0 \tag{9.14}$$

where ε is a small, positive constant. Figure 9.8 shows a self-excited oscillation for the case ε = 0.1 (van der Pol 1927). Even though the initial value of x is so small that it cannot be seen on the graph, the gradual onset of oscillations grows largely, and the oscillations are almost simple harmonic with period of approximately 2π.

We can see roughly how such self-excited oscillations come about by comparing Equation 9.14 with Equation 8.29 for the linearly damped oscillator. Note that the coefficient $\dot{x}$ in Equation 9.14 is not constant, but it is when the amplitude of motion $|x| > 1$, so the term as a whole might be expected to act as a damping mechanism when $|x|$ is reasonably large. However, if $|x| < 1$, the coefficient of $\dot{x}$ is negative, and we expect the term as a whole to have the opposite effect, causing growth of an oscillation, as would a negative value of k in Equation 8.29. Thus, there must be some value of amplitude for which the motion neither grows nor decays with time. In other words, there is a limit cycle in the phase plane.

To examine the behavior of the van der Pol oscillator in phase plane, we rewrite Equation 9.14 as a pair of first-order differential equations:

$$\dot{x} = y, \quad \dot{y} = \varepsilon(1 - x^2)y - x. \tag{9.15}$$

Thus, the integral curve is given by the equation

$$dy/dx = -x/y + \varepsilon(1 - x^2). \tag{9.16}$$

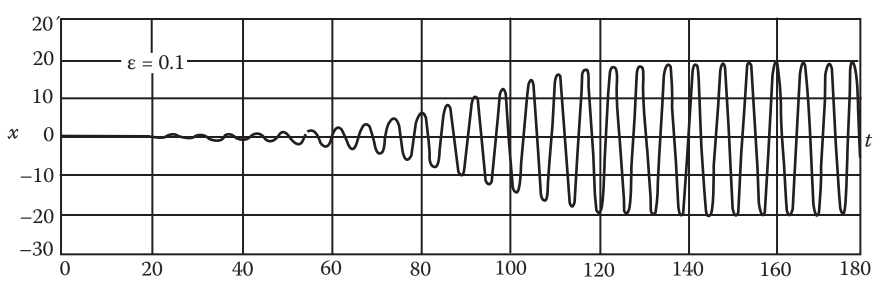


FIGURE 9.8 Self-excited oscillation.

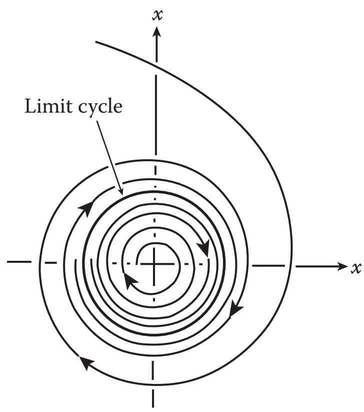


FIGURE 9.9 Phase diagram of a van der Pol oscillator for very small ε.

Equation 9.16 can be plotted for different values of the parameter ε. For ε = 0, we see that the phase path is a circle, and trajectories that lie outside the limit cycle spiral inward, and those that lie inside it spiral outward (Figure 9.9). But as ε is increased, the limit cycle will differ radically from a circle, indicating that while the limit cycle is certainly a periodic oscillation, it is no longer a simple harmonic one.

9.2 ELLIPTICAL INTEGRALS AND NONLINEAR OSCILLATIONS

The solution of certain types of nonlinear oscillation problems can be expressed in closed form by means of elliptic integrals. As an example, we consider a plane pendulum that is a simple pendulum with the string replaced by a massless, extensionless rod, so it can move in a vertical circle.

Example 9.2: A Plane Pendulum

The exact equation of motion for an idealized pendulum is

$$\ddot{\theta} + \frac{g}{l}\sin\theta = 0 \tag{9.17}$$

where θ can now take values up to 2π. When the value of θ is not small, the nonlinear effect becomes important. For complete accuracy, the full equation must be integrated. However, the system is conservative, so we can use the energy conservation equation ($T + V$ = constant) to obtain the solution. If we take the zero level of the potential energy along the horizontal, then

$$T = \frac{1}{2}ml^2\dot{\theta}^2, \quad V = -mgl\cos\theta$$

and we have

$$T + V = \frac{1}{2}ml^2\dot{\theta}^2 - mgl\cos\theta = E$$

or

$$\dot{\theta}^2 = \frac{2g}{l}\cos\theta + \frac{2E}{ml^2} = 2\frac{g}{l}\left(\cos\theta + \frac{E}{mgl}\right). \tag{9.18}$$

If θ_0 is the highest point of the motion, then

$$T(\theta_0) = 0,\ V(\theta_0) = -mgl\cos\theta_0 = E$$

and so

$$\frac{E}{mgl} = -\cos\theta_0.$$

Substituting this into Equation 9.18, we obtain

$$\dot{\theta}^2 = 4\frac{g}{l}\left(\sin^2\frac{\theta_0}{2} - \sin^2\frac{\theta}{2}\right) \tag{9.19}$$

where the trigonometric identity

$$\cos\theta = 1 - 2\sin^2(\theta/2)$$

has been used.

From Equation 9.19, we have

$$\mathrm{d}t = \frac{1}{2}\sqrt{\frac{l}{g}}\left(\sin^2\frac{\theta_0}{2} - \sin^2\frac{\theta}{2}\right)^{-1/2}\mathrm{d}\theta. \tag{9.20}$$

Equation 9.20 can be integrated to obtain the period τ. Two types of motion are possible: (1) oscillatory motion and (2) a complete circle. We shall only discuss the oscillatory motion, where the pendulum swings between the maximum angular displacement $\pm\,\theta_0$. $\dot{\theta} = 0$ when $\theta = \pm\,\theta_0$. Because the motion is symmetrical, the integral over θ from $\theta = 0$ to $\theta = \theta_0$ gives $\tau/4$; hence,

$$\tau = 2\sqrt{\frac{l}{g}}\int_0^{\theta_0}\left(\sin^2\frac{\theta_0}{2} - \sin^2\frac{\theta}{2}\right)^{-1/2}\mathrm{d}\theta. \tag{9.21}$$

The integral is an elliptical integral of the first kind. Now if we let

$$a = \sin\frac{\theta_0}{2} \quad \text{and} \quad z = \frac{\sin\frac{\theta}{2}}{\sin\frac{\theta_0}{2}} = a^{-1}\sin\frac{\theta}{2}$$

then

$$\mathrm{d}z = \frac{1}{2a}\cos\frac{\theta}{2}\mathrm{d}\theta = \frac{\sqrt{1-a^2z^2}}{2a}\mathrm{d}\theta.$$

In terms of z, Equation 9.21 becomes

$$\tau = 4\sqrt{\frac{l}{g}}\int_0^1[(1-z^2)(1-a^2z^2)]^{-1/2}\,\mathrm{d}z. \tag{9.22}$$

Numerical values for elliptical integrals can be found in various mathematical tables.

For oscillatory motion to result, $\theta_0 < \pi$, or equivalently, $\sin(\theta_0/2) = a < 1$. Thus, for this case, the integral in Equation 9.22 can be evaluated by expanding the factor $(1 - a^2z^2)^{-1/2}$:

$$(1-a^2z^2)^{-1/2} = 1+\frac{1}{2}(az)^2+\frac{3}{8}(az)^4+\dots\,.$$

Substituting this into Equation 9.22, we obtain the period

$$\begin{aligned}\tau &= 4\sqrt{\frac{l}{g}}\int_0^1(1-z^2)^{-1/2}\left[1+\frac{1}{2}(az)^2+\frac{3}{8}(az)^4\right]dz\\ &= 4\sqrt{\frac{l}{g}}\left[\frac{\pi}{2}+\frac{a^2}{2}\cdot\frac{1}{2}\cdot\frac{\pi}{2}+\frac{3a^4}{8}\cdot\frac{3}{8}\cdot\frac{\pi}{2}+\dots\right]\\ &= 4\sqrt{\frac{l}{g}}\left[1+\frac{a^2}{4}+\frac{9a^4}{64}+\dots\right].\end{aligned} \tag{9.23}$$

If a is large (i.e., near 1), many terms will be needed to produce a reasonably accurate result, but for small a, the expansion converges rapidly. As $a = \sin(\theta_0/2) \cong \theta_0/2 - \theta_0^2/410$, we have

$$\tau = 2\pi\sqrt{\frac{l}{g}}\left[1+\frac{\theta_0^2}{16}+\frac{11\theta_0^4}{3072}+\dots\right]. \tag{9.24}$$

Note that $\tau \to 2\pi(l/g)^{1/2}$ as $a \to 0$. We, therefore, see that a plain pendulum is not isochronous but is very nearly so only for small amplitudes of oscillation.

The phase diagram for the pendulum is easy to construct. When the oscillation amplitude is small ($\sin\theta \cong \theta$), the equation of the path in the phase plane $(\theta,\dot\theta)$ is a family of ellipses, and each path corresponds to a definite total energy of the oscillator. In the case of the finite oscillation amplitude, Equation 9.18 provides the necessary relationship, $\dot\theta = \dot\theta(\theta)$, for the construction of the phase diagram. For small values of θ and $\dot\theta$, the phase diagram should appear similar to that of the pendulum in the linear approximation. This can be seen directly from Equation 9.19, which reduces to

$$\left(\sqrt{l/g}\,\dot\theta\right)^2+\theta^2 \cong \theta_0^2$$

when θ and $\dot\theta$ are small angles. That is, in the phase plane $\left(\theta,\dot\theta/\sqrt{l/g}\right)$, the phase paths near $\theta = 0$ are approximate circles. But as θ approaches $\pm\pi$, the picture changes. For $-\pi < \theta < \pi$ and the total energy $E < mgl$ ($\cong E_0$), the situation is equivalent to a particle bound in the potential well $V(\theta) = -mgl\cos\theta$. For $E = E_0$ ($\theta_0 = \pi$), Equation 9.18 reduces to

$$\dot\theta = \pm 2\sqrt{g/l}\,\cos(\theta/2)$$

so the phase paths for $E = E_0$ are just cosine functions. There are two branches, depending on the direction of motion (Figure 9.10). Because the potential is periodic in θ, the same phase paths exist for regions $\pi < \theta < 3\pi$, $-3\pi < \theta < -\pi$, and so forth, and the point (0,0) and the set of points $(\pm 2n\pi, 0)$ in phase space, where n is an integer (0, +1, +2,...), are points of stable equilibrium for the pendulum; the pendulum is hanging down from the pivot. The set of points $[\pm(2n - 1)\pi, 0]$ are points of unstable equilibrium; the pendulum is standing on its pivot. In the vocabulary of chaotic dynamics, both sets of equilibrium points are called fixed points. Although the motion is still periodic for values of E that exceed E_0, it is no longer oscillatory; the pendulum executes a complete revolution about its pivot point.

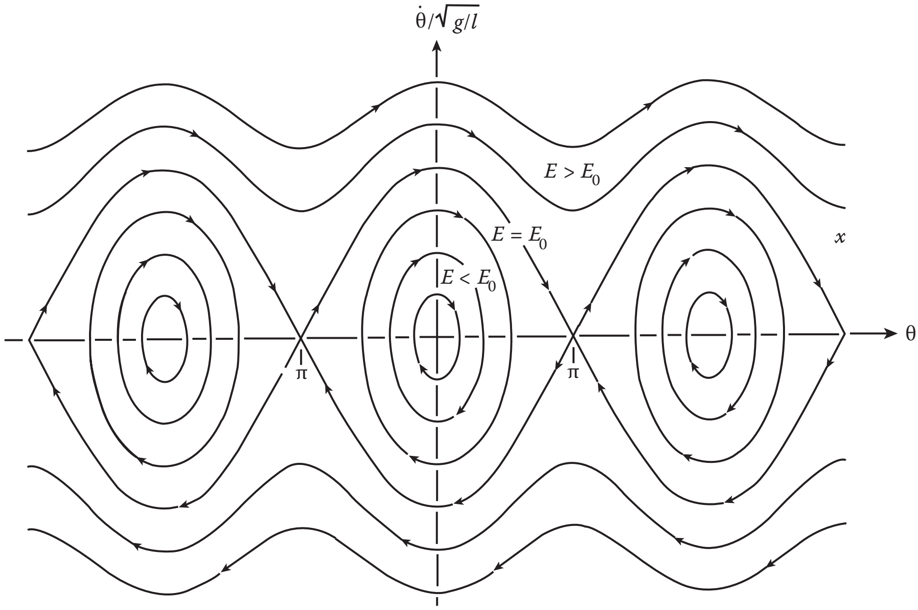


FIGURE 9.10 Phase portrait of a simple pendulum (no damping or driving force).

9.3 FOURIER SERIES EXPANSIONS

The Fourier series method is quite versatile and applies in any situation where the motion is periodic. As an example of application to nonlinear oscillations, we consider a free oscillation with nonlinear restoring forces, where the potential energy is an arbitrary function of x, $V = V(x)$. There is no loss in generality by setting $V(0) = 0$ at the equilibrium position $x = 0$. We further limit ourselves to a stable equilibrium where

$$V(x) > 0 \text{ for } x \neq 0 \text{ and } \mathrm{d}V(0)/\mathrm{d}x = 0.$$

Then Taylor's expansion of $V(x)$, for small oscillations, at the equilibrium position $x = 0$ takes the following form:

$$V(x) = \frac{1}{2}k_1x^2 + \frac{1}{3}k_2x^3 + \frac{1}{4}k_3x^4 \tag{9.25}$$

where

$$k_1 = \frac{\mathrm{d}^2V(0)}{\mathrm{d}x^2}, \quad k_2 = \frac{1}{2!}\frac{\mathrm{d}^3V(0)}{\mathrm{d}x^3}, \quad k_3 = \frac{1}{3!}\frac{\mathrm{d}^4V(0)}{\mathrm{d}x^4}.$$

The corresponding restoring force is

$$F(x) = -k_1x - k_2x^2 - k_3x^3 \tag{9.26}$$

and the equation of motion is

$$m\ddot{x} + k_1x + k_2x^2 + k_3x^3 = 0, \tag{9.27}$$

a nonlinear differential equation. This equation reduces to a simple harmonic motion when x is very small so that the term $(k_2x^2 + k_3x^3)$ becomes insignificant.

For many problems of physical importance, the potential function is either symmetrical or asymmetrical with respect to the point of equilibrium. We shall discuss the symmetrical case first.

9.3.1 Symmetrical Potential: $V(x) = V(-x)$

For this case, $V(x)$ is an even function and all derivatives of odd orders vanish:

$$\frac{\mathrm{d}^n V(0)}{\mathrm{d}x^n} = 0 \quad \text{for} \quad n = \text{odd}.$$

Thus, $k_2 = 0$ and Equation 9.27 reduces to

$$m\ddot{x} + k_1 x + k_3 x^3 = 0. \tag{9.28}$$

In the absence of the nonlinear term k_3x^3, the solution is a sinusoidal function with angular frequency $\omega_0 = (k_1/m)^{1/2}$. With the presence of the nonlinear term, we expect that the solution is of the following form:

$$x(t) = \sum_{n=0}^{\infty} [a_n \sin(n\omega t) + b_n \cos(n\omega t)] \tag{9.29}$$

where $\omega \neq \omega_0$, and many of the coefficients a_n and b_n can be shown to be zero by using the symmetry properties of the motion.

For simplicity and without loss of generality, we measure the time at the instant that $\dot{x}$ vanishes: $\dot{x}(0) = 0$. Under this initial condition, the displacement $x(t)$ must be a symmetric function of t:

$$x(t) = x(-t).$$

This means that the solution for $\pi \leq \omega t < 2\pi$ is a mirror image of the solution for $0 < \omega t < \pi$. This condition requires that the coefficients a_n vanish:

$$a_n = \frac{1}{\pi}\int_0^{2\pi} x(t)\sin(n\omega t)\mathrm{d}(\omega t) = 0 \quad \text{for all } n$$

and Equation 9.29 is then reduced to

$$x(t) = \sum_{n=0}^{\infty} b_n \cos(n\omega t). \tag{9.30}$$

Now, if we multiply both sides of Equation 9.30 by $\cos(n'\omega t)$ and integrate over $\mathrm{d}(\omega t)$ from $\omega t = 0$ to $\omega t = 2\pi$, we find

$$b_n = \frac{1}{\pi}\int_0^{2\pi} x(t)\cos(n\omega t)\mathrm{d}(\omega t) \tag{9.31}$$

where the following orthogonality condition has been used:

$$\int_0^{2\pi} \cos(n\omega t)\cos(n'\omega t)\mathrm{d}(\omega t) = \int_0^{2\pi} \sin(n\omega t)\sin(n'\omega t)\mathrm{d}(\omega t) = \pi\delta_{nn'}.$$

It is easy to show that $b_n = 0$ for all even n. For this reason, we set $n = 2\alpha$, α being an integer. Equation 9.31 then becomes

$$\begin{aligned} b_{2\alpha} &= \frac{1}{\pi}\int_0^{2\pi} x(t)\cos(2\alpha\omega t)\mathrm{d}(\omega t) \\ &= \frac{1}{\pi}\int_0^{2\pi} x(t)[\cos^2(\alpha\omega t) - \sin^2(\alpha\omega t)]\mathrm{d}(\omega t). \end{aligned}$$

As the oscillator oscillates symmetrically with respect to $x = 0$ (the equilibrium position), neither $\cos(n\omega t)$ nor $\sin(n\omega t)$ contributes to the integral. Thus,

$$b_{2\alpha} = 0 \text{ for all } \alpha.$$

The remaining coefficients, b_n for all odd n as well as the frequency ω of the motion, are to be determined from the differential equation. For this, we rewrite Equation 9.31 as

$$\begin{aligned} b_n &= \frac{1}{\pi}\int_0^{2\pi} x(t)\cos(n\omega t)\mathrm{d}(\omega t) \\ &= \frac{1}{\pi}\int_0^{2\pi} x(t)\frac{1}{n}\left[\frac{\mathrm{d}}{\mathrm{d}(\omega t)}\sin(n\omega t)\right]\mathrm{d}(\omega t). \end{aligned}$$

Integration by parts gives

$$\begin{aligned} b_n &= \frac{1}{n\pi}\left[x\sin(n\omega t)\Big|_0^{2\pi} - \int_0^{2\pi}\frac{\mathrm{d}x}{\mathrm{d}(\omega t)}\sin(n\omega t)\mathrm{d}(\omega t)\right] \\ &= -\frac{1}{n\pi\omega}\int_0^{2\pi}\frac{\mathrm{d}x}{\mathrm{d}(\omega t)}\sin(n\omega t)\mathrm{d}(\omega t). \end{aligned}$$

Integrating by parts once more, we obtain

$$\begin{aligned} b_n &= \frac{1}{n\pi^2\omega}\left[\frac{\mathrm{d}x}{\mathrm{d}t}\cos(n\omega t)\Big|_0^{2\pi} - \frac{1}{\omega}\int_0^{2\pi}\frac{\mathrm{d}^2x}{\mathrm{d}t^2}\cos(n\omega t)\mathrm{d}(\omega t)\right] \\ &= -\frac{1}{n\pi^2\omega^2}\int_0^{2\pi}\ddot{x}\cos(n\omega t)\mathrm{d}(\omega t) \end{aligned}$$

where $\dot{x} = 0$ at 0 and 2π has been used. Substituting for $\ddot{x}$ from the equation of motion yields

$$b_n = \frac{1}{m\pi n^2\omega^2}\int_0^{2\pi}(k_1x + k_3x^3)\cos(n\omega t)\mathrm{d}(\omega t) = \frac{k_1b_n}{mn^2\omega^2} + \frac{k_3}{m\pi n^2\omega^2}\int_0^{2\pi}x^3\cos(n\omega t)\mathrm{d}(\omega t).$$

Hence,

$$b_n = \frac{k_3}{\pi(mn^2\omega^2 - k_1)} \int_0^{2\pi} x^3 \cos(n\omega t)\mathrm{d}(\omega t). \tag{9.32}$$

By substituting $x(t) = b_1\cos\,\omega t + b_3\cos\,3\omega t + b_5\cos\,5\omega t + \cdots$ into the integral, we find

$$b_n = \frac{k_3}{\pi(mn^2\omega^2 - k_1)} \sum_{ijl} b_i b_j b_l \int_0^{2\pi} \cos(i\omega t)\cos(j\omega t)\cos(l\omega t)\cos(n\omega t)\mathrm{d}(\omega t). \tag{9.33}$$

By using the following trigonometric identities:

$$\cos A \cos B = \frac{1}{2}\cos(A+B) + \frac{1}{2}\cos(A-B)$$

$$\int_0^{2\pi} \cos(n\omega t)\cos(n'\omega t)\mathrm{d}(\omega t) = \pi\delta_{nn'}$$

the integration in Equation 9.33 can be carried out, and the final result is

$$\begin{aligned} b_n = \frac{k_3}{4(mn^2\omega^2 - k_1)} \sum_{ijl} b_i b_j b_l [&\delta_{n,i+j+l} + \delta_{n,i+j-l} \\ &+ \delta_{n,j+l-i} + \delta_{n,l+i-j} + \delta_{n,i-j-l} + \delta_{n,j-l-i} + \delta_{n,l-i-j}]. \end{aligned} \tag{9.34}$$

Equation 9.34 represents a set of simultaneous algebraic equations for the coefficients b_1, b_3,... and cannot be solved exactly. We must therefore make an additional approximation. Note that Equation 9.32 is a method of estimating the order of magnitude of the coefficients b_n. If the integral is replaced by its maximum value

$$|b_n| < \frac{|k_3|}{\pi(n^2\omega^2 m - k_1)} \int_0^{2\pi} x^3 \mathrm{d}(\omega t) = \frac{2|k_3|x^3}{n^2\omega^2 m - k_1}$$

where x is the amplitude of the motion, then we see that coefficients b_n decrease roughly in proportion to $1/n^2$. This information is very useful. For example, if we wish to obtain a solution good to the fifth order for a summary of Equation 9.34, we only need to keep the terms $i + j + l < 5$, and we, therefore, obtain

$$b_1 = \frac{k_3}{4(m\omega^2 - k_1)}(3b_1^3 + 3b_1^2 b_3) \tag{9.35a}$$

$$b_3 = \frac{k_3}{4(9m\omega^2 - k_1)}(b_1^3 + 6b_1^2 b_3) \tag{9.35b}$$

$$b_5 = \frac{k_3}{4(25m\omega^2 - k_1)} 3b_1^2 b_3. \tag{9.35c}$$

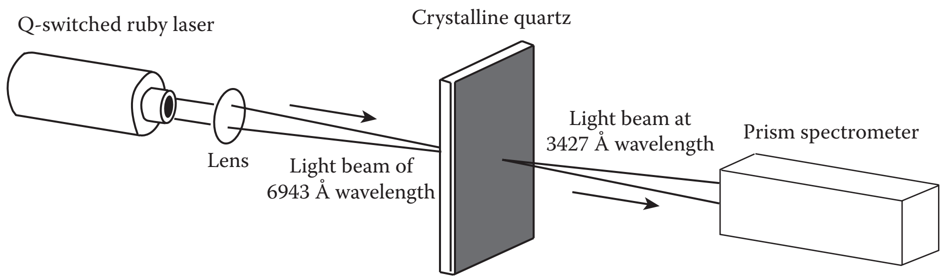


FIGURE 9.11 Generation of an optical harmonic. (From Franken, P.A. et al., *Physical Review*, 7, p. 118, 1961. With permission.)

There are four unknowns, b_1, b_3, b_5, and ω, in Equations 9.35a through 9.35c. We can solve three of them in terms of the remaining one, which is then determined by the initial conditions $x(0) = x_0$. We now solve ω, b_3, and b_5 in terms of b_1, and keeping terms not higher than the fifth order, we find

$$\omega^2 = \omega_0^2 + \frac{3k_3b_1^2}{4m} + \frac{3k_3^2b_1^4}{128mk_1} \tag{9.36a}$$

$$b_3 = \frac{k_3b_1^2}{32k_1}\left(1 - \frac{21k_3b_1^2}{32k_1}\right) \tag{9.36b}$$

$$b_5 = \frac{k_3^2b_1^5}{1024k_1^2}. \tag{9.36c}$$

The solution accurate to the fifth order is

$$x(t) = b_1\cos\omega t + \frac{k_3b_1^3}{32k_1}\left(1 - \frac{21k_3b_1^3}{32k_1}\right)\cos 3\omega t + \frac{k_3^2b_1^5}{1024k_1^2}\cos 5t. \tag{9.37}$$

Equation 9.37 contains harmonic displacements at the fundamental frequency and at the third and fifth harmonic frequencies. If we proceed to solve for terms involving orders higher than five, additional harmonics at 7ω, 9ω, and so on would appear in the solution. Such harmonic generation is of great significance because it permits us to synthesize from low-frequency signals to ever-higher frequency signals (with greatly diminished intensity). This has been known in electronics for decades, but a beautiful example comes to us from modern optics. Figure 9.11 shows the generation of an optical harmonic. The red incident light at a wavelength of 6943 Å emerges as a beam of green light of wavelength 3471.5 Å (halving the wavelength is equivalent to doubling the frequency). The intensity of the green light was about 1/10 that of the intensity of the incident red light. The cause of this second-harmonic generation lies in the form of the polarizability of quartz.

9.3.2 Asymmetrical Potential: $V(-x) = -V(x)$

If the potential function is not symmetric, k_2 will not vanish in Equation 9.25, and we must work with the equation of motion (Equation 9.24). Both periodic and nonperiodic solutions of Equation 9.27 exist; we consider here only the periodic solutions. If the same initial condition $x(0) = 0$ is used, then the requirement $x(t) = x(-t)$ remains true, and the solution will be of the form

$$x(t) = \sum_{n=0}^{\infty} b_n \cos(n\omega t). \tag{9.38}$$

The symmetry of the potential function means that both even and odd n are present in Equation 9.38. The determination of b_n can be made in the same way that we did for the symmetric case, with the result

$$b_n = \frac{1}{\pi(mn^2\omega^2 - k_1)} \int_0^{2\pi} x^2(k_2 + k_3 x)\cos(n\omega t)\mathrm{d}(\omega t). \tag{9.39}$$

If third-order accuracy is maintained, Equation 9.39 leads to

$$b_0 = -\frac{k_1}{k_3} b_1^2, \quad b_1 = \frac{1}{m\omega^2 - k_1}\left(2k_2 b_0 b_1 + k_2 b_1 b_2 + \frac{3}{4} k_3 b_1^3\right) \tag{9.40a}$$

$$b_2 = \frac{1}{2(4m\omega^2 - k_1)} k_2 b_1^2, \quad b_3 = \frac{1}{9m\omega^2 - k_1}\left(k_2 b_1 b_2 + \frac{1}{4} k_3 b_1^3\right) \tag{9.40b}$$

and the approximate Fourier series solution is

$$\begin{aligned} x(t) = -\frac{k_2}{k_1} b_1^2 + b_1 \cos\omega t + \frac{k_2}{6k_1} b_1^2 \cos 2\omega t \\ + \left(\frac{k_2^2}{48k_1} + \frac{k_3}{32k_1}\right) b_1^3 \cos 3\omega t \end{aligned} \tag{9.41}$$

with

$$\omega^2 = \omega_0^2 + \left(\frac{3k_3}{4m} - \frac{11}{6m}\frac{k_2^2}{k_1}\right) b_1^2. \tag{9.42}$$

It is interesting to note that the two terms in the coefficient of the third harmonic—which result from the presence of both the square and cube terms in Equation 9.27—are of the same order of magnitude. This is an interesting and important feature that is often overlooked.

9.4 THE METHOD OF PERTURBATION

From Newton's pioneering work in celestial mechanics to the modern space physicist's computerized orbital calculations, a form of perturbation theory has often been used to find approximate solutions to complicated equations of motion. We now consider a perturbation procedure for dealing with nonlinear differential equations. This procedure was originated by S.D. Poisson and extended by J.H. Poincare. Consider the differential equation of a one-dimensional nonlinear system to be of the form

$$\ddot{x} + \omega_0^2 x + \alpha f(x, \dot{x}) = 0 \tag{9.43}$$

where the quantity f is an analytic function of x and $\dot{x}$, and α is a small parameter. We consider the third term in Equation 9.43 to be a small perturbation and seek a solution in the form of a power series in the small parameter α:

$$x(t) = x_0(t) + \alpha x_1(t) + \alpha^2 x_2(t) + \alpha^3 x_3(t) + \dots . \tag{9.44}$$

From Equation 9.44, we obtain $\dot{x}$ and $\ddot{x}$ and substitute these into Equation 9.43, and equating the coefficients with the same powers of α, we obtain a number of differential equations that can be integrated recursively. The solution in the first order of approximation is correct, but if one attempts to obtain the solution in the next order of approximation (i.e., x_2), a serious problem may arise in the form of secular terms. These terms increase indefinitely as $t \to \infty$ and destroy the convergence of the series solution obtained. One of the methods for eliminating the secular terms is the Bogoliuboff–Kryloff procedure. In order to see the perturbation method and the Bogoliuboff–Kryloff procedure, we consider a simple example.

Example 9.3

Consider a particle moving in an asymmetric potential $V(x)$:

$$V(x) = \frac{1}{2}k_1x^2 + \frac{1}{3}k_2x^3 \tag{9.45}$$

and obtain a solution to the equation of motion to the second order.

Solution:

The equation of motion is

$$\ddot{x} + \omega_0^2 x + \alpha x^2 = 0 \tag{9.46}$$

where

$$\omega_0^2 = k_1/m, \quad \alpha = k_2/m. \tag{9.47}$$

Now α is, by hypothesis, a small quantity; the solution of Equation 9.46 is, therefore, expected to resemble that of a simple harmonic motion. Thus, we assume that the solution may be expressed as a power series of α:

$$x(t) = x_0(t) + \alpha x_1(t) + \alpha^2 x_2(t) + \dots \tag{9.48}$$

where x_0 is the solution of the simple harmonic motion, and x_1, x_2,… are to be determined.

Let us first limit ourselves to a first-order calculation, which means that we will further assume that α is sufficiently small so that the approximation

$$x(t) = x_0 + \alpha x_1(t) \tag{9.49}$$

is adequate. Substituting Equation 9.49 into Equation 9.46, we find

$$\ddot{x}_0 + \alpha\ddot{x}_1 + \omega_0^2 x_0 + \omega_0^2 \alpha x_1 + \alpha(x_0 + \alpha\ x_1)^2 = 0. \tag{9.50}$$

In accordance with our assumption, we neglect α^2 and α^3; then Equation 9.50 reduces to

$$(\ddot{x}_0 + \omega_0^2 x_0) + \alpha(\ \ddot{x}_1 + +\omega_0^2 x_1 + x_0^2) = 0. \tag{9.51}$$

Because Equation 9.51 is true for arbitrary α, each term in parentheses must vanish separately, and we have

$$\ddot{x}_0 + \omega_0^2 x_0 = 0 \tag{9.52}$$

$$\ddot{x}_1 + +\omega_0^2 x_1 = -x_0^2. \tag{9.53}$$

The solution of Equation 9.52 is

$$x_0(t) = A\cos(\omega_0 t + \delta).$$

For simplicity, we neglect phase δ from now on. Equation 9.53 then becomes

$$\ddot{x}_1 + \omega_0^2 x_1 = -A^2 \cos^2 \omega_0 t = -\frac{1}{2} A^2 (!+ \cos 2\omega_0 t). \tag{9.54}$$

The complementary solution of Equation 9.54 is $A\cos(\omega_0 t)$. We now seek a particular integral of Equation 9.54 in the form

$$x_1(t) = B\cos(2\omega_0 t) + C). \tag{9.55}$$

Substitution of this into Equation 9.54 gives

$$-3\omega_0^2 B \cos 2\omega_0 t + \omega_0^2 C = -\frac{1}{2} A^2 (1 + \cos 2\omega_0 t)$$

from which we find

$$B = A^2/6\omega_0^2, \quad C = -A^2/2\omega_0^2.$$

Thus, the solution of Equation 9.46 to the first order of α is

$$x(t) = A\cos\omega_0 t + \alpha \frac{A^2}{6\omega_0^2}(\cos 2\omega_0 t - 3). \tag{9.56}$$

The second harmonic ($2\omega_0$) of the fundamental frequency is generated in the solution. If terms involving higher powers of α had been included in Equation 9.45, higher harmonics such as $3\omega_0$, $4\omega_0$,… would have been generated.

We now proceed to obtain the solution to the next higher order of approximation:

$$x(t) = x_0 + \alpha x_1(t) + \alpha^2 x_2(t). \tag{9.57}$$

Substituting this into the equation of motion, Equation 9.43, gives the following three equations:

$$\ddot{x}_0 + \omega_0^2 x_0 = 0 \tag{9.58a}$$

$$\ddot{x}_1 + +\omega_0^2 x_1 + x_0^2 = 0 \tag{9.58b}$$

$$\ddot{x}_2 + +\omega_0^2 x_2 + 2x_0 x_1 = 0. \tag{9.58c}$$

As expected, Equations 9.58a and 9.58b are identical to Equations 9.52 and 9.53, respectively. Now,

$$x_0(t) = A\cos\omega_0 t, \quad x_1(t) = \frac{A}{6\omega_0^2}(\cos 2\omega_0 t - 3).$$

Substituting these into Equation 9.58c, we find

$$\ddot{x}_2 + \omega_0^2 x_2 = -\frac{A^3}{6\omega_0^2}\cos 3\omega_0 t + \frac{5A^3}{6\omega_0^2}\cos\omega_0 t \tag{9.59}$$

where the following trigonometric identity has been used:

$$\cos 3\omega_0 t + \cos\omega_0 t = 2\cos\omega_0 t\cos 2\omega_0 t.$$

Equation 9.59 is an inhomogeneous second-order linear differential equation of the general form

$$\ddot{x} + a\dot{x} + bx = f(t)$$

($a = 0$ in our case) whose solution has two parts, namely, a complementary solution (the solution of the homogeneous equation) and a particular integral. The particular integral is a term or terms that produce the right-hand side, $f(t)$, of the equation. If $f(t)$ and its first two derivatives (a second-order equation is being considered here) contain only linearly independent functions, then a good trial particular integral will be a linear combination of these functions. If the trial particular integral contains a term that already appears in the complementary solution, use the term multiplied by t; if this combination also appears in the complementary solution, use the term multiplied by t^2. We shall not prove it, because it is too tedious.

We now return to Equation 9.59. Because the term $\cos\omega_0 t$ on its right-hand side already occurs in the complementary solution, it is necessary to include a term proportional to $(t\cos\omega_0 t)$ in the particular integral:

$$x_2(t) = D\cos 3\omega_0 t + Et\sin\omega_0 t + F.$$

Substituting this into Equation 9.59, we find

$$\left(\frac{A^3}{6\omega_0^2} - 8\omega_0^2 D\right)\cos 3\omega_0 t + \left(2E\omega_0 - \frac{5A^3}{6\omega_0^2}\right)\cos\omega_0 t + \omega_0^2 F = 0.$$

Hence,

$$D = A^3/48\omega_0^4, \quad E = 5A^3/12\omega_0^3, \quad F = 0$$

and

$$x_2(t) = \frac{A^3}{48\omega_0^4}(\cos 3\omega_0 t + 20\omega_0 t\sin\omega_0 t). \tag{9.60}$$

Thus, the solution of Equation 9.46 to the second order of α^2 is

$$\begin{aligned} x(t) &= x_0(t) + \alpha x_1(t) + \alpha^2 x_2(t) \\ &= A\cos\omega_0 t + \alpha\frac{A^2}{6\omega_0^2}(\cos 2\omega_0 t - 3) \\ &\quad + \alpha^2\frac{A^3}{48\omega_0^4}(\cos 3\omega_0 t + 20\omega_0 t\sin\omega_0 t). \end{aligned} \tag{9.61}$$

The term $20\omega_0 t\sin\omega_0 t$ in Equations 9.60 and 9.61 is called a secular term, one that always increases with time. The reason for the appearance of such a term is easy to understand. When the nonlinear term is zero, the solution is periodic with a frequency, say, ω_0. But the inclusion of the nonlinear term may render the solution periodic, but with a new frequency ω that differs from ω_0 by a small amount δ: $\omega = \omega_0 + \delta$. Now, the expansion of $\sin \omega t = \sin(\omega_0 + \delta)t$ results in

$$\sin(\omega_0 + \delta)t = \sin\omega_0 t + t\delta\cos\omega_0 t - \frac{1}{2}t^2\delta^2\sin\omega_0 t + \ldots .$$

Therefore, if the approximation procedure is terminated after a finite number of terms, secular terms will inevitably appear. These terms are not physically meaningful because the bound motion of a particle in a conservative one-dimensional potential is periodic. In the case of the simple symmetric potential,

$$V(x) = \frac{1}{2}k_1x^2 + \frac{1}{4}k_3x^4,$$

the secular terms arise even in the first order.

9.4.1 Bogoliuboff–Kryloff Procedure and Removal of Secular Terms

One of the methods for eliminating secular terms in the approximate solutions is the procedure of Bogoliuboff and Kryloff (also known as Linstedt's method). In addition to expanding $x(t)$ in terms of the small parameter α (see Equation 9.48 of Example 9.3), we also expand the frequency ω_0 in terms of the small parameter α:

$$\omega_0^2 = \omega^2 + \alpha\beta_1 + \alpha^2\beta_2 + \ldots = \omega^2 + \sum_{j=1}^{n}\alpha^j\beta_j. \tag{9.48a}$$

The constant β's are to be determined so that the secular terms are eliminated. We now revisit Example 9.3 and apply this procedure to remove the secular terms.

Example 9.4

Use the Bogoliuboff–Kryloff procedure to eliminate the secular terms in the solution $x(t)$ of Example 9.3.

Solution:

We first copy the solution of Example 9.3, Equation 9.61, and label it Equation 9.62:

$$\begin{aligned} x(t) = A\cos\omega_0 t &+ \alpha\frac{A^2}{6\omega_0^2}(\cos 2\omega_0 t - 3) \\ &+ \alpha^2\frac{A^3}{48\omega_0^4}(\cos 3\omega_0 t + 20\omega_0 t\sin\omega_0 t). \end{aligned} \tag{9.62}$$

Next, substituting Equations 9.44 and 9.48a into Equation 9.46 and reducing the coefficients of like powers of α to zero, we have up to the second order in α

$$\ddot{x}_0 + \omega^2 x_0 = 0 \tag{9.63}$$

$$\ddot{x}_1 + \omega^2 x_1 = -x_0^2 - \beta_1 x_0 \tag{9.64}$$

$$\ddot{x}_2 + \omega^2 x_2 = -\beta_2 x_0 - \beta_1 x_1 - 2x_0 x_1. \tag{9.65}$$

The initial conditions are

$$x(0) = A \quad \dot{x}(0) = 0$$

which, in view of Equation 9.44, lead to

$$x_i(0) = A = 0 \quad \dot{x}_i(0) = 0 \quad i = 1,2,3,\ldots.$$

From Equation 9.63, the zero-order solution is

$$x_0 = A\cos \omega t.$$

Substituting this for x_0 in Equation 9.64 yields

$$\begin{aligned}\ddot{x}_1 + \omega^2 x_1 &= -A^2\cos^2\omega t - \beta_1 A\cos\omega t \\ &= -\frac{1}{2}A^2(\cos 2\omega t + 1) - \beta_1 A\cos\omega t.\end{aligned} \tag{9.66}$$

The cosωt term on the right-hand side (which already appears in the complementary solution) generates the troublesome secular term. We now try to eliminate this term by setting the coefficient of the cos(ωt) term equal to zero; hence, $\beta_1 = 0$, and Equation 9.66 then reduces to

$$\ddot{x}_1 + \omega^2 x_1 = -\frac{1}{2}A^2(\cos 2\omega t + 1).$$

Its solution, which satisfies the initial conditions, is

$$x_1(t) = \frac{A^2}{6\omega^2}(\cos 2\omega t - 3) \tag{9.67}$$

and, up to the first order, the frequency ω is equal to ω_0:

$$\omega = \omega_0. \tag{9.68}$$

For the second-order solution, we substitute x_0 and x_1 into Equation 9.65 and obtain

$$\ddot{x}_2 + \omega^2 x_2 = \left(-A\beta_2 + \frac{5A^3}{6\omega^2}\right)\cos\omega t - \frac{A^3}{6\omega^2}\cos 3\omega t \tag{9.69}$$

where the following trigonometric identities have been used:

$$\cos \omega t + \cos 3\omega t = 2\cos 2\omega t \cos \omega t.$$

Again, we require the coefficient of the cosωt term to be zero:

$$-A\beta_2 + \frac{5A^3}{6\omega^2} = 0 \quad \text{or} \quad \beta_2 = \frac{5A^2}{6\omega^2}$$

and Equation 9.69 reduces to

$$\ddot{x}_2 + \omega^2 x_2 = -\frac{A^3}{6\omega^2}\cos 3\omega t.$$

Its solution, which satisfies the initial conditions, is

$$x_2(t) = \frac{A^3}{48\omega^4}\cos 3\omega t. \tag{9.70}$$

The complete solution, up to the second order, is

$$x(t) = A\cos\omega t + \alpha\frac{A^2}{6\omega^2}(\cos 2\omega t - 3) + \alpha^2\frac{A^3}{48\omega^4}\cos 3\omega t. \tag{9.71}$$

It is obvious that the secular term, which is present in Equation 9.62, is absent in Equation 9.71. The frequency is, also up to the second order,

$$\omega_0 = (\omega^2 + \alpha^2 5A^2/6\omega^2)^{1/2}$$

or

$$\omega \cong \omega_0\left(1 + \alpha^2 5A^2/6\omega_0^4\right)^{-1/2}$$

where we have substituted ω_0 for ω in the radical. Expanding the radical yields

$$\omega = \omega_0\left(1 - \alpha^2\frac{5A^2}{12\omega_0^4}\right). \tag{9.72}$$

The Bogoliuboff–Kryloff procedure can be extended to any desired order of approximation.

We can see from Examples 9.3 and 9.4 that the method of perturbation amounts to a series expansion. It works with a nonlinear equation of the general form Equation 9.43. We now summarize, step by step, the method of perturbations:

1. The basic computational need is first to find a dimensionless expansion parameter, say, α, which characterizes the size of the nonlinear term f relative to the linear term $\omega_0^4 x$. Often, we can take α to be the ratio $f/\omega_0^4 x$, evaluated at $x = x_{max}$.
2. Rewrite the equation of motion so that α appears explicitly:

$$\ddot{x} + \omega_0^4 x = \frac{\alpha\omega_0^4 x_{max} f}{f_{max}}$$

   where f_{max} stands for f evaluated at $x = x_{max}$.
3. Consider the nonlinear term as a perturbation and assume that the solution can be expressed as a power series of α:

$$x(t) = \sum_{i=0}^{\infty}\alpha^i x_i(t).$$

4. Insert the power series into the equation of motion and collect terms of like order in α.
5. Reduce each set of terms of like order to zero.
6. Solve the equations sequentially, starting with the equation of zeroth order in α.

If secular terms arise, we can use the Bogoliuboff–Kryloff procedure to remove them. That is, in step 3, in addition to writing the solution $x(t)$ as a power series of α, we also expand the angular frequency ω_0 as a power series of α. Then, in step 4, insert both power series into the equation of motion and collect terms of like order in a. Repeat step 5. In step 6, choose the β_i (i.e., by setting the coefficients of the secular terms to zero) to remove secular terms.

Often, the first correction term in a power series expansion does remarkably well, and the inclusion of the second term usually satisfies all practical needs. If the first few terms do not give an approximation of the desired accuracy, the series converges too slowly, and the method of perturbation is of little use.

9.5 RITZ METHOD

In Chapter 3, we saw that Hamilton's principle is fundamental to the elegant variational formulation of mechanics, and it views the motion of a mechanical system as a whole and involves a search for the path in configuration space that yields a stationary value for the action integral:

$$\delta\int_{t_1}^{t_2} L(q,\dot{q})\mathrm{d}t = 0 \tag{9.73}$$

with $\delta q(t_1) = \delta q(t_2) = 0$. Ordinarily, it is used as a variational method to obtain Lagrange's and Hamilton's equations of motion, so we do not often think of it as a computational tool. But in other areas of physics, variational formulations are used in a much more active way. For example, the variational method for determining approximate ground state energies in quantum mechanics is very well known. In this section, we shall use the Ritz method to illustrate that Hamilton's principle can be used as a computational tool in classical mechanics. The Ritz method is a procedure for obtaining approximate solutions of problems expressed in variational form directly from the variational equation.

The Lagrangian is a function of the generalized coordinates and their time derivatives. The basic idea of the approximation method is to guess a solution for the q's that depends on time and a number of parameters. The parameters are adjusted so that Hamilton's principle, Equation 9.73, is satisfied. The Ritz method takes a special form for the trial solution. A complete set of functions $\{f(t)\}$ is chosen, and the solution is assumed to be a linear combination of a finite number of these functions. The coefficients in this linear combination are the parameters that are chosen to satisfy Hamilton's principle (Equation 9.70). Because the variations of the q's in Equation 9.73 must vanish at the endpoint of the integral, the variations of the parameters must be chosen so that this condition is satisfied.

To summarize, suppose a given system can be described by the following action integral:

$$I = \int_{t_1}^{t_2} L(q,\dot{q},t)\mathrm{d}t. \tag{9.74}$$

The Ritz method requires the selection of a trial solution, ideally in the form

$$q = \sum_{i=1}^{n} a_i f_i(t) \tag{9.75}$$

which satisfies the appropriate conditions at both the initial and final time and where the a's are undetermined constant coefficients and the f's are arbitrarily chosen functions. This trial solution is then substituted into Equation 9.74, and integration is performed so that we obtain an expression for the integral I in terms of the coefficients. The integral I is then made "stationary" with respect to the assumed solution by requiring that

$$\frac{\partial I}{\partial a_i} = 0 \tag{9.76}$$

after which the resulting set of n simultaneous equations is solved for the values of the coefficients a_i. To illustrate this method, we apply it to a simple example, the simple pendulum.

Example 9.5

Use the Ritz method to obtain an approximate solution to the equation of motion of a simple pendulum:

$$\ddot{\theta} + (g/l)\sin\theta = 0. \tag{9.77}$$

Solution:

When θ is small, and if accuracy to the cubic term in θ is required, Equation 9.77 reduces to

$$\ddot{\theta} + k\theta - k\varepsilon\theta^3 = 0 \tag{9.78}$$

where

$$k = g/l,\ \varepsilon = 1/3! = 1/6. \tag{9.79}$$

This can be derived from the Lagrangian L as follows:

$$L = \frac{1}{2}\dot{\theta}^2 - \frac{1}{2}k\theta^2 + \frac{1}{4}k\varepsilon\theta^4. \tag{9.80}$$

The motion is periodic, and we use this fact in choosing t_1 and t_2. It is convenient to choose

$$t_1 = 0,\ t_2 = 2\pi/\omega \text{ (i.e., a full period later).}$$

For the boundary conditions, we choose

$$\theta(0) = \theta(2\pi/\omega) = 0.$$

As a trial solution, we take

$$\theta(t) = a_1 \sin\omega t. \tag{9.81}$$

This satisfies the boundary conditions for any choice of the coefficient a_1. Furthermore, if a_1 is taken as the parameter, the variation of θ at t_1 and t_2 vanishes as required.

The next step is to calculate I, the action integral. First, we introduce the trial function θ and its time derivative into Equation 9.80 and integrate over the time interval (0, 2π/ω):

$$\begin{aligned} I &= \int_{t_1}^{t_2} L\,\mathrm{d}t \\ &= \int_0^{2\pi/\omega} \left[\frac{1}{2}a_1^2(\omega^2\cos^2\omega t - k\sin^2\omega t) + \frac{1}{4}k\varepsilon\, a_1^4 \sin^4\omega t\right]\mathrm{d}t \\ &= \frac{1}{2}\frac{a_1^2\pi}{\omega}(\omega^2 - k) + \frac{3\pi k\varepsilon a_1^4}{16\omega}. \end{aligned} \tag{9.82}$$

We now evaluate a_1 by requiring that

$$\frac{\partial I}{\partial a_1} = \frac{a_1\pi}{\omega}(\omega^2 - k) + \frac{3\pi\, k\varepsilon\, a_1^3}{4\omega} = 0.$$

This relates the amplitude of the oscillation to the frequency. Although the spirit of the approximation says to solve for a_1, it is more natural to solve for ω^2 as a function of the amplitude a_1. We, therefore, have

$$\omega^2 = k(1 - 3\varepsilon\, a_1^2/4). \tag{9.83}$$

For the next approximation, we adopt the following linear combination as a trial function:

$$\theta = a_1 \sin\,\omega t + a_2 \sin\;3\omega t. \tag{9.84}$$

Substitute it and its time derivative into Equation 9.77 and integrate over the time interval (0, $2\pi/\omega$). This leads to an expression for the integral of the system that reads

$$\begin{aligned} I = \frac{\pi\omega}{2\omega}\left[a_1^2 + 9a_2^2\right] - \frac{\pi\, k}{2\omega}\left[a_1^2 + a_2^2\right] \\ - \frac{k\varepsilon\,\pi}{4\omega}\left[\frac{3}{4}a_1^4 - a_1^3 a_2 + 3a_1^2 a_2^2 + \frac{1}{4}a_2^4\right]. \end{aligned} \tag{9.85}$$

From Equation 9.76, we then have

$$\frac{\partial I}{\partial a_1} = -\frac{3k\varepsilon}{4\omega}a_1^3 + \frac{3k\varepsilon}{4\omega}a_3 a_1^2 - \frac{3k\varepsilon}{2\omega}a_2^2 a_1 + (\omega - k/\omega)a_1 = 0 \tag{9.86}$$

$$\frac{\partial I}{\partial a_2} = -\frac{k\varepsilon}{4\omega}a_2^3 - \frac{3k\varepsilon}{2\omega}a_1^2 + \left(9\omega - \frac{k}{\omega}\right)a_2 + \frac{k\varepsilon}{4\omega}a_1^3 = 0 \tag{9.87}$$

and solving these equations in terms of the system parameters completes the problem. Obviously, the pair of Equations 9.86 and 9.87 is not a very easy solution to do by hand. It is necessary to set up a numerical solution by, say, the Newton–Raphson method on a computer. Any improvement in accuracy that we obtain by employing a multiple-term trial solution is at the expense of a substantial increase in computational effort.

9.6 METHOD OF SUCCESSIVE APPROXIMATION

We now consider an approximation method for solving those nonlinear equations that contain time t explicitly. Such systems are called nonautonomous ones. The general method of solution is as follows: Consider a nonlinear equation

$$m\ddot{x} = -kx - F(x) + f\cos\omega t \tag{9.88}$$

where $F(x)$ is a small nonlinear term, and ($f\cos\omega t$) is the driving force of frequency ω. We seek the solution as a series of successive approximations:

$$x(t) = x_1(t) + x_2(t) + x_3(t) + \ldots\,. \tag{9.89}$$

As a first approximation, we choose a simple trial function corresponding to $F(x) = 0$ as $x_1(t)$. We then substitute x_1 into the right-hand side of Equation 9.88 and solve for x_1. In the next approximation, the solution $x_1 + x_2$ is put into the right-hand side of Equation 9.88. This iterative process is repeated as often as desired. As an illustrative example, consider the Duffing's equation, but solve it only as far as the second approximation.

Example 9.6: Duffing's Equation

If we replace $F(x)$ in Equation 9.88 by $-\varepsilon x^3$, the resulting equation is known as the Duffing equation:

$$m\ddot{x} + kx - \varepsilon x^3 = f\cos\omega t \tag{9.90}$$

or

$$\ddot{x} = -\omega_0^2 x + \frac{\varepsilon}{m}x^3 + \frac{f}{m}\cos\omega t. \tag{9.91}$$

Considering that the nonlinear term is small and can be neglected, we obtain

$$\ddot{x}_1 + \omega_0^2 x_1 = \frac{f}{m}\cos\omega t. \tag{9.92}$$

This is a simple harmonic oscillator driven by an external force of frequency ω, and the solution can be written as

$$x_1 = A\cos\omega t \tag{9.93}$$

where, for simplicity, we take the initial phase to be zero. This is the solution of Equation 9.91 to a first approximation. Putting this value of x into the right-hand side of Equation 9.88, we obtain

$$\ddot{x} = \left(\frac{f}{m} - A\omega_0^2 + \frac{3\varepsilon A^3}{4m}\right)\cos\omega t + \frac{\varepsilon A^3}{4m}\cos 3\omega t \tag{9.94}$$

where we have used the trigonometric identity

$$\cos^3\omega t = \frac{3}{4}\cos\omega t + \frac{1}{4}\cos 3\omega t.$$

Integrating twice gives, with the integration constants set equal to zero,

$$x_2 = -\frac{1}{\omega^2}\left(\frac{f}{m} - A\omega_0^2 + \frac{3\varepsilon A^3}{4m}\right)\cos\omega t - \frac{\varepsilon A^3}{36m\omega^2}\cos 3\omega t. \tag{9.95}$$

It is easy to see from the solution 9.92 that the effect of the nonlinear term in Equation 9.90 is to generate a third harmonic of the driving frequency. If we continue the iterative procedure, higher approximations generate higher odd harmonics of ω.

The next step is to again substitute the result (Equation 9.95) into the right-hand side of Equation 9.91. We shall stop at the second approximation. It is easily seen that we may write

$$x_1 + x_2 = A\cos\omega t + B\cos 3\omega t \tag{9.96}$$

and assume that it is the solution of Equation 9.91. Then, substituting into both sides of Equation 9.91, we obtain

$$-A\omega^2\cos\omega t - 9B\omega^2\cos 3\omega t = \frac{f}{m}\cos\omega t - \omega_0^2(A\cos\omega t + B\cos 3\omega t) + \frac{\varepsilon}{m}(A\cos\omega t + B\cos 3\omega t)^3. \tag{9.97}$$

Now, because ε is a small parameter and the term $B\cos 3\omega t$ is a correction of the term $A\cos\omega t$, we may neglect $B\cos 3\omega t$ in the last term of Equation 9.97 without much loss of physics. Then we obtain

$$A\omega^2\cos\omega t + 9B\omega^2\cos 3\omega t = \left(-\frac{f}{m} + A\omega_0^2 - \frac{3\varepsilon}{4m}A^3\right)\cos\omega t + \left(B\omega_0^2 - \frac{\varepsilon}{4m}A^3\right)\cos 3\omega t. \tag{9.98}$$

Comparing the coefficients of cos ωt and cos $3\omega t$, we obtain

$$\frac{f}{m} + \left(\omega^2 - \omega_0^2\right)A + \frac{3\varepsilon}{4m}A^3 = 0 \tag{9.99a}$$

$$B = \frac{\varepsilon A^3}{4m(\omega_0^2 - 9\omega^2)}. \tag{9.99b}$$

Hence, an approximate solution of Equation 9.88 is

$$x(t) = A\cos\omega t + \frac{\varepsilon A^3}{4m(\omega_0^2 - 9\omega^2)}\cos 3\omega t \tag{9.100}$$

where A can be obtained by solving Equation 9.99a, which may be written as

$$\frac{3\varepsilon}{4m\omega_0^2}A^3 = -\frac{f}{m\omega_0^2} + \left(1 - \frac{\omega^2}{\omega_0^2}\right)A. \tag{9.101}$$

It is easy to see that the roots of Equation 9.101 are given by the points of intersection of the cubic parabola

$$y = \frac{3\varepsilon}{4m\omega_0^2}A^3 \tag{9.102a}$$

with the straight line

$$y = \left(1 - \frac{\omega^2}{\omega_0^2}\right)A - \frac{f}{m\omega_0^2}. \tag{9.102b}$$

9.7 MULTIPLE SOLUTIONS AND JUMPS

A nonlinear system may well be capable of oscillating in several different ways even for the same set of parameters. The initial conditions will then determine which type of oscillation is actually observed.

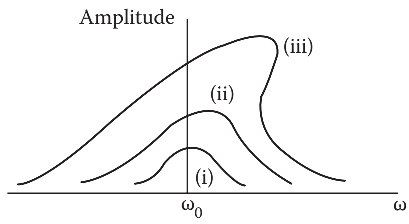


FIGURE 9.12 Sketch of oscillation of the amplitude versus the forcing frequencies ω.

A forced oscillator with damping provides a case in point. Equation 7.67 is a forced linear oscillator with damping from which we will not see multiple solutions and jumps. Now, if the spring constant k depends on x, $k = k(x)$. If the spring behaves the same way in compression as it does in extension, then $k(-x) = -k(x)$, that is, $k(x)$ is an old function of x. If we expand $k(x)$ in a Taylor series about $x = 0$ to two terms, we then obtain $k(x) = \alpha x + \beta x^3$ because there can be no x^2 term. The equation of motion of such a forced linear oscillator with damping then has the form of the Duffing's equation (see Example 9.6):

$$\ddot{x} + r\dot{x} + \alpha x + \beta x^3 = A\cos\omega t$$

where $A\cos\omega t$ is the external driving force, and r is the damping coefficient, which was denoted by 2β in Chapter 8. The coefficient $\beta = k'''(x)/6$ may be positive or negative, depending on the nature of the spring. We will take $\beta > 0$ in what follows.

When the forcing amplitude A is small, so that the response x is also quite small, the cubic term in the Duffing's equation is more or less negligible, and the system is essentially linear. The amplitude of the resulting steady oscillation is then greatest when ω is close to the natural frequency $\omega_0 = (\omega)^{1/2}$, although small damping ensures that the amplitude is finite even when $\omega = \omega_0$ as shown in (i) in Figure 9.12, which is a sketch of oscillation of the amplitude versus the forcing frequencies ω.

At somewhat larger forcing amplitudes A, the cubic term in the Duffing's equation becomes significant, and the response curve develops a pronounced asymmetry about $\omega = \omega_0$ with the strongest response shifting to a higher forcing frequency ω [Figure 9.12, (ii)]. For larger forcing still, the response curve can actually turn over [Figure 9.12, (iii)], and there is then a range of forcing frequencies ω for which three different oscillations—all with frequency ω—are possible.

We now take a closer look at (iii) of Figure 9.13. As ω increases, amplitude $A(\omega)$ increases to its peak until it reaches $\omega = \omega_2$, where the amplitude suddenly decreases by a large factor. As ω decreases from large values, the amplitude slowly increases until $\omega = \omega_1$, where the amplitude suddenly jumps, that is, it approximately doubles (Figure 9.13). The amplitude between ω_1 and ω_2 depends on whether ω is increasing or decreasing (hysteresis effect).

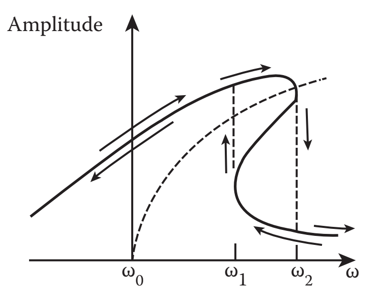


FIGURE 9.13 "Jump" phenomenon.

9.8 CHAOTIC OSCILLATIONS

For almost four centuries, following the lead of Galileo, Newton, and others, physicists have focused on the predictable, effectively linear responses of classical dynamic systems, which usually have linear and nonlinear properties. At the turn of the 20th century, the renowned French mathematician and physicist H. Poincare first recognized the possibility of completely irregular or chaotic behavior of solutions of nonlinear differential equations that are characterized by an extreme sensitivity to initial conditions. Since the rediscovery of this chaotic effect by Lorentz in meteorology in the early 1960s and with the widespread use of computers, the field of nonlinear dynamics has grown tremendously.

The motion of many well-defined nonlinear dynamic systems may not always be predicted far into the future, and they can display highly irregular or seemingly random outputs, so the state of the system is essentially unknown after a very short time. A chaotic system can resemble a stochastic system (i.e., a system subject to random external forces). However, the source of the irregularity for chaos is part of the intrinsic dynamics of the system not unpredictable by outside influences.

Chaos (or, more generally, chaotic dynamics) has many practical applications and arises in celestial mechanics, convection in fluids, oscillations in lasers, heating of plasmas by electromagnetic waves, the determinations of the limits of weather forecasting, chemical reactions, neural networks, biological rhythms, and many other examples. The study of chaotic phenomena is interdisciplinary and has brought together scientists from physics, mathematics, chemistry, meteorology, medicine, biology, and other disciplines.

This section is intended only to help the reader develop an ability to comprehend and judge chaotic motion. To get a feel for chaotic motion, let us revisit the van der Pol's equation. We saw in Section 9.1 that the original van der Pol oscillator is self-limiting and does not have chaotic solutions. Now, we impose a harmonic disturbance, say, $f\cos\omega t$, upon the van der Pol oscillator:

$$\ddot{x} - \varepsilon(1 - x^2)\dot{x} + x = f\cos\omega t \tag{9.103}$$

where f is the amplitude of the external periodic force and ω is the driving frequency. The behavior of this periodically forced van der Pol oscillator is quite different. For example, $f = 5$, $\varepsilon = 5$, and $\omega = 2.466$, which lead to a bounded non-periodic oscillation as depicted in Figure 9.14 (Steeb et al. 1989). The irregular and unpredictable time evolution of such chaotic behavior is different from random behavior. Randomness is reserved for problems in which we do not know the input force exactly or we only know some statistical measures of the parameters. Radioactive nuclei behave randomly, and the throw of a die is random. However, there is no element of randomness in the path of a chaotic system—each time step depends only on the motion of previous times in a well-defined way. But very small differences in initial conditions can lead to enormously different results. Two paths that are initially arbitrarily close in phase space will diverge exponentially in time from each other. (The well-known Lyapunov exponents characterize the rate of this divergence.)

9.8.1 Some Helpful Tools for an Understanding of Chaos

Toward an understanding of chaos, a variety of analytic and computational tools have been developed. Phase space, the Poincare section, the power spectrum, and the logistic map are just a few.

We have seen the usefulness of phase space. Each phase point represents a possible state of a dynamic system. As the system changes over time, the phase point traces out a path or trajectory in phase space. The paths of a deterministic dynamic system corresponding to similar energies will pass very close to each other but never intersect. This non-crossing property of the phase paths derives from the fact that past and future states of a deterministic dynamic system are uniquely prescribed by the system state at a given time. Another important feature of the phase space is that the phase volume of a conservative dynamic system is preserved as time passes (the Liouville theorem). A nonconservative dynamic system does not have such a phase volume preservation property.

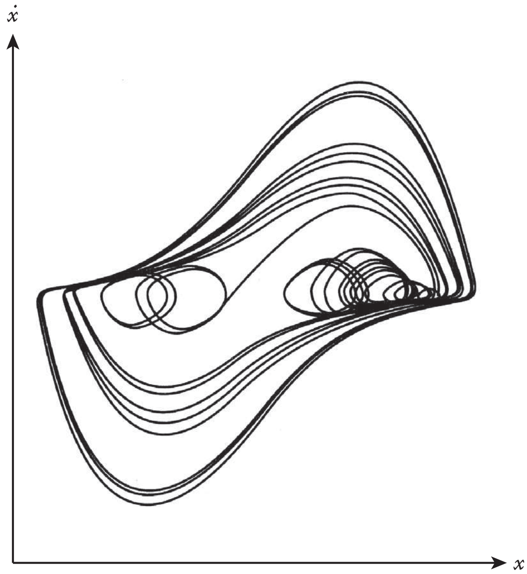


FIGURE 9.14 Phase diagram of a forced Van der Pol oscillator.

Another important difference between conservative and nonconservative dynamic systems is that the latter have attractors while the former does not.

An attractor is a set of points in phase space toward which a dynamic system evolves as time t gets very large. This might be a fixed point, as for the damped oscillator (Figures 7.14 and 7.15); a limit cycle, as for the van der Pol oscillator (Figure 9.9); or an extremely complex set of points. Nonlinear dynamic systems may have chaotic attractors, which are sometimes called strange and are examples of fractals.

The Poincare section is a snapshot of the motion in the phase space, taken at regular time intervals. This way of studying chaotic motion is better than watching a trajectory wander around in phase space. To get a good grasp on this concept, we go back to the simple pendulum swinging back and forth in a small arc near the bottom of its swing. In phase space, this motion appears to generate an ellipse around the stable point $(\dot{\theta},\theta)$ as shown in Figure 9.7, where the phase diagram is a set of points $(\dot{\theta}_0,\ \theta_0)$, $(\dot{\theta}_1,\ \theta_1)$, $(\dot{\theta}_2,\ \theta_2)$,... generated by solving the differential equation that describes the system. Now, if we add a time axis perpendicular to the $\dot{\theta}-\theta$ coordinates, we will generate a curve in $\dot{\theta}-\theta-t$ space. If the orbit is a circle in the $\dot{\theta}-\theta$ plane, it traces out a helix in $\dot{\theta}-\theta-t$ space as time progresses. For an elliptical orbit, as for our simple pendulum, the path in $\dot{\theta}-\theta-t$ space is a squashed helix. Imagine that time, like a clock, runs in a circle, and we choose the circumference of the time circle to be exactly one period of our dynamic system. For our simple pendulum, the t-axis completes its circle when $t = 2\pi\sqrt{L/g}$, and we choose $L/g = 1$, giving it a period of $T = 2\pi$. If we examine the motion of the simple pendulum on the screen of a video monitor, we can take the screen as the $\dot{\theta}-\theta$ plane. Furthermore, we can imagine the time axis poking straight out of the screen from the point (0,0), circling around the monitor, and piercing the screen exactly at the same point (0,0) as illustrated in Figure 9.15. The trajectory of the pendulum in $\dot{\theta}-\theta-t$ space will pierce the monitor screen at the same $(\dot{\theta},\ \theta)$ point each time the pendulum completes its cycle. We can plot a $(\dot{\theta},\ \theta)$ point at $t = 0$, $t = 2$, $t = 4$, and so on. And we call the set of these $(\dot{\theta},\ \theta,\ 2n\pi)$ points a Poincare map. It is obvious that the Poincare map for a simple pendulum is a single point. If we add damping to the simple pendulum, the Poincare section will be a series of points that leads to the attractor at (0,0). Now, if a driving force is added, then the natural period of the pendulum is no longer very important. In this case, it is convenient to choose the period of the driving force as the circumference of

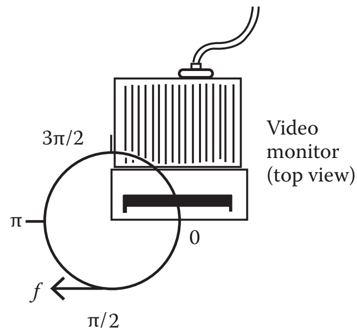


FIGURE 9.15 Video monitor.

the time axis. For example, if the driving force is sinusoidal with angular frequency $F = F_0 \cos\omega t$, then the period T of the driving force is given by $T = 2\pi/\omega$. If we want n points per cycle of the driving force, we can set the step size h with the formula $h = 2\pi/(n\omega)$.

The power spectrum is computed using Fourier analysis to display the frequency composition of time variation of the dynamic variables.

Example 9.7

Find the Poincare map of period 2π for the solution of the following equation:

$$64\ddot{x} + 16\dot{x} + 65x = 64\cos t \tag{9.104}$$

which starts from (0,0) at time $t = 0$.

Solution:

The general solution for Equation 9.104 has two parts:

$$x(t) = x_c + x_p$$

where x_p is a particular integral to Equation 9.103, and x_c is the complementary solution that is the general solution to the corresponding homogeneous equation:

$$64\ddot{x} + 16\dot{x} + 65x = 0. \tag{9.105}$$

A solution of the form exp(pt) always exists for Equation 9.105, so we write $x_c = A\exp(pt)$. The characteristic equation is

$$64p^2 + 16p + 65x = 0$$

which has two roots

$$p_1 = -1/10 + i \text{ and } p_2 = -1/10 - i.$$

Thus, we have

$$x_c = A_1\exp(p_1 t) + A_2\exp(p_2 t) = \exp(-1/10)(C\cos t + D\sin t).$$

To find the particular integral (which represents the forced periodic response), we write

$$x_p = A\cos t + B\sin t.$$

Substituting this into Equation 9.104, we obtain

$$64\ddot{x} + 16\dot{x} + 65x - 64\cos t = (A + 16B - 64)\cos t + (B - A)\sin t = 0$$

from which we have

$$A + 16B - 64 = 0 \text{ and } B - 16A = 0.$$

Solving for A and B, we find $A = 0.249$ and $B = 3.9104$, and the general solution $x(t)$ is

$$x(t) = \exp(-t/8)(C\cos t + D\sin t) + A\cos t + B\sin t.$$

Thus,

$$\dot{x}(t) = \exp(-t/8)(-C\sin t + D\cos t) - (1/8)\exp(-t/8)(C\cos t + D\sin t) + A\cos t + B\sin t.$$

The initial conditions imply that $C = -0.249 = -A$ and $D = -4.015$. The Poincare map of this solution is given by (Figure 9.16)

$$x_n = A\left[1 - \exp\left(-\frac{1}{4}n\pi\right)\right], \quad y_n = \dot{x}_n = B\left[1 - \exp\left(-\frac{1}{4}n\pi\right)\right], \quad n = 1, 2, \ldots.$$

As $n \to \infty$, the sequence of the dots approaches the fixed point at $(A, B) = (0.249, 3.9104)$.

Chaotic motions of even a relatively simple dynamic system are governed by nonlinear differential equations. Mathematically and computationally, nonlinear differential equations are difficult to solve. Fortunately, even an elementary model system can give insight into the mechanisms leading to chaotic behavior. These are stated in the form of difference equations, and a typical difference equation is of the form

$$x_{n+1} = f(\alpha, x_n) \tag{9.106}$$

where α is a model-dependent parameter, n denotes the time sequence of a system, and x denotes a physical observable of the system; thus, x_n refers to the nth value of x, and it is a real number restricted in some real number interval. The relationship 9.106 describes the progression of a nonlinear system at a particular moment by investigating how the $(n+1)$th state (or iterate) depends on the nth state. An example of such a simple, nonlinear behavior is $x_{n+1} = (2x_n + 3)^2$.

The relationship 9.106 is called mapping. The function $f(\alpha, x_n)$ generates the value of x_{n+1} from x_n, and the collection of points generated is said to be a map of the function itself (known as a logistic map). The equations, which are often nonlinear, are amenable to numerical solution by iteration, starting with x_1. We will restrict ourselves here to one-dimensional maps, but two-dimensional and higher-order equations are possible.

Mapping can best be understood by looking at an example. Let us consider the logistic equation, a simple one-dimensional equation given by

$$f(\alpha, x) = \alpha x(1-x), \; 0 \le \alpha \le 4$$

so that the iterative (or difference) equation becomes

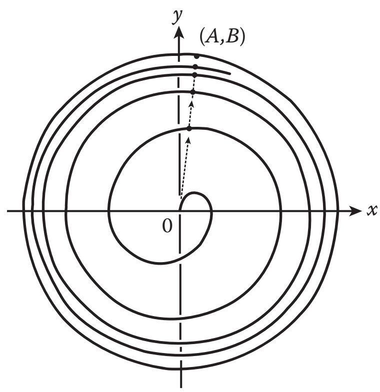


FIGURE 9.16 Poincare map.

$$x_{n+1} = \alpha x_n(1 - x_n),\ 0 \le \alpha \le 4. \quad (9.107)$$

This nonlinear, one-dimensional iteration equation was first used by P.F. Verhulst in 1845 to model the development of a breeding population whose generations do not overlap, and it was then used by T. Bessoir and A. Wolf in 1991 for a biological application example of studying the population growth of fish in a pond. In a rather qualitative sense, the sample logistic equation 9.107 is representative of many dynamic systems in biology, chemistry, and physics.

Because $0 \le \alpha \le 4$, $0 \le x_n \le 1$ implies $0 \le x_{n+1} \le 1$, and so we can assume that x_n is restricted in the interval [0,1]. The quadratic function $f(\alpha,x) = \alpha x(1-x)$ has one maximum in the interval [0,1] and is zero at the endpoints $f(\alpha,0) = 0 = f(\alpha,1)$. The maximum at $x_m = ½$ is determined from $f'(\alpha, x_m) = \alpha(1 - 2x_m) = 0$, where $f(\alpha,½) = \alpha/4$.

The results of the logistic equation are most easily observed by graphic means. The iteration x_{n+1} is plotted versus x_n in Figure 9.17 for a value of $\alpha = 2.0$. Starting with an initial value x_1 on the horizontal (x_n) axis, we move up until we intersect with the curve $x_{n+1} = 2x_n(1 - x_n)$ and then move to the left where we have x_2 on the vertical axis (x_{n+1}). We then start with this value of x_2 on the horizontal axis and repeat the process to find x_3 on the vertical axis. This process can be repeated for more iteration. We can make our job easier by adding the 45° line ($x_{n+1} = x_n$) to the same graph (Figure 9.18). Then, after initially intersecting the curve from x_1, we move horizontally to intersect with the 45° line to find x_2 and then move up vertically to find the next iterative value of x_3. This process can go on and reach the same result as in Figure 9.15.

We see from Figure 9.15 or 9.16 that x_i converges toward the fixed point (0.5, 0.5). At the fixed point x^*, which is an attractor, the iteration stops so that

$$f(\alpha,x^*) = x^*(1 - x^*) = x^*, \text{ that is, } x^* = 1 - 1/\alpha.$$

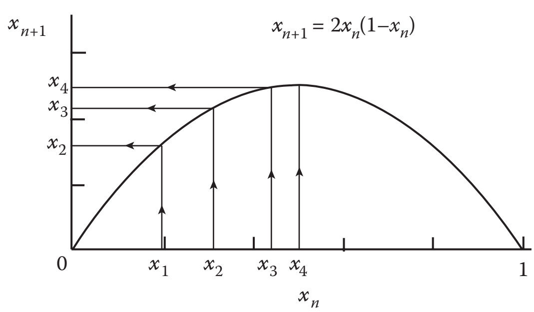


FIGURE 9.17 Logistic map.

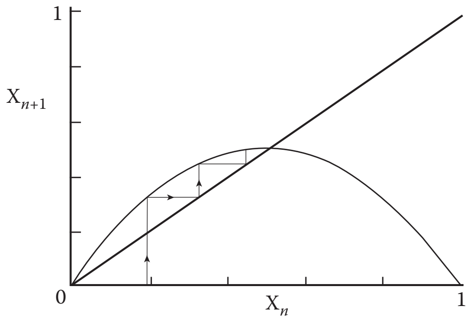


FIGURE 9.18 Logistic map of Figure 9.18, with the 45° line added.

That is, the fixed point is the intersection of the quadratic map function with the 45° line (the line $x_{n+1} = x_n$). Now, for $\alpha = 2$, $x^* = 0.5$.

For any initial value x_1 satisfying $f(\alpha, x_1) < x^*$, or $0 < x_1 < 1/\alpha$, x_i converge to the attractor x^*; the interval $[0, 1/\alpha]$ defines a basin of attraction for the fixed point x^*. Note that the fixed point $x^* = 1 - 1/\alpha$ is in the interval [0,1] only when $\alpha \geq 1$. For $\alpha = 2$, $x^* = 0.5$.

In practice, we want to study the behavior of the system when the model parameter α is varied. The attractor x^* is stable provided the slope is

$$|df(\alpha,x^*)/dx^*| < 1 \text{ or } \alpha(1 - 2x^*) < 1.$$

Accordingly, the fixed point $x^* = 1 - 1/\alpha$ is stable when $1 < \alpha < 3$.

We next explore the region $3 < \alpha < 4$. There are no stable fixed points when $\alpha > 3$; we consider periodic points. A point x_1 is defined as a periodic point of period n for $f(\alpha,x)$ if $f^{(n)}(x_1) = x_1$ but $f^{(i)}(x_1) \neq x_1$ for $1 < i < n$, where we abbreviate $f^{(1)}(x) = f(\alpha, x)$, $f^{(2)}(x) = f(\alpha,f(\alpha,x))$ for the second iterate, etc. For the fixed points x_2^*, we have

$$x_2^* = f(\alpha, f(\alpha, x_2^*)$$

or

$$x_2^* = \alpha^2 x_2^*\left[1 - \alpha x_2^*(1 - x_2^*)\right]$$

and we see that there are three fixed points; each one is a point of period 2. We can establish that one is unstable and two are stable. We see that as the value of α is raised past a critical value of 3, the transition from one stable fixed point to a pair of stable period-2 points occurs. This transition is known as a bifurcation or period doubling. It can be proved that for $\alpha \approx 3.45$, each branch of fixed points bifurcates again so that $x_4^* = f^{(4)}(x_4^*)$, that is, has a period 4. This increasing period doubling is one of the roads to chaos (Figure 9.19).

9.8.2 Conditions for Chaos

Chaotic motion is not a very uncommon phenomenon. There are two necessary conditions for chaotic motion to occur:

(1) The associated phase space of the system must be at least three dimensional.
(2) The equations of motion contain a nonlinear term that couples several of the variables.

We can state the first condition for chaos to occur in another way: Chaos cannot occur in a two-dimensional, autonomous system

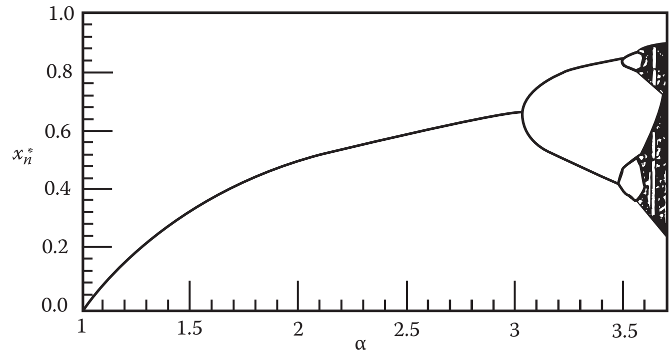


FIGURE 9.19 Bifurcation diagram of the logistic map.

$$\dot{x} = f(x, y), \quad \dot{y} = g(x, y). \tag{9.108}$$

Now suppose that we have determined the equilibrium points (if any) of the system (Equation 9.108), which are the points at which both $f(x,y) = 0$ and $g(x,y) = 0$. Next, suppose that a phase path starts at some point and cannot leave a certain bounded region of the phase plane $(\dot{x}, x)$. Poincare and Bendixon show that the phase path must eventually (1) terminate at an equilibrium; (2) return to the original point, giving a closed path; or (3) approach a limit cycle. This is known as the Poincare–Bendixon theorem. Thus, according to the Poincare–Bendixon theorem, chaotic motion is ruled out.

We can understand the fact that three variables are required for chaos to occur: Three-space is sufficient to allow for (1) divergence of trajectories, (2) confinement of the motion to a finite region of the phase space of the dynamic variables, and (3) uniqueness of the trajectory.

Now let us go back to Equation 9.103, which is of the second order and not autonomous. We can recast it in the form of Equation 9.106, an autonomous first-order system:

$$\dot{x} = y, \quad \dot{y} = \varepsilon(1 - x^2)y - x + f\cos\omega t, \quad \dot{t} = 1.$$

We see that the associated phase space is three-dimensional. Note also that we are seeing in Figure 9.12 only the projection onto the x, y-plane of a chaotic phase path that rises up out of the page, that is, in the positive t-direction, and never, in fact, crosses itself at all.

9.8.3 Routes to Chaos

It is obvious that the above two necessary conditions do not guarantee chaos; they only make its existence possible. Routes to chaos are important for greater understanding of chaos and for practical purposes. Identifying pre-chaotic patterns or behavior might help us anticipate the occurrence of chaos. All the possible routes to chaos probably have not yet been discovered. The methods that have received most of the attention in the literature thus far are the following three: period doubling, intermittency, and quasi-periodicity.

Period doubling is a systematic cascading progression to chaos and is the most extensively studied type of transition.

Intermittency consists of orderly periodic motion (regular oscillations with no period doubling) interrupted by occasional bursts of chaos or noise at irregular intervals (Figure 9.20). In mathematical modeling, the periodic motion (limit cycle) typically shows up under relatively low values of the control parameter. Gradually increasing the control parameter brings infrequent chaotic bursts in the time series. These bursts set in abruptly, rather than gradually. With further increase of the

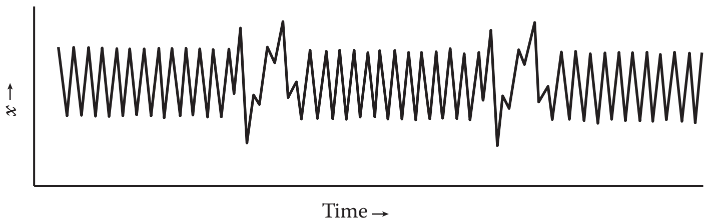


FIGURE 9.20 Orderly periodic motion interrupted by occasional bursts of chaos.

control parameter, chaotic bursts are more frequent and last longer until the pattern eventually becomes completely chaotic.

Quasi-periodicity is motion or behavior caused by two or more simultaneous periodicities whose different frequencies are out of phase (not commensurate) with one another. Because the frequencies are independent and lack a common denominator, the motion never repeats itself exactly. However, it can almost repeat itself or seem, at first glance, to repeat itself. Hence we dub it quasi-periodic. It does not show up readily on a time-series graph and usually requires more sophisticated mathematical techniques to be seen. The periodicities characterize trajectories that trace out a torus or doughnut in phase space.

9.8.4 Lyapunov Exponentials

The sensitive dependence on initial conditions is a characteristic of chaotic behavior. One method to quantify this sensitive dependence uses the Lyapunov characteristic exponent (named after A.M. Lyapunov, 1857–1918, a Russian mathematician). We can compute this exponent easily for a one-dimensional map, such as the logistic map. Now, if a system is allowed to evolve from two slightly differing initial states, x and $x + \varepsilon$, we want to know the difference between the eventual values of x_n after n iterations from the two initial values. The Lyapunov exponential λ represents the coefficient of the average exponential growth per unit time between the two states. After n iterations, the difference d_n between the two x_n values is approximately

$$d_n \approx \varepsilon e^{\lambda n}.$$

We see that if λ is negative, slightly separate trajectories converge and the evolution is not chaotic, but if λ is positive, nearby trajectories diverge and chaotic motion results.

Let us consider a specific one-dimensional map given by $x_{n+1} = f(x_n)$. The difference between two initially nearby states after the nth step is given by

$$f^n(x + \varepsilon) - f^n(x) \approx \varepsilon e^{n\lambda}$$

or

$$\log_e\left[\frac{f^n(x+\varepsilon) - f^n(x)}{\varepsilon}\right] \approx n\lambda.$$

For small ε, this expression becomes

$$\lambda \approx \frac{1}{n}\log_e\left|\frac{\mathrm{d}f^{(n)}}{\mathrm{d}x}\right|.$$

The value of $f^n(x_0)$ is obtained by iterating the function $f(x_0)$ n times:

$$f^n(x_0) = f(\ldots(f(x_0)\ldots).$$

We use the chain rule for the derivative of the nth iterate and obtain

$$\left.\frac{\mathrm{d}f^n(x)}{\mathrm{d}x}\right|_{x_0} = \left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_n-1}\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_n-2}\cdots\left.\frac{\mathrm{d}f}{\mathrm{d}x}\right|_{x_0}.$$

We take the limit as $n \to \infty$ and finally obtain

$$\lambda = \lim_{n\to\infty}\frac{1}{n}\sum_{i=0}^{n-1}\log_e\left|\frac{\mathrm{d}f(x_i)}{\mathrm{d}x}\right|.$$

For n-dimensional maps, there are n Lyapunov exponents. Only one of them need be positive for chaos to occur.

PROBLEMS

1. Show that the superposition principle does not hold for a nonlinear differential equation, such as

$$\frac{\mathrm{d}^2x}{\mathrm{d}t^2} + cx^2 = 0$$

   where c is a constant.
2. Consider the system in Figure 9.21. Assume that the tension T in the wire is linear with its elongation δ. If the initial tension is T_0, derive the equation of motion and linearize the problem.
3. Construct a phase diagram for the potential in Figure 2.12.

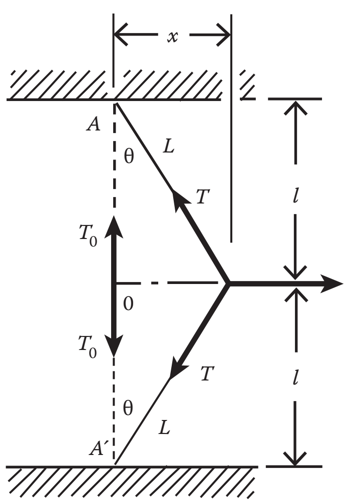


FIGURE 9.21 Nonlinear system.

4. If the amplitude of the motion of a pendulum is not small, the free horizontal motion may be represented approximately by

$$\ddot{x} + \frac{g}{b}x - \frac{g}{2b^3}x^3 = 0$$

where the notation is that of Figure 7.2. Verify this equation. If a force $F_0 \cos \omega t$ is also acting, find an approximate steady-state expression for $x(t)$, considering $(g/b)x^3$ to be small.

5. Lord Rayleigh used the equation

$$\ddot{x} - (a - b\dot{x}^2)\dot{x} + \omega_0^2 x = 0$$

in his discussion of nonlinear effects in acoustic phenomena. Show that this equation can be converted to the following equation (van der Pol's equation):

$$\ddot{y} - \frac{a}{y_0^2}(y_0^2 - y^2)\dot{y} + \omega_0^2 y = 0 \quad \text{with} \quad y = y_0\sqrt{3b/a}\ \dot{x}.$$

6. Consider the equation of motion of a particle

$$\ddot{x} + \alpha x + \gamma x^3 = 0; \quad \alpha > 0, \quad \gamma > 0.$$

Sketch the potential function. Find a Fourier series solution for the case where energy has a small positive value.

7. If a simple pendulum is oscillating in a medium and experiences a resistive force proportional to the velocity, then its equation of motion becomes

$$\ddot{\theta} + \lambda\dot{\theta} + \frac{g}{b}\sin\theta = 0$$

where b is the length of the pendulum and λ is the coefficient of friction. Discuss the stability of such a pendulum near its equilibrium point.

8. Show that a secular term arises in the first-order solution to the equation

$$\ddot{x} + \omega_0^2 x = \lambda x^2$$

where λ is a small quantity. Remove the secular term from the solution by using the Bogoliuboff–Kryloff procedure.

9. Consider the motion of the anharmonic oscillator whose potential energy is

$$V(x) = \frac{1}{2}kx^2 + \frac{1}{3}m\lambda x^3$$

under the combined action of the two sinusoidal driving forces

$$F = F_1 \cos \omega_1 t + F_2 \cos \omega_2 t.$$

Show that the solution of the equation of motion contains sinusoidal terms whose frequencies are equal to the sum or difference of the driving frequencies. This phenomenon is known as intermodulation. What are possible values of ω_1 and ω_2 for resonance?

10. Solve the following equation of motion for an anharmonic oscillator

$$m\ddot{x} + kx - \lambda x^2 = 0$$

by the method of successive approximations and the method of perturbations.
11. Consider the system in Figure 9.21 again. Assume a periodic force $F_0\cos\omega t$ is acting on the mass. Gravity is neglected. Find the equation of motion and solve by the method of successive approximation or the method of perturbation.
12. Use the Ritz method to analyze the forced oscillations of a harmonic oscillator incorporating a nonlinear spring whose deflection characteristic is cubic (Figure 9.22):

$$m\ddot{y} + ky + \lambda y^3 = F_0 \sin \omega t.$$

13. A particle of mass m is placed on the inside of a smooth paraboloid of revolution having the equation $cz = x^2 + y^2$ at a point P, which is at height h above the horizontal plane (the x,y-plane). Assuming that the particle starts from rest, find
    (a) the speed of the particle at the vertex 0
    (b) the time taken to reach 0
    (c) the period for small oscillations

    If the oscillations are not small, show that the period is given by (Figure 9.23)

$$\tau = 2\pi \frac{c+4h}{2g}\{1-(1/2)^2 k^2 - (1\cdot 3/2\cdot 4)^2 k^4/3$$
$$- (1\cdot 3\cdot 5/2\cdot 4\cdot 6)^2 k^6/6 \ + \ldots\}.$$

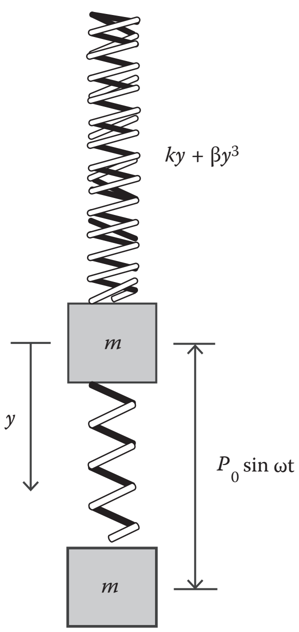


FIGURE 9.22 Forced oscillations of a harmonic oscillator incorporating a nonlinear spring.

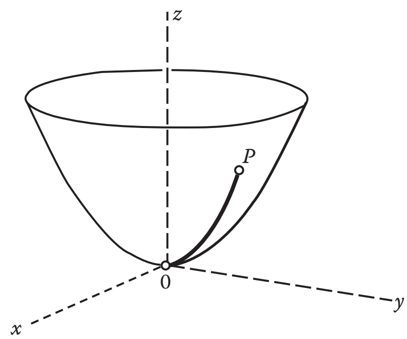


FIGURE 9.23 Particle placed on the inside of a smooth paraboloid of revolution.

14. A very important type of nonlinear equation was extensively studied by van der Pol in connection with relaxation oscillations in vacuum tube circuits. The van der Pol equation may be written as

$$\ddot{x} - \varepsilon(1 - x^2)\dot{x} + x = 0$$

where ε is a small, positive parameter. Solve the van der Pol equation by the perturbation method.
15. Consider the system

$$\dot{x} = y + \varepsilon x(1 - x^2 - y^2), \quad \dot{y} = -x + \varepsilon y(1 - x^2 - y^2)$$

where ε is a positive constant. Prove that it has a simple limit cycle by changing to the variables r, θ, where $x = r\cos\theta$ and $y = r\sin\theta$.
16. Can chaos occur in a one-dimensional nonautonomous system $\dot{x} = f(x,t)$?
17. For any mapping function $x_{n+1} = f(x_n)$, show that a fixed point x^* will be stable if

$$\left|f'(x^*)\right| < 1, \text{ where } f'(x^*) \equiv \left.\frac{df}{dx}\right|_{x^*}.$$

18. For the iterated quadratic map $x_{n+2} = f(f(x_n)) \equiv f_2(x_n)$, show that its fixed points satisfy

$$(x^*)^3 - 2(x^*)^2 + \frac{\alpha+1}{\alpha}x^* - \frac{\alpha^2-1}{\alpha} = 0.$$

REFERENCES

Steeb, W.H., Huang, J.C., and Gou, Y.S. A comment on the chaotic behavior of Van del Pol equation with an external periodic excitation, *Z. Naturfors.*, 44a, 160, 1989.

van der Pol, B. On relaxation-oscillations, *The London, Edinburgh and Dublin Phil. Mag. & J. of Sci.*, 2(7), 978–992, 1927.

10 Collisions and Scatterings

Information and knowledge in the domain of microscopic physics are obtained by two means: spectroscopy and collisions or scattering. In spectroscopy, we analyze the electromagnetic radiation emitted or absorbed in the transitions between bound states of the systems. In scattering experiments, we probe the interaction between various particles by performing experiments in which one particle is scattered by another. For example, it was Rutherford's experiments on the scattering of alpha particles that provided the physical basis for the nuclear atom model and for quantum mechanics. Many of the features of the nuclear force, such as its range, strength, and spin dependence, have been deduced from data gathered from nucleon–nucleon scattering. Scattering of electrons from nuclei as well as nucleons has helped in determining the charge distribution in the latter. Scattering experiments have become essential to the exploration of the interaction between the various elementary particles. In this chapter, we will study the collision problems and the classical description of scattering.

For a two-particle collision or scattering problem, the motion of one particle relative to the other can be completely solved once the force law is known. When the nature of the interaction or force is not known, important results can still be obtained by using the laws of conservation of momentum and energy. In collisions or scatterings where no external forces are involved and internal forces satisfy Newton's third law, the total linear momentum of the system is conserved as we have already seen in Chapter 2. The conservation of momentum is a fundamental law and is a Galilean invariant (i.e., it holds in all inertial frames). If the kinetic energy is also conserved in collisions, such collisions are called elastic collisions. In inelastic collisions, kinetic energy is not conserved. If $\vec{P}$ and $\vec{K}$ are the initial momentum and kinetic energy before the collision, and $\vec{P}'$ and $\vec{K}'$ are the final momentum and kinetic energy after collision, then for

elastic collisions: $\vec{P} = \vec{P}'$ and $\vec{K} = \vec{K}'$
inelastic collisions: $\vec{P} = \vec{P}'$ and $\vec{K} \neq \vec{K}'$.

It should be emphasized that conservation of kinetic energy alone cannot guarantee that the collision is an elastic one. Why? Because conservation of kinetic energy is not a Galilean invariant; that is, it can hold relative to one inertial frame without holding relative to other inertial frames. However, if the momentum is also conserved, then the kinetic energy is conserved in all inertial frames if it is conserved in any one inertial frame.

10.1 DIRECT IMPACT OF TWO PARTICLES

Consider a collision between two moving bodies, such as two smooth spheres of masses m_1 and m_2. We limit the collision to a direct head-on impact. That is, the velocity vectors of the approaching masses lie along the line of impact, shown in Figure 10.1. In this chapter, we limit ourselves to collisions involving particles. In the present example of the sphere, rotational motions are not introduced, and the problem is essentially that of the collision of two particles. Rotational motions of extended bodies will be examined in Chapter 12.

The momentum conservation holds, and so we have

$$\vec{P}_1 + \vec{P}_2 = \vec{P}_1' + \vec{P}_2'$$

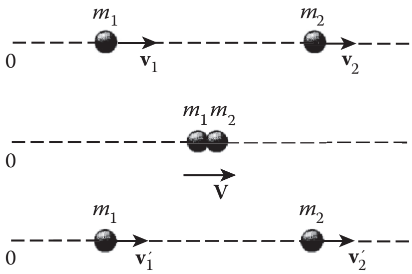


FIGURE 10.1 Head-on collision of two particles.

or

$$m_1\vec{v}_1 + m_2\vec{v}_2 = m_1\vec{v}_1' + m_2\vec{v}_2' \tag{10.1}$$

where the subscripts 1 and 2 refer to the two bodies, and the primes indicate the respective momenta and velocities after collision.

In practice, all colliding bodies have elasticity. As the bodies come into contact, each deforms until maximum deformation occurs. From the maximum deformation to the point of just separating is the restitution period. The total time of impact—deformation plus restitution—is quite short. The instant of maximum deformation occurs when both bodies are traveling with the same velocity V. Applying the impulse-momentum equation to m_1 for the deformation period, we have

$$I_{d1} = m_1V - m_1v_1 \tag{10.2}$$

where I_{d1} represents the linear impulse of m_1 during deformation. Similarly, for m_2

$$I_{d2} = m_2V - m_2v_2 - I_{d1}. \tag{10.3}$$

Solving Equations 10.2 and 10.3 for v_1 and v_2, we obtain

$$v_1 = -\frac{I_{d1}}{m_1} + V \quad \text{and} \quad v_2 = \frac{I_{d2}}{m_2} + V \tag{10.4}$$

and so

$$v_2 - v_1 = I_{d1}\left(\frac{1}{m_1} + \frac{1}{m_2}\right) = I_{d1}\frac{m_1 + m_2}{m_1m_2}. \tag{10.5}$$

We next apply the impulse-momentum equation to m_1 and m_2 for the restitution period. Each mass possesses velocity V at the beginning of this period and individual velocities v_1' and v_2' at its conclusion. For m_1, we have

$$I_{r1} = m_1v_1' - m_1V \tag{10.6}$$

where I_{r1} represents the impulse during the restitution for m_1. For m_2, we have

$$I_{r2} = m_2v_2' - m_2V = -I_{r1}. \tag{10.7}$$

Solving Equations 10.6 and 10.7 for v_1' and v_2', we obtain

$$v_1' = \frac{I_{r1}}{m_1} + V \quad \text{and} \quad v_2' = -\frac{I_{r2}}{m_2} + V \tag{10.8}$$

and so

$$v_2' - v_1' = -I_{r1}\frac{m_1 + m_2}{m_1 m_2}. \tag{10.9}$$

The relationship between the impulse of deformation I_{d1} and the impulse of restitution I_{r1} depends on many factors, such as geometry, material properties, and velocity. Experiments tell us that they are related by the equation

$$I_{r1}/I_{d1} = e \tag{10.10}$$

where e, called the coefficient of restitution (or, sometimes, the coefficient of elasticity), is experimentally determined and is approximately a constant for a given pair of colliding substances. Dividing Equation 10.9 by Equation 10.5, we obtain, with the help of Equation 10.10,

$$e = -\frac{v_2' - v_1'}{v_2 - v_1} = -\frac{\text{velocity of separation}}{\text{velocity of approach}}. \tag{10.11}$$

This result is often called Newton's collision rule.

For elastic collisions, the system's kinetic energy is conserved:

$$\frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 = \frac{1}{2}m_1 v_1'^2 + \frac{1}{2}m_2 v_2'^2.$$

The last equation can be rewritten as, after algebraic manipulation,

$$m_1(v_1 + v_1')(v_1 - v_1') = m_2(v_2 + v)(v_2 - v_2'). \tag{10.12}$$

Dividing Equation 10.12 by Equation 10.1,

$$v_1 + v_1' = v_2 + v_2'$$

or

$$v_2' - v_1' = -(v_2 - v_1). \tag{10.13}$$

Comparing this with Equation 10.11, we see that the coefficient of restitution e is equal to unity for an elastic collision.

For direct central collisions other than elastic, $e < 1$, there is a loss of kinetic energy associated with the collision, so an energy-loss term must be added to the kinetic energy equation:

$$\frac{1}{2}m_1 v_1^2 + \frac{1}{2}m_2 v_2^2 = \frac{1}{2}m_1 v_1'^2 + \frac{1}{2}m_2 v_2'^2 + Q.$$

Here, the quantity Q represents the net loss or gain in kinetic energy, which occurs during the collision. If there is an energy loss, then Q is positive. This is called an exoergic collision. If there is an energy gain, Q is negative. This collision is called an endoergic collision. This would occur, for example, if an explosive were present on one of the bodies at the point of contact.

When the collision is perfectly inelastic, the velocity of separation is zero; the two bodies stick together and move with the same final velocity. For this case, $e = 0$, which can be confirmed by substituting $v_2' = v_1'$ into Equation 10.11.

In collision problems of the type discussed, the two final velocities are usually unknown, and two independent equations are therefore required. One such equation is provided by momentum conservation, Equation 10.1. For an elastic collision, the second equation is provided by the energy conservation, Equation 10.12 or its equivalent, Equation 10.11, with $e = 1$. For cases in which e lies between zero and unity, Equation 10.11 is used.

So far, our discussion has been limited to head-on collisions in which the motion is confined to a single straight line. In practice, most collisions occur obliquely; that is, the velocity vectors of the approaching body do not lie along the line that connects the two mass centers (the mass-center line). In this general case, the motion is no longer confined to a single straight line; the vector form of the momentum equations must be employed. It is found empirically that Newton's collision rule, Equation 10.11, applies to the velocity components that lie along the mass-center line. The following two examples illustrate these points.

Example 10.1

A particle of mass m_1 with initial velocity V strikes a particle of mass m_2 that is initially at rest. If the collision is elastic, show that the two particles emerge from the collision at a right angle to one another, provided that they have the same mass.

Solution:

Momentum conservation gives

$$\vec{p}_1 = \vec{p}' + \vec{p}' \tag{10.14}$$

or

$$m_1\vec{v}_1 = m_1\vec{v}_1' + m_2\vec{v}_2'. \tag{10.14a}$$

The energy balance condition is

$$\frac{1}{2}m_1v_1^2 = \frac{1}{2}m_1v_1'^2 + \frac{1}{2}m_2'^2 + Q \tag{10.15}$$

or

$$\frac{p_1^2}{2m_1} = \frac{p_1'^2}{2m_1} + \frac{p_2'^2}{2m_2} + Q. \tag{10.15a}$$

Here, as before, the primes indicate the velocities and momenta after the collision, and Q represents the net energy that is lost or gained during the collision.

If $m_1 = m_2 = m$, then Equation 10.15a becomes

$$p_1^2 = p_1'^2 + p_2'^2 + 2mQ. \tag{10.16}$$

Taking the dot product of each side of Equation 10.14 with itself, we obtain

$$p_1^2 = p_1'^2 + p_2'^2 + 2\vec{p}_1' \cdot \vec{p}_2'. \tag{10.17}$$

Substituting Equation 10.16 into Equation 10.17, we obtain

$$\vec{p}_1' \cdot \vec{p}_2' = mQ.$$

For an elastic collision, $Q = 0$, then

$$\vec{p}_1' \cdot \vec{p}_2' = 0.$$

This last equation says that the two particles emerge from the collision at a right angle to one another.

Example 10.2

A smooth sphere of mass M is tied to a fixed point by a light, inextensible string. Another sphere of mass m with initial velocity v_1 in a direction making an angle θ with the string makes a direct impact with M (Figure 10.2). Find the velocity with which M begins to move after the collision. The coefficient of restitution is e.

Solution:

Because the string AB is inextensible, the sphere M is constrained to move in a circle about point A, and the velocity v_2', with which M begins to move after the collision, is perpendicular to the line AB. The impulse between the spheres is along the line of mass centers, so the direction of motion of m is unaltered by the collision, and v_1' is in the same direction as v_1. The conservation of the component of momentum at right angles to AB gives

$$Mv_2' + mv_1' \sin\theta = mv_1 \sin\theta \tag{10.18}$$

and Newton's rule gives

$$v_1' - v_2' \sin\theta = -ev_1. \tag{10.19}$$

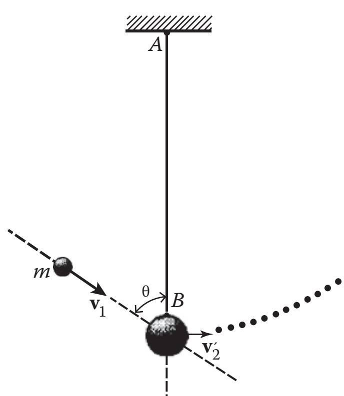


FIGURE 10.2 Oblique collision of two spheres.

Solving Equations 10.18 and 10.19 for v_2', we obtain

$$v_2' = \frac{(e+1)mv_1 \sin\theta}{M + m\sin^2\theta}.$$

We can also solve Equations 10.18 and 10.19 for v_1':

$$v_1' = -v_1 \frac{m\sin^2\theta - eM}{M + m\sin^2\theta}.$$

10.2 SCATTERING CROSS SECTIONS AND RUTHERFORD SCATTERING

Most scattering events of significance involve motion in three dimensions, and this adds considerably to the complexity of the analysis. A typical experimental arrangement is shown schematically in Figure 10.3. A collimated beam of particles of mass m_1 is incident upon a fixed scattering target of mass m_2. During scattering, the interaction is sampled over a wide range of distances and, possibly, relative velocities. For that reason, it is not surprising that a measured scattering distribution can usually be fitted with a variety of theoretical models. In the following paragraph, we shall derive some of the features of potential scattering without ever committing ourselves to a specific potential for the scattering centers. Because scattering encompasses a wide variety of processes, we narrow down the number of possibilities as follows:

1. The interaction between the particles can be described by a potential, which is spherically symmetrical and decreases rapidly as we go away from the center of force, vanishing at an infinite distance away from it.
2. The scattering process is elastic. That is, neither of the participating particles changes its rest mass, and no particles are created or annihilated during the collision.
3. Target particles and incident particles are different. This means that we can always tell whether a particle emerging from the target is a scattered incident particle or a recoiling target particle.

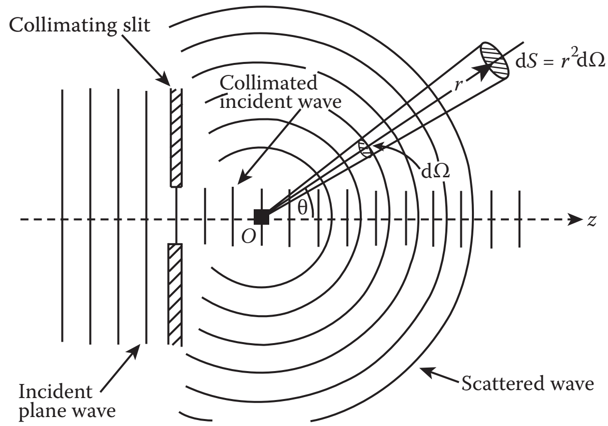


FIGURE 10.3 Typical scattering experimental arrangement.

10.2.1 Scattering Cross Sections

In the analysis of the scattering experimental results, an extremely useful concept is the scattering cross section, which depends on the nature of the interaction between the incident particle and the target and on their individual properties. Thus, a measured cross section will reveal this information, and comparing the predicted and measured cross section can test a theory about the behavior of the incident and target particles.

In a typical scattering experiment, the number of particles involved is quite large so that, even classically, it is impractical to measure the details of each individual trajectory. What we measure is the initial velocity of the particles (they are usually prepared to be monoenergetic) and their final velocity. That is, if the direction of the incident beam is taken to be at $\theta = 0$ (Figure 10.3), then we measure the number of particles that scatter into a solid angle $\mathrm{d}\Omega$ at (θ,ϕ). Hence, if N is the number of incident particles crossing the unit area normal to the beam in unit time, and $\mathrm{d}N$ of them scatter by an angle (θ,ϕ) into the solid angle $\mathrm{d}\Omega$ in the unit time, then the differential cross section is defined to be

$$\sigma(\theta,\phi) = \frac{\mathrm{d}\sigma}{\mathrm{d}\Omega}(\theta,\phi) = \frac{1}{N}\frac{dN}{\mathrm{d}\Omega} \tag{10.20}$$

in the limit $\mathrm{d}\Omega > 0$. N is often called the incident intensity or flux density. The differential cross section is a useful quantity because it probes the shape of the potential or the nature of the force—the larger the differential cross section, the stronger the force. If the potentials are spherically symmetrical, then physical observables cannot have an azimuthal dependence (no ϕ dependence). We can, in such cases, define the differential cross section for scattering into a ring around the beam axis by integrating out the azimuthal angle, that is,

$$\sigma(\theta) = \int \mathrm{d}\phi\sigma(\theta,\phi) = 2\pi\sigma(\theta,\phi). \tag{10.21}$$

We see that the differential cross section has the dimensions of an area (which arises from the "per unit area" in the denominator in Equation 10.20). Indeed, σ may be interpreted as the area of a disc with the center at the scattering center, which is placed perpendicular to the path of the incident beam and which must be struck by this beam in order that the particles be scattered into the solid angle $\mathrm{d}\Omega$ at the angle θ. We may relate this to the classical motion of the particle by means of the impact parameter $b(\theta)$ as shown in Figure 10.4. This is the distance away from the center of force where the line of the initial particle path must lie in order that the particle be deflected classically through an angle θ. A beam of particles will have a range of impact parameters, and the outgoing particles will then be scattered through a corresponding range of angles.

It is clear that the smaller the impact parameter, the greater will be the effect of the force and hence the scattering angle. If we call the initial magnitude of velocity v_0 for the beam, then

$$E = \frac{1}{2}mv_0^2$$

or

$$v_0 = \sqrt{2E/m}$$

and the angular momentum of the incident particle is

$$\ell = mv_0 b = \sqrt{2mE}b.$$

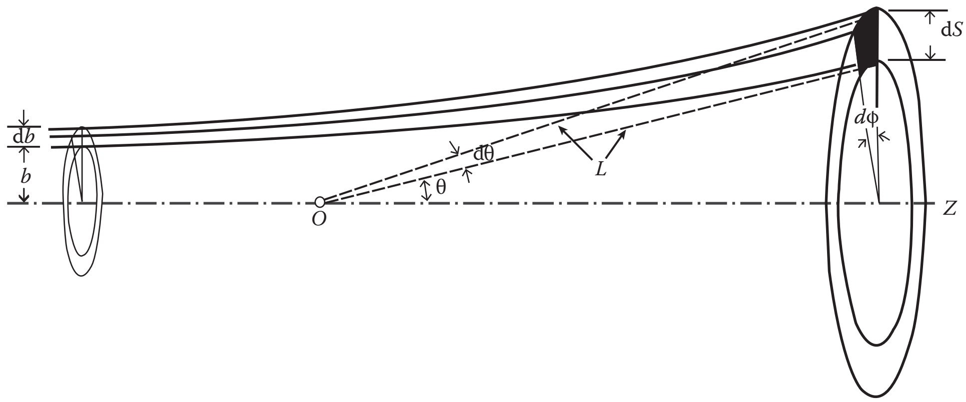


FIGURE 10.4 Scattering of an incident beam of particle by a center of force.

Those particles falling in the range of impact parameters between b and $(b + \mathrm{d}b)$ will be scattered (into the detector) through angles between θ and $(\theta + \mathrm{d}\theta)$ or into the solid angle $\mathrm{d}\Omega$ about the direction θ. If the surface area of the detector is dS, then the solid angle $\mathrm{d}\Omega$ subtended by dS at the scattering center is

$$\mathrm{d}\Omega = \frac{\mathrm{d}S}{L^2} = \frac{(L\mathrm{d}\theta)L\sin\theta\mathrm{d}\phi}{L^2} = \sin\theta\mathrm{d}\theta\mathrm{d}\phi = 2\pi\sin\theta\mathrm{d}\theta.$$

The number of particles per unit time crossing a ring of area $2\pi b\ \mathrm{d}b$ (about the central axis and with plane perpendicular to the beam) is $(2\pi b\ \mathrm{d}b)N$; these particles will all be scattered into $\mathrm{d}\Omega$. Substituting these into Equation 10.1, we obtain

$$\sigma(\theta) = -\frac{b}{\sin\theta}\frac{\mathrm{d}b}{\mathrm{d}\theta}. \tag{10.22}$$

The minus sign is introduced in Equation 10.3 because an increase db in the impact parameter means less force is exerted on the particle, resulting in a decrease $\mathrm{d}\theta$ in the scattering angle. It is this quantity $\sigma(\theta)$ that is measured in practice. A particle detector on a swinging arm is moved around, changing θ, and the particle count at each position is recorded. The principal task then is to relate the scattering angle as a function of the impact parameter, $\theta(b)$, to the force law or the corresponding potential energy function $V(r)$. We now consider Rutherford's α-particle scattering as an example.

10.2.2 Rutherford's α-Particle Scattering Experiment

Up to Rutherford's time, the popular view of the atom was embodied in Thomson's "plum pudding" model. According to this, there was a positive charge spread over the volume of the atom (the pudding) and sufficient negatively charged electrons dotted around (the plums) to produce an atom with no overall electrical charge. Now, the alpha particle consists of two protons and two neutrons, that is, a helium nucleus. Its positive charge makes it sensitive to the electric fields inside the atom and hence the internal charge distribution.

According to Thomson's model, one would not expect the alpha particles incident on a thin foil to be deflected significantly. To Rutherford's surprise, however, there was considerable scattering. Some alpha particles were even reflected, that is, scattered through 180°. Rutherford realized that the positive charge of the atom had to be concentrated at its center, and this suggested that the electrons orbited around it like a miniature solar system.

Now, our next test is to calculate the scattering cross section resulting from the Coulomb interaction. An alpha particle will follow a hyperbolic path. We have learned in Chapter 5 that, for an attractive force ($k > 0$), when energy E is positive and the eccentricity ε is greater than 1, the orbit is hyperbolic; for a repulsive force ($k < 0$), we must have $E > 0$ because the kinetic energy cannot be negative, and potential energy $V = -k/r$ is positive. The orbit is also a hyperbola given by

$$r = \frac{\alpha}{-1 + \varepsilon\cos\phi}. \tag{10.23}$$

We can combine these two equations into one:

$$r = \frac{\alpha}{\pm 1 + \varepsilon\cos\phi} \tag{10.24}$$

where the positive sign for $k > 0$ is the attractive force, and the negative sign for $k < 0$ is the repulsive force. Equation 10.24 represents the two branches of the same hyperbola as shown in Figure 10.5 in which point O is the center of force. In terms of Cartesian coordinates $x = r\cos\theta$ and $y = r\sin\theta$, Equation 10.24 becomes

$$\frac{(x - a\varepsilon)^2}{a^2} - \frac{y^2}{b^2} = 0 \tag{10.25}$$

where

$$a = \frac{\alpha}{\varepsilon^2 - 1} = \frac{|k|}{2E}, \quad b = \frac{\alpha}{\sqrt{\varepsilon^2 - 1}} = a\sqrt{\varepsilon^2 - 1} = \frac{l}{\sqrt{2E}}. \tag{10.26}$$

Equation 10.24 represents a hyperbola with its center C at $x = a\varepsilon$, $y = 0$, and asymptotes $y = \pm b(x - a\varepsilon)/a$. At great distances, the orbit merges into the asymptotes CZ and CY; θ_s is, therefore, the scattering angle.

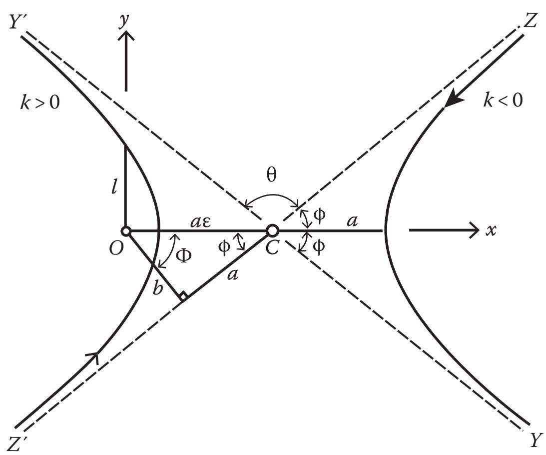


FIGURE 10.5 Hyperbolic orbit for $k < 0$ and $k > 0$. O is the center of force.

In a Rutherford scattering, we shall consider the motion of a particle in an inverse-square repulsive field. From Figure 10.5, we see that

$$\theta + 2\varphi = 2\pi$$

or

$$\theta/2 = \pi/2 - \varphi$$

and so

$$\sin(\theta/2) = \cos\varphi.$$

From Equation 10.23, we see that $r \to \infty$ or $1/r \to 0$. That is, $-1 + \varepsilon\cos\varphi = 0$ or

$$\cos\varphi = 1/\varepsilon$$

Accordingly,

$$\sin(\theta/2) = \cos\phi = 1/\varepsilon.$$

Now, relating this to the impact parameter b and the energy E, we have the expression of the eccentricity (Equation 6.38a):

$$\varepsilon = \left(1 + \frac{2E\ell^2}{mk^2}\right)^{1/2}$$

and the expression of the angular momentum

$$\ell = b(2mE)^{1/2}.$$

From the last three expressions, we obtain

$$\varepsilon = \left(1 + \frac{4E^2b^2}{k^2}\right)^{1/2} = \cos ec\frac{\theta}{2}.$$

Squaring this,

$$1 + \frac{4E^2b^2}{k^2} = \cos ec^2\frac{\theta}{2} = 1 + \cot^2\frac{\theta}{2}$$

which gives

$$\cot\frac{\theta}{2} = \frac{2Eb}{k} \tag{10.27}$$

or

$$b = \frac{k}{2E}\cot\frac{\theta}{2}.$$

Differentiating with respect to θ_s, we obtain

$$\frac{db}{d\theta} = -\frac{k}{4E}\frac{1}{\sin^2(\theta/2)} \tag{10.28}$$

so that $\sigma(\theta)$ becomes, by Equation 10.22,

$$\sigma(\theta) = \frac{1}{2}\left(\frac{k}{2E}\right)^2 \frac{\cot(\theta/2)}{\sin\theta}\frac{1}{\sin^2(\theta/2)}$$

or

$$\sigma(\theta) = \frac{1}{4}\left(\frac{k}{2E}\right)^2 \frac{1}{\sin^4(\theta/2)} \tag{10.29}$$

where $k = 2Z_e'^2$.

This is the famous Rutherford scattering cross section, originally derived by Rutherford to explain the experimental results of Geiger and Marsden on the scattering of alpha particles by heavy nuclei (gold, $Z' = 79$). In his derivation, this cross section is strongly dependent on both the velocities of the incident particle and the scattering angle. It also increases rapidly with increasing charge, $k = ZZ'e^2$. Thus, we expect the number of particles scattered to increase as Z'^2 with increasing atomic number. Because $\sigma(\theta_s)$ is independent of the sign of k, the form of scattering distribution is the same for an attractive force as for a repulsive one.

In the derivation of the Rutherford formula, an implicit assumption was made that no incident particle interacts with more than one target nucleus, which is valid if the scattering angle is not too small. In actual experiments, incident particles, particularly electrons, are deflected many times while passing through the metal target foil; hence, the net angle of deflection is the result of a statistical accumulation of small single scattering events. The theory of multiple scattering is complicated and is beyond the scope of this text. A beautiful, simplified treatment of multiple scattering is given in *Nuclear Physics* by Fermi (1949).

The Rutherford formula can be expressed in terms of momentum transferred to the striking alpha particle. In scattering off a nucleus in the target, the momentum

$$\vec{q} = \vec{p}_f - \vec{p}_0 \tag{10.30}$$

is transferred to the alpha particle. Taking the square of $\vec{q}$, we have

$$\vec{q}\cdot\vec{q} = q^2 = p_f^2 + p_0^2 - 2\vec{p}_f\cdot\vec{p}_0 = p_f^2 + p_0^2 - 2p_f p_0\cos\theta_s$$

which relates the magnitude of $\vec{q}$ to the scattering angle θ_s. If the mass M of the target nucleus is much heavier than the mass m of the alpha particle, then the energy conservation condition

$$\frac{p_0^2}{2m} = \frac{p_f^2}{2m} + \frac{q^2}{2M}$$

reduces to

$$p_f = p_0 = mv$$

where v is the velocity of the incident alpha particle. Thus,

$$q^2 = 2p_0^2(1-\cos\theta) = 4m^2v^2\sin^2\frac{\theta}{2} = 2m\left(4E\sin^2\frac{\theta}{2}\right)$$

and the Rutherford differential scattering cross section becomes

$$\sigma(\theta) = \left(\frac{2mk}{q^2}\right)^2. \tag{10.31}$$

Let us return to Equation 10.29. If we want the probability that any interaction whatsoever will take place, it is then necessary to integrate $\sigma(\theta)$ over all possible scattering angles; the resulting quantity is the total scattering cross section σ_t:

$$\sigma_t = \int \sigma(\theta)\,d\Omega = \int_{\phi=0}^{2\pi}\int_{\theta=\theta_0}^{\pi} \sigma(\theta)\sin\theta\,d\theta\,d\phi. \tag{10.32}$$

θ_0 is the least scattering angle. If we attempt to calculate the total cross section for a Coulomb scattering by substituting Equation 10.29 into Equation 10.32, we get an infinite result as $\theta_0 \to 0$:

$$\sigma_t = 4\pi\left(\frac{ZZ'e^2}{2mv^2}\right)^2\left(\frac{1}{\sin^2(\theta_0/2)} - 1\right) \to \infty \text{ as } \theta_0 \to 0.$$

This is a consequence of the infinite range of the Coulomb force but not a feature peculiar to the Coulomb field. Classically, a force field that extends to infinity would give an infinite scattering cross section. But quantum theory gives a finite value. The reason for this is that in classical theory all deflections are counted, no matter how small, whereas in quantum theory, only those deflections that exceed the limits of the uncertainty principle are counted.

10.2.3 Cross Section Is Lorentz Invariant

A question often asked by physics majors in studying collisions and scatterings is are cross sections Lorentz invariant? Lorentz invariance is a concept developed in special relativity. In special relativity, time is not absolute, and objects contract in the direction of motion (the Lorentz contraction); time and space are treated on equal footing. In special relativity, we deal with four-dimensional space–time (ct, x, y, z). The three-dimensional Galilean transformations in Newtonian mechanics are replaced by the four-dimensional Lorentz transformations. And a quantity that is not changed under a Lorentz transformation is known as a Lorentz invariant quantity. The cross section is found by dividing the number of events by the number of particles that were incident per unit area transverse to the direction of relative motion. The number of particles and transverse areas are invariant; thus, the cross section is Lorentz invariant.

10.3 LABORATORY AND CENTER-OF-MASS FRAMES OF REFERENCE

In the study of scatterings, physical measurements are generally made in the laboratory coordinate system (L system, for short), in which the incident particle (m_1) moves in and the target particle (m_2) is initially at rest. The center-of-mass system (C system), in which the center of mass of the colliding particles is always at rest, is a more convenient system for describing particle collisions, and theoretical calculations are generally done in terms of quantities referred to the C system. Measurements made in the L system must be reduced to the C system to compare with theory. Hence, it is important to consider the problem of conversion from one coordinate system to the other. Figure 10.6 illustrates the difference of the two systems for non-relativistic collisions between two particles. Scattering of two particles as viewed in the laboratory system is indicated in Figure 10.6a, and Figure 10.6b indicates scattering of two particles as viewed in the center of mass system. In what follows, we shall use the subscript L to denote quantities in the L system. For a non-relativistic collision between a beam of incident particle of mass m_1 and a target object of mass m_2, if $m_2 \gg m_1$ (as in the scattering of electrons by atoms or molecules), the L and C systems nearly coincide and are the same in the limit $m_1/m_2 \to 0$. If m_2 is not very much larger than m_1, the theoretical description of the collision is best carried out in the C system. In such a system, the total linear momentum is zero, of course, and the two particles always move with equal and opposite momenta. The connection between the two scattering angles θ_L and θ can be obtained by considering the transformation between the L and C systems. We use the following quantities in the two systems:

$\vec{r}_1$ and $\vec{v}_1$ are the position and velocity, respectively, of the incident particle 1 in the L system.
$\vec{r}_1'$ and $\vec{v}_1'$ are the position and velocity, respectively, of particle 1 in the C system.
$\vec{R}$ and $\vec{R}'$ are the position and (constant) velocity, respectively, of the center of mass in the L system.

At any instant, by definition

$$\vec{r}_1 = \vec{R} + \vec{r}_1'$$

and so

$$\vec{v}_1 = \dot{\vec{R}} + \vec{v}_1'. \tag{10.33}$$

Figure 10.7 graphically indicates this vector relation evaluated after the scattering has taken place, at which time $\vec{v}_1$ and $\vec{v}_1'$ make the angles θ_L and θ, respectively, with the vector $\bar{R}$ lying along the initial direction. From the diagram, we obtain the equation

$$\tan\theta_L = \frac{v_1'\sin\theta}{v_1'\cos\theta + \dot{R}} = \frac{\sin\theta}{\cos\theta + \dot{R}/v_1'}. \tag{10.34}$$

We can express $\dot{R}/v_1'$ in terms of m_1 and m_2. Now, $\vec{v}_1'$ is related to the relative velocity $\dot{\vec{r}}$

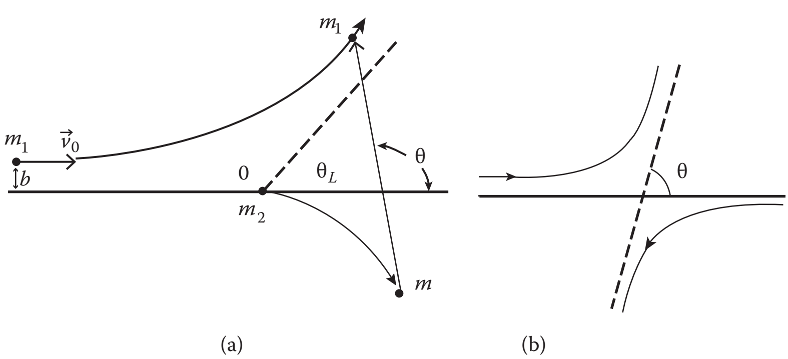


FIGURE 10.6 Scattering of two particles: (a) as viewed in the laboratory system; (b) as viewed in the center of mass system.

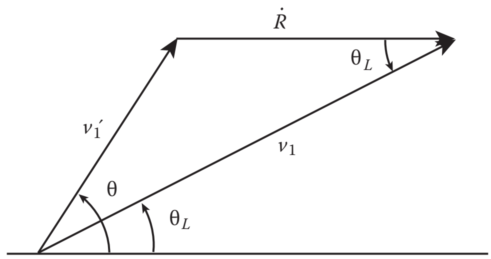


FIGURE 10.7 Vector relation between $\vec{v}_1$, $\vec{v}_1'$, and $\dot{\vec{R}}$ after the scattering.

$$\vec{v}_1' = \frac{\mu}{m_1}\dot{\vec{r}}$$

where μ is the reduced

$$\mu = \frac{m_1 m_2}{m_1 + m_2}$$

and

$$\vec{r} = \vec{r}' - \vec{r}.$$

As the system is conservative, the relative velocity after scattering must have the same magnitude as the initial velocity $\vec{v}_0$. Hence, after scattering

$$v_1' = \frac{\mu}{m_1} v_0. \tag{10.35}$$

The constant velocity of the center of mass can be found from the conservation law for the total linear momentum:

$$(m_1 + m_2)\dot{\vec{R}} = m_1 \vec{v}_0$$

or

$$\dot{\vec{R}} = \left(\mu/m_2\right)\vec{v}_0. \tag{10.36}$$

Substituting Equation 10.36 and Equation 10.35 into Equation 10.34, we obtain

$$\tan\theta_L = \frac{\sin\theta}{\cos\theta + m_1/m_2}. \tag{10.37}$$

When $m_1 \ll m_2$, the two angles are approximately equal, and the massive scatter particle m_2 suffers little recoil and are practically as a fixed center of mass.

The scattering cross sections are also different in the C and L systems. To transfer the scattering cross section from the C system to the L system or vice versa, we make use of the fact that the total number of particles scattered into a unit solid angle are the same in both systems, that is,

$$N\sigma(\theta)\mathrm{d}\Omega(\theta) = N\sigma(\theta_L)\mathrm{d}\Omega(\theta_L)$$

or

$$\sigma(\theta_L) = \sigma(\theta)\frac{\sin\theta \mathrm{d}\theta}{\sin\theta_L \mathrm{d}\theta_L} = \sigma(\theta)\frac{\mathrm{d}\cos\theta}{\mathrm{d}\cos\theta_L} \tag{10.38}$$

where $\mathrm{d}\Omega(\theta)$ and $\mathrm{d}\Omega(\theta_L)$ represent the same solid angle element but measured in the C frame and L frame, respectively.

In principle, the derivative $\mathrm{d}\theta/\mathrm{d}\theta_L$ can be evaluated by using Equation 10.12, but it is quite a job because of the presence of the m_1/m_2 ratio in the equation's denominator. Let us consider two special cases: (1) $m_2 \gg m_1$ and (2) $m_2 = m_1$. For the first case, the corrections are small as for the Rutherford α-particle scattering, where m_1 is 4 atomic units and m_2 may be 100 atomic units or larger. But if the two masses are equal, as in an *n-p* scattering, then the effects are quite large. In this case, Equation 10.12 becomes

$$\tan\theta_L = \frac{\sin\theta}{\cos\theta + 1} = \tan\frac{\theta}{2}$$

and so

$$\theta_L = \frac{\theta}{2} \quad (m_1/m_2 = 1) \tag{10.39}$$

which indicates that the maximum scattering angle in the L system is 90°. Equation 10.38 reduces to

$$\sigma(\theta_L) = 4\cos\theta_L\sigma(2\theta_L),\ (m_1/m_2 = 1). \tag{10.40}$$

Scattering can only be described classically provided that the wavelength of the particle is much smaller than the dimensions of the target, in which case diffraction effects are negligible. If this condition is not satisfied, we cannot speak in terms of particle trajectories. For instance, to define an impact parameter, we must know the lateral position of the particle relative to the target, but this introduces an uncertainty in the transverse momentum. The description of a particle trajectory requires the precise simultaneous knowledge of both these quantities, and this would violate the uncertainty principle. Accordingly, we must resort to a quantum mechanical description.

Example 10.3

A beam of α particles strikes a helium target. If a certain incident α-particle is scattered through an angle of 30° in the laboratory frame, find the kinetic energy of both the incident particle and the target particle.

Solution:

Assume that the target particle is at rest and that the collision is elastic. With $m_1 = m_2$, we have $\theta_l + \phi_l = 110°$. Hence, the momentum conservation gives

$$p_1' = p_1\cos\theta_l + p_2'\cos\phi_l \quad \text{and} \quad 0 = p_1'\sin\theta_l - p_2'\sin\phi_l$$

in which $\theta_l = 30°$. Solving for p_1' and p_2', we find

$$p_1' = p_1 \cos 30° = \sqrt{3} p_1/2 \quad \text{and} \quad p_2' = p_1 \sin 30° = p_1/2.$$

Thus, the kinetic energies after impact are

$$T_1' = \frac{p_1'^2}{2m_1} = \frac{3p_1^2}{8m_1} = \frac{3T}{4} = 3 \text{ MeV}$$

and

$$T_2' = \frac{p_2'^2}{2m_2} = \frac{p_1^2}{4m_1} = \frac{T}{4} = 1 \text{ MeV}.$$

10.4 NUCLEAR SIZES

The Rutherford alpha particle scattering experiment provided the first evidence that nuclei are of finite size, much smaller than the atomic dimensions. In this α particle scattering experiment, the distance of closest approach is found from Equation 10.23 to be

$$r_{\min} = \frac{\alpha}{\varepsilon - 1} \tag{10.41}$$

where

$$\alpha = \ell^2/mk$$

and

$$\varepsilon = \left(1 + \frac{2E\ell^2}{mk^2}\right)^{1/2}.$$

From the last two expressions, we obtain

$$\alpha = \frac{k}{2E}(\varepsilon^2 - 1).$$

Substituting this into Equation 10.41, we obtain

$$r_{\min} = -\frac{k}{2E}(1 + \varepsilon). \tag{10.42}$$

We can express this in terms of the scattering angle θ_s. To this purpose, we revisit Equation 10.23, which gives, as $r \to \infty$,

$$-1 + \varepsilon \cos \varphi = 0 \text{ or } \cos \varphi = 1/\varepsilon.$$

As $\theta_s = \pi - 2\varphi$, we can write the last expression as

$$\varepsilon = \frac{1}{\sin(\theta/2)} \tag{10.43}$$

and finally we can express r_{min} in terms of the scattering angle θ_s:

$$r_{\text{min}} = \frac{k}{2E}\left(1 + \frac{1}{\sin(\theta/2)}\right). \tag{10.44}$$

Thus, to investigate the structure of the atom at small distances, we have to use high velocity particles and examine the large-angle scattering (this corresponds to particles with small impact parameters). If the positive nuclear charges were spread out over a large volume, the force would be the inverse-square law force down to a distance equal to the radius of the charge distribution. Beyond this point, the force would decrease as we go to even smaller distances. Consequently, the particles that penetrate to within this distance would experience a weaker force than the inverse-square law predicts and would be scattered through smaller angles. The region down to $r_{\text{min}} = k/E$ is probed at $\theta = \pi$. By studying the backward scattering events, Rutherford found that the upper limit of the radius of the gold nucleus was 11^{-12} cm, well under the 11^{-8} cm radius of the atom.

In more recent years, experiments on the sizes of nuclei have employed electrons of several hundred megaelectron volts to more than 1 GeV and neutrons of 20 MeV and up. In every case, it has been found that the radius R of a nucleus can be expressed as

$$R = R_0 A^{1/3} \text{ and } R_0 \cong 1.2 \times 11^{-13}\,\text{cm}$$

where A is the mass number of the sample nucleus. The indefiniteness in R_0 is, in addition to experimental error, a result of the fact that electrons and neutrons interact differently with nuclei. The value of R_0 is slightly smaller when it is deduced from electron scattering.

10.5 SMALL-ANGLE SCATTERING (OPTIONAL)

It is apparent that Rutherford scattering has a pronounced maximum for small scattering angles. This maximum, as pointed out earlier, is associated with large impact parameters: particles passing each other at great distances deflect slightly while large-impact parameters predominate in the expression for the differential scattering cross section because they involve a larger area. In case only the angular dependence of the cross section in the limit of a small scattering angle is desired, the calculation of the differential scattering cross section can be much simplified. As shown in Figure 10.8 for small-angle scattering, the scattering angle θ is approximately given by

$$\sin\theta \approx \theta = p'_y/p'. \tag{10.45}$$

p' is the magnitude of the momentum of the particle after scattering, and p'_y is the component of p' perpendicular to the incident direction. We may calculate p'_y from Newton's second law:

$$p'_y = \int_{-\infty}^{\infty} F_y\,\mathrm{d}t = \int_{-\infty}^{\infty} \frac{ZZ'e^2}{r^2}\cos\alpha\,\mathrm{d}t = \int_{-\infty}^{\infty} \frac{ZZ'e^2 b}{(b^2 + v^2t^2)^{3/2}}\,\mathrm{d}t.$$

This can be rewritten as

$$p'_y = \frac{2ZZ'e^2}{bv}\int_0^{\infty} \frac{\mathrm{d}x}{(1+x^2)^{3/2}}$$

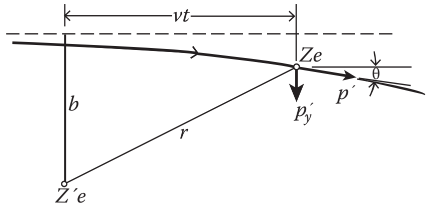


FIGURE 10.8 Small-angle scattering.

where $x = vt/b$. The integral can be performed by substituting $x = \tan\gamma$. The result is $p'_y = \dfrac{2ZZ'e^2}{bv}$.

Substituting this into Equation 10.45, we find the scattering angle θ:

$$\theta \cong \frac{p'_y}{p'} = \frac{2ZZ'e^2}{mv^2 b} \tag{10.46}$$

which gives the relationship between θ and the impact parameter b for small-angle scattering. The Rutherford formula for small-angle scattering is obtained from Equation 10.22:

$$\sigma(\theta) = \frac{d\sigma}{d\Omega} = \frac{2ZZ''e^2}{mv^2}\frac{1}{\theta^4}. \tag{10.47}$$

Caution should be exercised when dealing with small-angle scattering. In experiments, weakly deflected particles are not all detected as having been deflected. This is because the incident beam already possesses a certain scattering, which means deflection angles lying within this scattering angle cannot be detected.

Example 10.4: Scattering by a Rigid Sphere

Consider the scattering of particles by a rigid sphere of radius R. There is no interaction as long as a particle does not hit the sphere. Thus, the potential outside the sphere is zero. If the sphere is not penetrable, the potential inside the sphere must be infinity:

$$\begin{aligned} U(r) &= 0, \quad r > R \\ &= \infty \quad r \le R \end{aligned}$$

Solution:

When the incident particle approaches the sphere with an impact b greater than R, it will not be scattered; thus,

$$\theta = 0 \text{ for } b > R.$$

A particle with impact parameter $b < R$ incident at an angle α with the normal to the surface of the sphere will be scattered off on the other side of the normal at the same angle α resulting from the conservation law of energy and momentum. We observe from Figure 10.9 that

$$\sin\alpha = \frac{b}{R}.$$

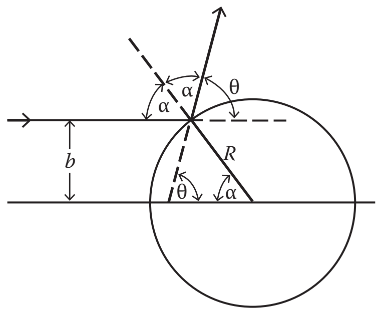


FIGURE 10.9 Scattering by a rigid sphere.

So

$$\theta = \pi - 2\alpha = \pi - 2\sin^{-1}\frac{b}{R}$$

which gives the impact parameter b as a function of the scattering angle θ for $b < R$:

$$b = R\sin\frac{\pi-\theta}{2} = R\cos\frac{\theta}{2} \quad \text{and} \quad b < R$$

and from Equation 10.22, the differential scattering cross section is

$$\sigma(\theta) = R^2/4$$

which is a constant, independent of θ and the incident energy. Similarly, the total scattering cross section is also a constant:

$$\sigma_t = \int \sigma(\theta)\mathrm{d}\Omega = \int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi} \sigma(\theta)\sin\theta \ \mathrm{d}\theta \ \mathrm{d}\phi = \pi R^2.$$

Quantum theory gives different results for rigid sphere scattering. In the low incident energy approximation, the quantum results are

$$\sigma(\theta) = R^2 \text{ and } \sigma_t = 4\pi R^2$$

while in the high incident energy approximation, the differential scattering cross section is rather difficult to find. The total cross section is twice that of the classical result:

$$\sigma_t = 2\pi R^2.$$

The discrepancy between the classical and quantum results for rigid sphere scattering cannot be explained satisfactorily in the framework of classical mechanics.

Example 10.5: Scattering by a Spherical Potential Well

A central force potential frequently encountered in nuclear physics is the spherical potential well, defined by

$$\begin{aligned} U(r) &= 0 \text{ and } r > R \\ &= -U_0, \ r < R. \end{aligned}$$

Solution:

When the impact parameter b of an incident particle is greater than R, it will not be scattered. If a particle approaching the potential well with an impact parameter $b < R$, it will be deflected on entering and leaving the well at points a and b, respectively. The particle enters the well at an angle α with the normal at a and leaves the well at point b at the same angle α with the normal at b. Point b' is the distance of the trajectory ab from the force center O. An inspection of Figure 10.10 shows that

$$\gamma = \alpha - \beta \text{ and } \theta = \gamma + \gamma = 2\gamma.$$

Now, the conservation of energy and angular momentum give, respectively,

$$E = \frac{1}{2m}p^2 = \frac{1}{2m}p'^2 + (-U_0)$$

$$pb = p'b'$$

where p is the momentum of the particle outside the potential well, and p' is the momentum inside the well.

Eliminating p' from the two previous equations, we find

$$b' = \frac{pb}{p'} = \frac{pb}{(p^2 + 2mU_0)^{1/2}}. \tag{10.48}$$

The scattering angle θ can be expressed in terms of b and b':

$$\theta = 2\gamma = 2(\alpha - \beta) = 2\left(\sin^{-1}(b/R) - \sin^{-1}(b'/R)\right)$$

from which we obtain

$$\sin^{-1}\frac{b}{R} = \frac{\theta}{2} + \sin^{-1}\frac{b'}{R}$$

or

$$\begin{aligned}\frac{b}{R} &= \sin\left(\frac{\theta}{2} + \sin^{-1}\frac{b'}{R}\right)\\ &= \sin\frac{\theta}{2}\cos\left(\sin^{-1}\frac{b'}{R}\right) + \cos\frac{\theta}{2}\sin\left(\sin^{-1}\frac{b'}{R}\right).\end{aligned} \tag{10.49}$$

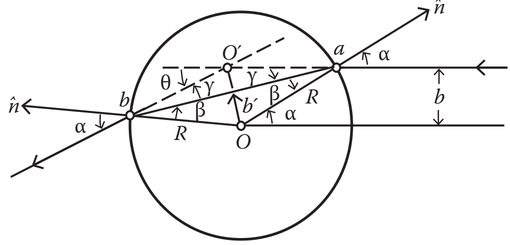


FIGURE 10.10 Scattering by a spherical potential well.

The quantity $\cos(\sin^{-1}b/R)$ needs some special attention. Let $y = \sin^{-1}(b'/R)$; then $\sin(y) = b'/R$, and

$$\cos(y) = (1 - \sin^2 y)^{1/2} = (1 - b'^2/R^2)^{1/2}.$$

That is,

$$\cos\left(\sin^{-1}(b'/R)\right) = \left(1 - b'^2/R^2\right)^{1/2}$$

and Equation 11.68 becomes

$$\frac{b}{R} = \left(1 - \frac{b'^2}{R^2}\right)^{1/2} \sin\frac{\theta}{2} + \frac{b'}{R}\cos\frac{\theta}{2}$$

from which we obtain

$$b^2 + b'^2 - 2bb'\cos\frac{\theta}{2} = R^2 \sin^2\frac{\theta}{2}. \tag{10.50}$$

Eliminating b' from this equation and from Equation 10.50, we obtain the impact parameter b as a function of the scattering angle θ:

$$b = \frac{R(p'/p)\sin(\theta/2)}{\sqrt{1 + (p'/p)^2 - 2(p'/p)\cos(\theta/2)}}.$$

The range of θ varies from $\theta = 0$ to $\theta = \theta_{max}$, where θ_{max} is determined by setting $b = R$ in Equation 10.50.

The differential scattering cross section is

$$\sigma(\theta) = \frac{(p'/p)^2 R^2 \left[(p'/p)\cos(\theta/2) - 1\right]\left[p'/p - \cos(\theta/2)\right]}{4\cos(\theta/2)\left[1 + (p'/p)^2 - 2(p'/p)\cos(\theta/2)\right]}. \tag{10.51}$$

The total cross section, obtained by integrating $\sigma(\theta)$ over all angles within the cone $\theta < \theta_{max}$, is, of course, equal to the geometrical cross section πR^2.

It is interesting to see that the scattering produced by such a potential is identical with the refraction of light rays by a sphere of radius R and relative index of refraction

$$n = \frac{p'}{p} = \sqrt{\frac{E + U_0}{E}}.$$

This equivalence explains why it was possible to explain refraction phenomena both by Huygens' wave theory and by Newton's mechanical corpuscular theory.

PROBLEMS

1. Consider an isolated system of two masses m_1 and m_2 moving under a mutual force F.
   (a) Write down the equations of motion of the two bodies and show that these equations lead to the law of conservation of momentum.
   (b) In general, the force F must be a function of the relative distance $x = x_1 - x_2$ and the relative velocity $\dot{x} = \dot{x}_1 - \dot{x}_2$. Show that when F is a function of x, a potential function $U(x)$ can be introduced (so F is conservative).

(c) For collision problems, force F is generally small except when the bodies are very close together. The potential function is then a constant (say, zero) for large values of x and rises very sharply for small values (Figure 10.11a). An ideal impulsive conservative force corresponds to potential function with a step (Figure 10.11b). So long as the initial kinetic energy is less than the height of the step, the bodies will bounce off one another. Show that, for elastic collision, the relative velocity is just reversed in the collision:

$$v_2 - v_1 = u_1 - u_2$$

where u_1, u_2 are the initial velocities, and v_1, v_2 are the final velocities. Show that, when the second body is initially at rest ($u_2 = 0$)

$$v_1 = \frac{m_1 - m_2}{m_1 + m_2} u_1$$

$$v_2 = \frac{2m_1}{m_1 + m_2} u_1 .$$

These equations indicate that if the masses are equal, then the first body is brought to rest by the collision, and its velocity is transferred to the second body. If $m_1 > m_2$, the first body continues in the same direction, with reduced velocity, whereas if $m_1 < m_2$, it rebounds in the opposite direction. In the limit where m_2 is much larger than m_1, $v_1 = -u_1$ (i.e., m_1 rebounds from a fixed potential barrier).

2. A mass m strikes a smooth flat surface as shown in Figure 10.12. If the coefficient of restitution is e, determine the angle of rebound ϕ, the final velocity v_f, and the energy loss per impact in terms of initial velocity u and the angle of incidence θ.
3. Two spheres of mass m and M, respectively, collide obliquely. Find their velocities after impact in terms of their velocities before impact.
4. A gun is fired horizontally point-blank at a block of wood, which is initially at rest on a horizontal floor. The bullet becomes imbedded in the block, and impact causes the system to slide a certain distance s before coming to rest. Given m (mass of the bullet), M (mass of the block), and μ (coefficient of sliding friction between the block and the floor), find the muzzle velocity of the bullet.
5. A wedge of mass M rests with one face on a horizontal plane and with the inclined face at an angle α to the horizontal. It is hit on its inclined face by a particle of mass m moving along the plane with speed u in a direction perpendicular to the edge of the wedge. Show that the speed of the wedge immediately after impact is $mu \sin^2\alpha/(M + m \sin^2\alpha)$.

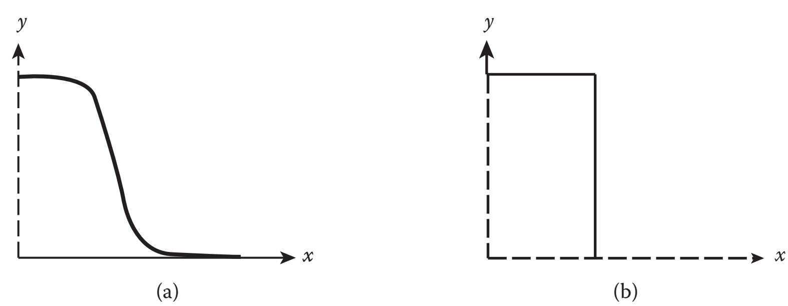


FIGURE 10.11 Scattering by potential wells. (a) smooth potential well, (b) step potential well.

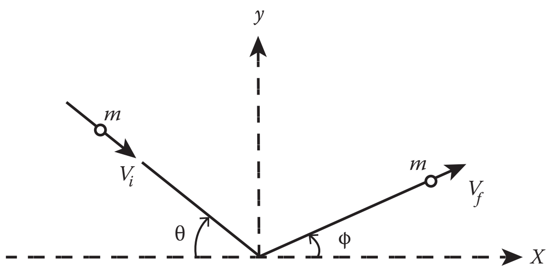


FIGURE 10.12 Mass m strikes a smooth flat surface.

The particle then slides up the face of the wedge and back to the horizontal plane. Find the final velocities of the particle and the wedge. It is to be assumed that there are no frictional forces and that the impact occurring is inelastic; in other words, relative motion in the direction of the impact is destroyed.

6. A particle of mass m lies at the middle, A, of a hollow tube of length $2b$ and mass M. The tube, which is closed at both ends, lies on a smooth table. The coefficient of restitution between m and M is e. Let m be given an initial velocity v_0 along the tube.
   (a) Find the velocities of m and M after the first impact.
   (b) Find the loss in energy during the first impact.
   (c) Find the time required for m to arrive back at A traveling in the original direction.
7. (a) In the system of Problem 6, determine the velocities relative to the center of mass before and after the first collision.
   (b) Find the loss in energy during the first impact in terms of a reference system with its origin at the center of mass.
   (c) How far has the center of mass traveled during the time mass m has arrived at A [i.e., during the time given by part (c) of Problem 6]?
8. In the elastic scattering of neutrons by protons ($m_n \cong m_p$) at relatively low energies, it is found that the energy distribution of the recoiling protons in the laboratory system is constant up to a maximum energy, which is the energy of the incident neutrons. Find the angular distribution of the scattering in the center of mass system.
9. Show that the energy distribution of particles recoiling from an elastic collision is directly proportional to the differential scattering cross section in the center of a mass system.
10. A beam of alpha particles, each of which has an energy of 4×10^6 ev, is projected toward thin-target gold foils.
    (a) Compute the cross section for scattering through an angle of $\pi/2$ or more.
    (b) The target gold foils in Rutherford experiment had a thickness of about 5×10^{-5} cm. An area of 1 cm^2 times that thickness yields a volume density of about 10^{16} atoms. If this volume is bombarded with an incident flux of 10^4 alpha particles/cm$^2 \cdot$ s, how many backward scatterings would we expect to observe in a minute?
11. Atoms, though overall neutral, attract each other. This phenomenon can be described adequately by a $1/r^6$ potential energy

$$V(r) = -V_0 \frac{r_0^6}{r^6}$$

where r_0 is a fixed length of atomic size, and V_0 is positive constant with the dimensions of energy. Suppose now a beam of hydrogen atoms passes into some cesium vapor. A cesium hydride molecule is formed if, we assume, the distance between a hydrogen and cesium atom is ever as small as r_0. The hydrogen atoms in the beam have energy E. The mass ratio is $m_H/m_{cs} = 1/133 \ll 1$. Calculate the cross section for forming a cesium hydride molecule.

12. The interaction between an atom and an ion at distances greater than contact is given by the potential energy

$$V(r) = -\frac{C}{r^4}$$

where $C = (e^2/2)\, P^2$, e is the ion charge, and P is the polarizability of the atom. Make a sketch of the effective energy versus the radial coordinate. If the total energy of the ion exceeds the maximum value of the effective potential energy, the ion spirals inward to the atom. Find the cross section for an ion of velocity v_0 to strike an atom. You can assume that the ion is much lighter than the atom.

13. A uniform beam of particles each of mass m and energy E is scattered by a fixed center of force. For a repulsive force $F = K/r^3$, find the differential scattering cross section. What will be the total cross section? Give physical discussion on its behavior.
14. If, in Example 4 (scattering by a rigid sphere), the energy lost by the scattered particle to the sphere is ε, show that

$$d\sigma_{cm}(\varepsilon) = \sigma_{cm}(\varepsilon) d\Omega = \frac{\pi R^2}{\varepsilon_{max}} d\varepsilon.$$

15. A fixed force center scatters a particle of mass m according to the force law

$$F = K/r^n, \quad n > 0.$$

If the initial velocity of the particle is u_0, determine the differential cross section for small-angle scattering.

REFERENCE

Fermi, E. Lecture notes of E. Fermi compiled by Jay Orear, *Nuclear Physics,* University of Chicago Press, Chicago, Illinois, USA, 1949.

11 Motion in Non-Inertial Systems

Newton's laws of motion apply only to an inertial frame of reference. In a non-inertial system, the equation of motion (Newton's second law) is modified by additional terms arising from the non-inertial effect of the system. In this chapter, we turn to the use of non-inertial systems. By introducing non-inertial systems, we can simplify many problems; in this sense, we gain one more computational tool. Moreover, consideration of non-inertial systems enables us to explore some of the conceptual difficulties of classical mechanics.

The motion of the non-inertial system may be accelerated translation, rotational, or a combination of both. It is often very convenient or sometimes necessary to study the behavior of a dynamic system in an accelerated coordinate system. For example, a coordinate system fixed on the rotating Earth is the most convenient one to use in describing the motion of a particle near the Earth's surface.

The reader should remember that the Galilean transformations only relate to observations in different inertial systems. According to the Galilean transformations, all systems translating uniformly relative to an inertial system are inertial.

11.1 ACCELERATED TRANSLATIONAL COORDINATE SYSTEM

Consider two coordinate systems O and O': their orientations in space are fixed, and their respective coordinate axes are parallel to each other (Figure 11.1). The O system is an inertial system in which Newton's laws of motion are valid. The respective position vectors $\vec{r}(t)$, $\vec{r}'(t)$, and $\vec{R}(t)$ are related by

$$\vec{r} = \vec{r}' + \vec{R}. \tag{11.1}$$

If the origin O' is moving relative to the origin O, which we take as fixed, the relation between the velocities relative to the two systems is obtained by differentiating Equation 11.1.

$$\frac{d\vec{r}}{dt} = \frac{d\vec{r}'}{dt} + \frac{d\vec{R}}{dt} \quad \text{or} \quad \vec{v} = \vec{v}' + V \tag{11.2}$$

where $\vec{v}$ and $\vec{v}'$ are the velocities of the moving particle P relative to O and O', and $\vec{V}$ is the velocity of O' relative to O. Differentiating Equation 11.2 once, we obtain the relationship between relative accelerations:

$$\frac{d\vec{v}}{dt} = \frac{d\vec{v}'}{dt} + \frac{d\vec{V}}{dt} \quad \text{or} \quad \vec{a} = \vec{a}' + \vec{A} \tag{11.3}$$

where $\vec{a}$ and $\vec{a}'$ are the accelerations of the particle P relative to O and O', and $\vec{A}$ is the acceleration of O' relative to O.

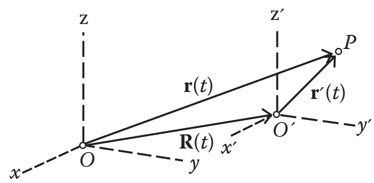


FIGURE 11.1 Initial and accelerated system.

Newton's laws of motion are valid in the unprimed O system:

$$m\vec{a} = \vec{F} \tag{11.4}$$

where $\vec{F}$ is the physical force acting on the particle. Substitution of Equation 11.3 into Equation 11.4 gives

$$m\vec{a}' = \vec{F} - m\vec{A}$$

or

$$m\frac{d^2\vec{r}'}{dt^2} = \vec{F} - m\frac{d^2\vec{R}}{dt^2}. \tag{11.5}$$

That is, Newton's second law is modified by the additional term $m\vec{A}$ in the O' system. This additional term $-mA$ is called a fictitious (or inertial) force; it can be altered or made to vanish merely by altering the state of the observer, that is, by the choice of a coordinate system. In contrast to this, a real physical force cannot be eliminated by the choice of a coordinate system. The real forces are a result of interactions with other objects, and they are independent of the state of the observer or the coordinate frame in which the motion might be described. This distinction is somewhat modified in the general theory of relativity. Because inertial systems have no unique position in the general theory of relativity, the fictitious forces in classical mechanics are regarded as on the same footing with the real force in general relativity, so the same law of motion holds in all coordinate systems.

Example 11.1

Consider the motion of the bob of a simple pendulum with a moving support (Figure 11.2), which accelerates translationally with acceleration $\vec{A}$. Show that the equation of motion for the angular motion of the pendulum bob can be written as

$$\ddot{\theta} + [g - A_x(t)]\frac{\sin\theta}{b} = -\frac{A_y\cos\theta}{b}$$

where θ, b, and A are as depicted in Figure 11.2.

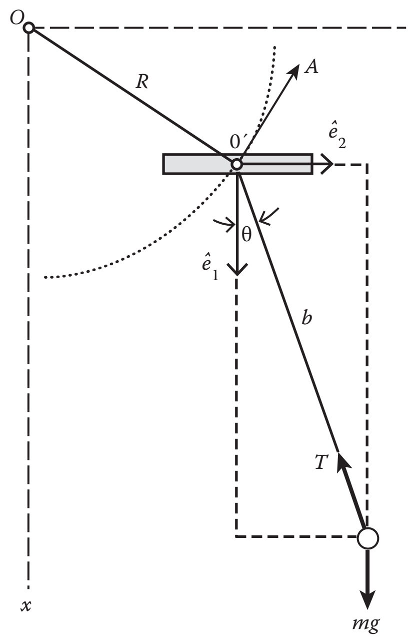


FIGURE 11.2 Pendulum with a support that moves with acceleration *A*.

Solution:

As shown in Figure 11.2, *Oxy* is the unprimed fixed system with the *x*-axis positive downward, and the point of suspension of the simple pendulum lies in the *xy*-plane.

The origin *O*′ of the primed system is placed at the instantaneous location of the point of suspension, and the angular motion of the pendulum bob is restricted to be in the *xy*-plane. The equation of motion of the bob in the moving primed system is

$$m\frac{d^2\vec{r}'}{dt^2} = \vec{T} + mg\hat{e}_1 - m\vec{A} \tag{11.6}$$

where $\vec{T}$ is the tension and $\vec{A}$ is

$$\vec{A} = \frac{d^2\vec{R}}{dt^2} = A_x\hat{e}_1 + A_y\hat{e}_2. \tag{11.7}$$

Because the motion of the bob is confined to the *xy*-plane, it is convenient to write the equation of motion (Equation 11.6) in plane polar coordinates (r, θ), where θ is as depicted in Figure 11.2, and *r* points from the point of suspension to the bob. Using Equation 1.23,

$$\begin{aligned}\ddot{\vec{r}} &= (\ddot{r} - r\dot{\theta}^2)\hat{e}_r + (2\dot{r}\dot{\theta} + r\ddot{\theta})\hat{e}_\theta \\ &= -b\dot{\theta}^2\hat{e}_r + b\ddot{\theta}\hat{e}_\theta\end{aligned}$$

the relationship

$$\begin{aligned}\hat{e}_r &= \hat{e}_1\cos\theta + \hat{e}_2\sin\theta \\ \hat{e}_\theta &= -\hat{e}_1\sin\theta + \hat{e}_2\cos\theta\end{aligned}$$

and also

$$T_\theta = 0,$$

we obtain

$$mb\ddot{\theta} = -m(g - A_x)\sin\theta - mA_y\cos\theta \tag{11.8}$$

$$-mb\dot{\theta}^2 = T_r + m(g - A_x)\cos\theta - mA_y\sin\theta. \tag{11.9}$$

The angular motion of the pendulum bob is determined by Equation 11.8, which can be rewritten as

$$\ddot{\theta} + \left(\frac{g}{b} - \frac{A_x}{b}\right)\sin\theta = -\frac{A_y}{b}\cos\theta. \tag{11.10}$$

We now examine this equation in two different cases:

(1) The point of suspension O' moving horizontally with an acceleration $\ddot{y}$.
We now have $A_x = 0$ and $Ay = \ddot{y}$. Accordingly, Equation 11.10 reduces to

$$\ddot{\theta} + \frac{g}{b}\sin\theta = -\frac{\ddot{y}}{b}\cos\theta. \tag{11.11}$$

For small angular displacement (i.e., $\sin\theta \cong \theta$) and a sinusoidal motion of the support ($y = y_0\cos\omega t$), Equation 11.11 becomes

$$\ddot{\theta} + \omega_0^2\theta = \frac{y_0}{b}\omega^2\cos\omega t$$

where $\omega_0 = (g/b)^{1/2}$. This equation is identical in mathematical form with the undamped forced harmonic oscillator that was treated in Chapter 8 on linear oscillations.

(2) The point of suspension O' moving vertically with a uniform acceleration.
In this case, $A_y = 0$ and $A_x =$ constant k. Equation 11.10 becomes

$$\ddot{\theta} + \frac{g-k}{b}\sin\theta = 0.$$

If $g > k$, the natural frequency of small oscillation is

$$\omega_0 = \sqrt{(g-k)/b}.$$

When $g = k$, $\omega_0 = 0$ and the pendulum falls freely and behaves as if the gravitational field has vanished. This is another example of the principle of equivalence: over a small region of space–time, a non-inertial reference is equivalent to a certain gravitational field. Thus, gravity can be made to disappear (or appear) locally by a suitable coordinate transformation.

11.2 DYNAMICS IN ROTATING COORDINATE SYSTEM

We now consider the motion of a particle P described with respect to a primed system $O'x_1'x_2'x_3'$, which experiences angular velocity $\vec{\omega}$ and translation $\dot{\vec{R}}$ relative to the unprimed system $Ox_1x_2x_3$ as in Figure 11.3. The three unit vectors $\hat{e}_1'$, $\hat{e}_2'$, and $\hat{e}_3'$ associated with axes x_1', x_2', and x_3', respectively, are not indicated in Figure 11.3; they do not change in magnitude but do change in direction as $O'x_1'x_2'x_3'$ rotates. An inspection of Figure 11.3 shows that $\vec{r}$ and $\vec{r}'$ are related by the following equation:

$$\vec{r} = \vec{R} + \vec{r}' = \vec{R} + \sum_{j=1}^{3} x_j'\hat{e}_j'. \tag{11.12}$$

Upon differentiation, we obtain

$$\left(\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right)_0 = \left(\frac{\mathrm{d}\vec{R}}{\mathrm{d}t}\right)_0 + \left(\frac{\mathrm{d}\vec{r}'}{\mathrm{d}t}\right)_0 = \left(\frac{\mathrm{d}\vec{R}}{\mathrm{d}t}\right)_0 + \sum_{j=1}^{3}\left(\frac{\hat{e}_j'\mathrm{d}x_j'}{\mathrm{d}t}\right)_0 + \sum_{j=1}^{3}\left(\frac{x_j'\mathrm{d}\hat{e}_j'}{\mathrm{d}t}\right)_0 \tag{11.13}$$

where the designation 0 is explicitly included to indicate that the quantity in the parentheses is measured in the inertial system $Ox_1x_2x_3$. The second term on the right-hand side is just the velocity of the particle P relative to the rotating system $O'x_1'x_2'x_3'$ observed by someone seated in and rotating with the system:

$$\left(\frac{\mathrm{d}\vec{r}'}{\mathrm{d}t}\right)_{0'} = \sum_{j=1}^{3}\frac{\mathrm{d}x_j'}{\mathrm{d}t}\hat{e}_j' \equiv \frac{\delta\vec{r}'}{\delta t}. \tag{11.14}$$

From now on, we shall reserve the notation "$\delta/\delta t$" for the rotating system $O'x_1'x_2'x_3'$.

The third term on the right-hand side is the apparent velocity of P resulting from the rotation of the $O'x_1'x_2'x_3'$ axes, and it can be put into a more familiar form: $\vec{\omega}\times\vec{r}'$. In order to show this, we first write the angular velocity $\vec{\omega}$ in terms of its components:

$$\vec{\omega} = \sum_{j=1}^{3}\omega_{x_j'}\hat{e}_j' \tag{11.15}$$

where $\omega_{x_1'}$ is the angular velocity with which the x_2' and x_3' axes are rotating about a fixed axis in the x_1'-direction, similarly to $\omega_{x_2'}$ and $\omega_{x_3'}$. $\omega_{x_1'}$ does not change the direction of $\hat{e}_1'$; $\omega_{x_2'}$ produces, in a short time, $\mathrm{d}t$, a change $-\hat{e}_3'\omega_{x_2'}\mathrm{d}t$ in $\hat{e}_1'$ in the negative x_3'-direction as depicted in Figure 11.4. Similarly,

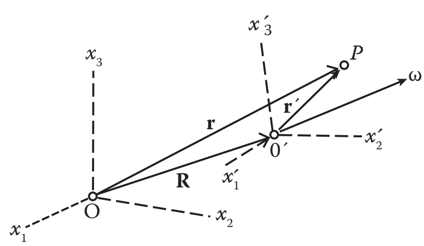


FIGURE 11.3 Translating and rotating systems.

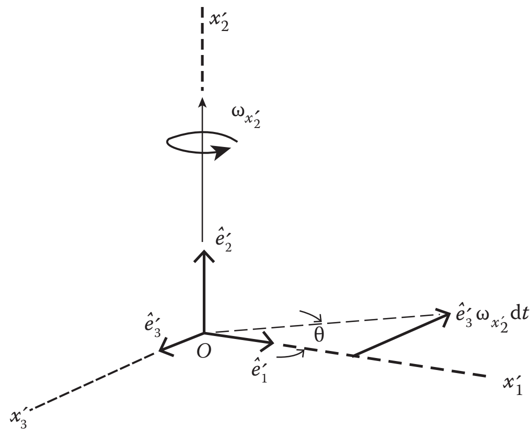


FIGURE 11.4 Rotation about x_2'-direction changes unit vector e_1' in the negative x_1'-direction.

the change in $\hat{e}_1'$ in a short time dt resulting from $\omega_{x_3'}$ is $\hat{e}_2'\omega_{x_3'}\mathrm{d}t$. Thus, the total change in $\hat{e}_1'$ in time dt is $\mathrm{d}\hat{e}_1' = (\hat{e}_2'\omega_{x_3'} - \hat{e}_3'\omega_{x_2'})\mathrm{d}t$ or

$$\frac{\mathrm{d}\hat{e}_1'}{\mathrm{d}t} = \hat{e}_2'\omega_{x_3'} - \hat{e}_3'\omega_{x_2'} = \vec{\omega}\times\hat{e}_1'. \tag{11.16}$$

Similarly,

$$\frac{\mathrm{d}\hat{e}_2'}{\mathrm{d}t} = \hat{e}_3'\omega_{x_1'} - \hat{e}_1'\omega_{x_3'} = \vec{\omega}\times\hat{e}_2' \tag{11.17}$$

$$\frac{\mathrm{d}\hat{e}_3'}{\mathrm{d}t} = \hat{e}_1'\omega_{x_2'} - \hat{e}_2'\omega_{x_1'} = \vec{\omega}\times\hat{e}_3'. \tag{11.18}$$

Substituting Equations 11.16 through 11.18 into the third term on the right-hand side of Equation 11.13, we obtain

$$\sum_{j=1}^{3}\left(\frac{x_j'\mathrm{d}\hat{e}_j'}{\mathrm{d}t}\right)_0 = \vec{\omega}\times\sum_{j=1}^{3}\hat{e}_j'x_j' = \vec{\omega}\times\vec{r}'. \tag{11.19}$$

Alternatively, we can obtain Equation 11.19 in the following way: because we are investigating, from the inertial frame $Ox_1x_2x_3$, the change in $\vec{r}'$ resulting only from rotation itself, we can consider the particle P at rest in the rotating frame. The rotation brings it from $\vec{r}'$ to $\vec{r}' + \mathrm{d}\vec{r}'$. The vector $\mathrm{d}\vec{r}'$, as shown in Figure 11.5, is perpendicular to the plane containing $\vec{r}'$ and the axis of rotation and its length is

$$\mathrm{d}r' = r'\sin\alpha\ \mathrm{d}\phi.$$

These two properties of $\mathrm{d}\vec{r}'$ are correctly represented if we write $\mathrm{d}\vec{r}'$ as an axial vector $\mathrm{d}\vec{r}' = \mathrm{d}\vec{\phi}\times\vec{r}'$.

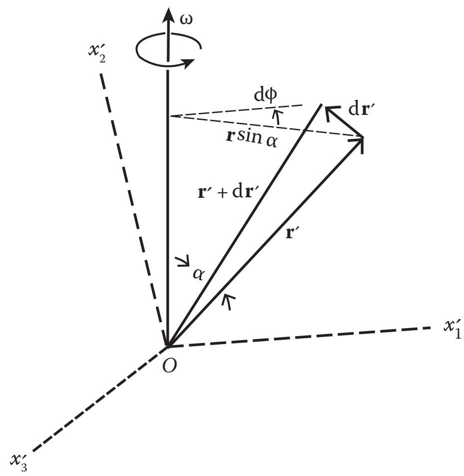


FIGURE 11.5 Infinitesimal rotation dϕ about $\hat{\lambda}$. The vector $\vec{r}'$ is assumed to be fixed in the rotating frame in this construction.

From the last equation, we obtain the time derivative of $\vec{r}'$ resulting from rotation:

$$\left(\frac{\mathrm{d}\vec{r}'}{\mathrm{d}t}\right)_0 = \sum_{j=1}^{3}\left(\frac{x_j'\mathrm{d}\hat{e}_j'}{\mathrm{d}t}\right)_0 = \frac{\mathrm{d}\vec{\phi}}{\mathrm{d}t}\times\vec{r}' = \vec{\omega}\times\vec{r}'$$

which is the same result as is given by Equation 11.19.

Combining Equations 11.19 and 11.14 with Equation 11.13, we find

$$\begin{aligned}\left(\frac{\mathrm{d}\vec{r}'}{\mathrm{d}t}\right)_0 &= \sum_{j=1}^{3}\left(\frac{\hat{e}_j'\mathrm{d}x_j'}{\mathrm{d}t}\right)_0 + \sum_{j=1}^{3}\left(\frac{x_j'\mathrm{d}\hat{e}_j'}{\mathrm{d}t}\right)_0 \\ &= \frac{\delta\vec{r}'}{\delta t} + \vec{\omega}\times\vec{r}' \end{aligned} \tag{11.20}$$

and

$$\begin{aligned}\left(\frac{\mathrm{d}\vec{r}}{\mathrm{d}t}\right)_0 &= \left(\frac{\mathrm{d}\vec{R}}{\mathrm{d}t}\right)_0 + \left(\frac{\mathrm{d}\vec{r}'}{\mathrm{d}t}\right)_0 \\ &= \left(\frac{d\vec{R}}{dt}\right)_0 + \frac{\delta\vec{r}'}{\delta t} + \vec{\omega}\times\vec{r}' \end{aligned} \tag{11.21}$$

We can rewrite Equation 11.21 as

$$\vec{v}_0 = \vec{V} + \vec{v}' + \vec{\omega}\times\vec{r}' \tag{11.22}$$

where $\vec{v}_0$ = velocity of the particle P relative to the inertial system $Ox_1x_2x_3$; $\vec{V} = (\mathrm{d}\vec{R}/\mathrm{d}t)_0$ = velocity of the moving origin relative to the inertial system; $\vec{v}' = \delta\vec{r}'/\delta t$ = velocity of the particle P relative to the rotating system; $\vec{\omega}\times\vec{r}'$ = velocity of the particle P as measured in the system $Ox_1x_2x_3$ resulting from the rotation of the primed system $O'x_1'x_2'x_3'$.

It is important to recognize that the vector Equation 11.20 is a general result. Any arbitrary vector, say, $\vec{Q}$, is used in place of the radius vector $\vec{r}'$, and the form of the result would be the same:

$$\left(\frac{\mathrm{d}\vec{Q}}{\mathrm{d}t}\right)_0 = \frac{\delta\vec{Q}}{\delta t} + \vec{\omega}\times\vec{Q}. \tag{11.23}$$

From this, it follows that the angular acceleration is the same in both coordinate systems:

$$\left(\frac{\mathrm{d}\vec{\omega}}{\mathrm{d}t}\right)_0 = \frac{\delta\vec{\omega}}{\delta t} + \vec{\omega}\times\vec{\omega} = \frac{\delta\vec{\omega}}{\delta t} \equiv \dot{\vec{\omega}}.$$

A higher-order time derivative can easily be evaluated with the help of Equation 11.23. For the second derivative, we have

$$\vec{a}_0 = \left(\frac{\mathrm{d}\vec{v}_0}{\mathrm{d}t}\right)_0 = \left(\frac{\mathrm{d}\vec{V}}{\mathrm{d}t}\right)_0 + \left(\frac{\mathrm{d}\vec{v}_0'}{\mathrm{d}t}\right)_0 + \left(\frac{\mathrm{d}\vec{\omega}}{\mathrm{d}t}\right)_0 \times\vec{r}' + \vec{\omega}\times\left(\frac{\mathrm{d}\vec{r}'}{\mathrm{d}t}\right)_0. \tag{11.24}$$

The usefulness of Equation 11.23 is now apparent. With its help, we have

$$\left.\begin{aligned} (\mathrm{d}\vec{v}'/\mathrm{d}t)_0 &= \delta\vec{v}'/\delta t + \vec{\omega}\times\vec{v}' = \vec{a}' + \vec{\omega}\times\vec{v}' \\ (\dot{\vec{\omega}})_0\times\vec{r}' &= \frac{\delta\vec{\omega}}{\delta t}\times\vec{r}' + (\vec{\omega}\times\vec{\omega})\times\vec{r}' = \dot{\vec{\omega}}\ \times\vec{r}' \\ \vec{\omega}\times(\mathrm{d}\vec{r}/\mathrm{d}t)_0 &= \vec{\omega}\times\vec{v}' + \vec{\omega}\times(\vec{\omega}\times\vec{r}') \end{aligned}\right\}. \tag{11.25}$$

Substituting Equation 11.25 into Equation 11.24, we obtain

$$\vec{a}_0 = \vec{A} + \vec{a}' + 2\vec{\omega}\times\vec{v}' + \vec{\omega}\times(\vec{\omega}\times\vec{r}') + \dot{\vec{\omega}}\times\vec{r}' \tag{11.26}$$

where $\vec{a}_0 = (\mathrm{d}\vec{v}_0/\mathrm{d}t)_0$ is the acceleration of the particle P relative to the inertial system $Ox_1x_2x_3$; $\vec{A} = (\mathrm{d}\vec{V}/\mathrm{d}t)_0$ is the acceleration of the moving origin O' relative to the inertial system $Ox_1x_2x_3$; $2\vec{\omega}\times\vec{v}'$ = Coriolis acceleration, which is present only when the particle P is moving in the rotating frame, but it is independent of the position of the particle relative to the rotating frame; $\dot{\vec{\omega}}\times\vec{r}'$ = acceleration of the particle P resulting from the angular acceleration of the rotating axes (it is called the transverse or azimuthal acceleration); $\vec{\omega}\times(\vec{\omega}\times\vec{r}')$ = centripetal acceleration of the particle P resulting from the rotation of the moving axes about an axis of rotation (it is present even when the particle is at rest relative to the rotating frame); and $\vec{a}' = \delta\vec{v}'/\delta t$ is the acceleration of the particle P relative to the rotating system $O'x_1'x_2'x_3'$.

Solving Equation 11.26 for $\vec{a}'$ and multiplying by m, mass of the particle, we find

$$m\vec{a}' = \vec{F} - m\left[\vec{A} + 2\vec{\omega}\times\vec{v}' + \vec{\omega}\times(\vec{\omega}\times\vec{r}') + \dot{\vec{\omega}}\times\vec{r}'\right] \tag{11.27}$$

where $\vec{F}$ is the physical force acting on the particle P, and $\vec{F} = m\vec{a}_0$; because the unprimed system $Ox_1x_2x_3$ is assumed to be an inertial system, Newton's law of motion holds in it.

It is obvious from Equation 11.27 that Newton's second law does not take the familiar simple form any more. Now, the physical force $\vec{F}$ is modified by the presence of fictitious (or pseudo or inertial) forces:

Coriolis force $\vec{F}_{cor} = -2m\vec{\omega} \times \vec{v}'$
centrifugal force $\vec{F}_{cf} = -m\vec{\omega} \times (\vec{\omega} \times \vec{r}')$
transverse (or azimuthal) force $\vec{F}_{az} = -m\dot{\vec{\omega}} \times \vec{r}'$
translational force $\vec{F}_{tr} = -m\vec{A}$. (11.28)

These are not physical forces and can be altered or made to vanish merely by the choice of a reference system. We now take a close look at these fictitious forces.

11.2.1 Centrifugal Force

Because $\vec{\omega} \cdot \vec{F}_{cf} = 0$, the centrifugal force is perpendicular to the axis of rotation and moves radially away from the axis of rotation as depicted in Figure 11.6. Its magnitude is

$$\left| m\vec{\omega} \times (\vec{\omega} \times \vec{r}') \right| = m\omega^2 r' \sin\theta = m\omega^2 \rho \tag{11.29}$$

where

$$\rho = r' \sin\theta.$$

It can be shown that centrifugal force corresponds to a potential energy $(m/2)[(\vec{\omega} \cdot \vec{r}')^2 - \omega^2 r'^2]$. To show this, we use the vector identity $\vec{A} \times (\vec{B} \times \vec{C}) = (\vec{A} \cdot \vec{C})\vec{B} - (\vec{A} \cdot \vec{B})\vec{C}$ to rewrite $\vec{F}_{cf}$:

$$\vec{F}_{cf} = -m\vec{\omega} \times (\vec{\omega} \times \vec{r}') = -m[(\vec{\omega} \cdot \vec{r}')\vec{\omega} - \omega^2 \vec{r}'] \tag{11.30}$$

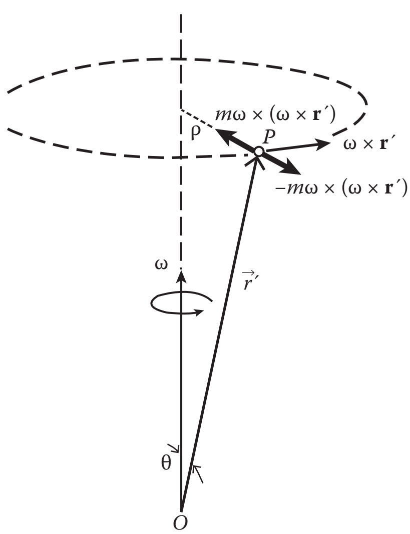


FIGURE 11.6 Centrifugal force.

This expression can be verified as the negative gradient of a scalar function V_{cf}:

$$\vec{F}_{cf} = -\nabla V_{cf} \tag{11.31}$$

with

$$V_{cf} = \frac{m}{2}[(\vec{\omega}\cdot\vec{r}')^2 - \omega^2 r'^2]. \tag{11.32}$$

It should be noted that this potential is no less fictitious than the fictitious force $\vec{F}_{cf}$ itself.

Example 11.2: The Conical Pendulum

The vertical rod *AB* as shown in Figure 11.7 is rotating with constant angular velocity $\vec{\omega}$. A massless inextensible string of length *b* is attached to the rod at point *A*, and its other end has a mass *m* attached. Find (1) the tension *T* in the string and (2) the angle θ that the string makes with the rod when the particle is in equilibrium.

Solution:

The unit vectors $\hat{i}$ and $\hat{k}$, as shown in Figure 11.7, rotate with the rod. The position vector $\vec{r}$ of *m* points from *A* to *m*:

$$\vec{r} = b\sin\theta\hat{i} - b\cos\theta\hat{k}.$$

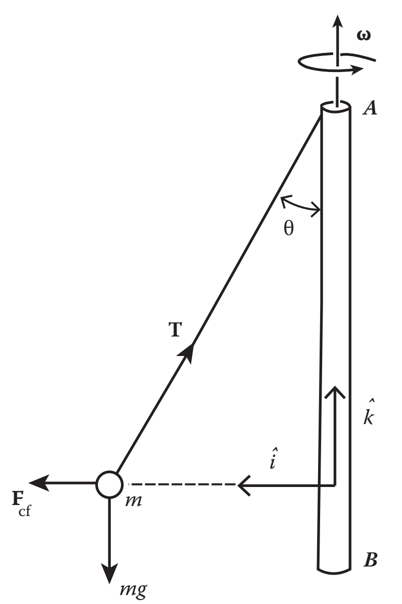


FIGURE 11.7 Conical pendulum.

There are three forces acting on the particle:

(1) The weight of the particle $\vec{W} = -mg\hat{k}$
(2) The tension in the string $\vec{T} = -T\sin\theta\hat{i} + T\cos\theta\hat{k}$
(3) The centrifugal force

$$-m[\vec{\omega}\times(\vec{\omega}\times\vec{r})] = -m\left\{(\omega\hat{k})\times\left[(\omega\hat{k})\times(b\sin\theta\hat{i} - b\cos\theta\hat{k})\right]\right\}$$
$$= m\omega^2 b\sin\theta\hat{i}.$$

When the particle is in equilibrium, the net force acting on it is zero:

$$-mg\hat{k} - T\sin\theta\hat{i} + T\cos\theta\hat{k} + m\omega^2\sin\theta\hat{i} = 0$$

or, by rearrangement,

$$(m\omega^2 b\sin\theta - T\sin\theta)\hat{i} + (T\cos\theta - mg)\hat{k} = 0$$

from which it follows that

$$m\omega^2\, b\sin\theta - T\sin\theta = 0$$

$$T\cos\theta - mg = 0.$$

Solving these two equations, we find

$$T = m\omega^2 b$$

and

$$\theta = \cos^{-1}(g/\omega^2 b).$$

Example 11.3: Surface of a Rotating Liquid

A bucket of liquid spins with angular speed ω. Find the shape of the liquid's surface.

Solution:

In a coordinate system rotating with the bucket, the liquid is at rest, and $\vec{v}' = 0$. Using Equation 11.27, the equation of motion of a small mass m of the liquid on the surface is

$$m\frac{\delta^2\vec{r}'}{\delta t^2} = \vec{F}' + m\vec{g} - m\vec{\omega}\times(\vec{\omega}\times\vec{r}') \tag{11.33}$$

where the Coriolis and the transverse terms have been dropped because of the vanishing of $\dot{\vec{\omega}}$ and $\vec{v}'$, and the force $\vec{F}'$, being the pressure gradient, is normal to the surface. At equilibrium, $\delta^2\vec{r}'/\delta t^2 = 0$, and Equation 11.33 reduces to

$$\vec{F}' + m\left[\vec{g} - \vec{\omega}\times(\vec{\omega}\times\vec{r}')\right] = 0. \tag{11.34}$$

The quantity in the square brackets can be considered as the effective gravity

$$\vec{g}_{eff} = \vec{g} - \vec{\omega} \times (\vec{\omega} \times \vec{r}'). \tag{11.35}$$

In terms of ρ, $\vec{g}_{eff}$ is given by

$$\vec{g}_{eff} = -g\hat{e}_3 + \rho\omega^2\hat{e}_\rho \tag{11.36}$$

where $\hat{e}_\rho$ is a unit vector along ρ. Because $\vec{F}'$ is normal to the liquid surface, by Equation 11.34, $\vec{g}_{eff}$ is also normal to the liquid surface.

From the geometry of Figure 11.8, we see that

$$\tan\theta = \frac{dx_3'}{d\rho} = \frac{m\rho\omega^2}{mg}. \tag{11.37}$$

Integration gives

$$x_3' = \frac{\omega^2}{2g}\rho^2 + \text{constant} = \left(\frac{\omega^2}{2g}\right)(x_1'^2 + x_2'^2) + \text{constant} \tag{11.38}$$

This is a paraboloid of revolution about the x_3' axis.

The solution in Equation 11.38 can also be found with the help of energy. Recall that the centrifugal force corresponds to a potential energy V_{cf}:

$$V_{cf} = \frac{m}{2}[(\vec{\omega} \cdot \vec{r}')^2 - \omega^2 r'^2].$$

If the angular velocity $\vec{\omega}$ is chosen to lie along the x_3' axis of the rotating frame, the centrifugal potential then reduces to a simpler form

$$V_{cf} = -\frac{m}{2}\left(x_1'^2 + x_2'^2\right)\omega^2.$$

Now, consider a particle of mass m on the liquid surface. Unless the liquid surface is an equipotential surface, the particle tends to move toward regions of lower potential energy. Hence, if

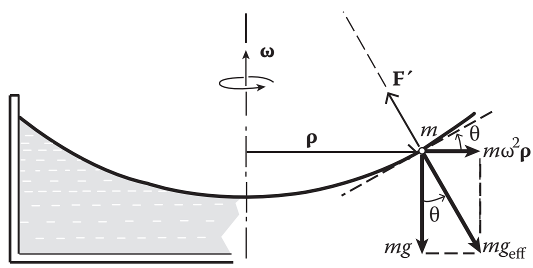


FIGURE 11.8 Parabolic surface of a spinning pail of water.

the liquid is in equilibrium under the gravitational and centrifugal forces, its surface must be an equipotential surface:

$$gx_3' - \frac{1}{2}\left(x_1'^2 + x_2'^2\right)\omega^2 = \text{constant}$$

which is the same result as given by Equation 11.38.

11.2.2 The Coriolis Force

The Coriolis force is of special interest. It is at right angles to the axis of rotation and to the velocity vector of the particle in the rotating frame. And it vanishes when the particle is at rest in the rotating frame so that $\vec{v}' = 0$. We can best see this by looking at the following simple example.

Example 11.4: Physics on a Rotating Table

A horizontal table rotates uniformly counterclockwise with an angular velocity $\vec{\omega}$ with respect to an inertial frame. An object of mass *m* moves on the table with velocity $\vec{v}$. Assuming no friction between the table and the object, what is the object's path as seen in the rotating frame?

Solution:

The origins of the inertial and rotating frames are fixed, so $\vec{A} = 0$. The angular velocity has the constant value

$$\vec{\omega} = \omega\hat{e}_3$$

implying that $\dot{\vec{\omega}} = 0$, and so Equation 11.27 reduces to

$$\vec{a} = \vec{F}/m - 2\vec{\omega}\times\vec{v} - \vec{\omega}\times(\vec{\omega}\times\vec{r}).$$

It should be noted that $\vec{a}$ and $\vec{v}$ are $\vec{a}'$ and $\vec{v}'$ of Equation 11.27. For the object moving on the table, both $\vec{r}$ and $\vec{v}$ lie in the plane of the table, perpendicular to $\vec{\omega}$, and the Coriolis term

$$-2\vec{\omega}\times\vec{v} = -2\omega\hat{e}_3\times\vec{v}$$

always generates an acceleration perpendicular to $\vec{v}$. The acceleration, in this instance, is to the "right" if we look along the object's path, following the motion. The centripetal acceleration term on the right-hand side may be expanded as

$$\vec{\omega}\times(\vec{\omega}\times\vec{r}) = -[(\vec{\omega}\cdot\vec{r})\vec{\omega} - \omega^2\vec{r}] = +\omega^2\vec{r}$$

when $\vec{r}$ is perpendicular to $\vec{\omega}$, as it is here. This term generates a radially outward acceleration. Some paths are shown in Figure 11.9. As seen from the inertial frame, every one of these paths is a straight line!

The problem of an ice skater or an equivalent, such as a student sitting on a rotating stool, can also be treated in terms of the Coriolis force. The usual argument goes this way: In the absence of external torque, angular momentum is conserved, so if the moment of inertia *I* is increased or decreased, the angular velocity $\vec{\omega}$ is correspondingly decreased or increased to hold the angular

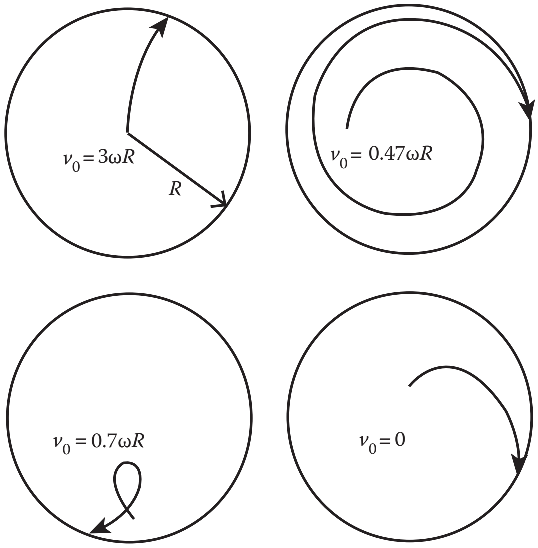


FIGURE 11.9 Free-force motion as seen in the rotating frame. The initial speed is indicated in each drawing.

momentum $\vec{L} = I\vec{\omega}$ constant. If desired, we can ask what the force is that causes the change in angular velocity. The forces involved in moving the skater's arms inward or outward cannot cause the change in angular velocity. However, as the skater moves his or her arms outward or inward, the radial velocity of his or her arms gives rise to the fictitious Coriolis force that acts tangentially (i.e., at right angles to his or her arms). Thus, it is the Coriolis force that causes the change in angular velocity:

$$\vec{F}_{cor} = -2m\vec{\omega} \times \vec{v}' = -\nabla V_{cor}$$

with

$$\begin{aligned} V_{cor} &= 2m\vec{\omega} \cdot (\vec{r}' \times \vec{v}') = 2\vec{\omega} \cdot (\vec{r}' \times m\vec{v}') \\ &= 2\vec{\omega} \cdot (\vec{r}' \times \vec{p}') = 2\vec{\omega} \cdot \vec{L}' \end{aligned} \tag{11.39}$$

where $\vec{L}' = \vec{r}' \times \vec{p}'$ is the skater's angular momentum in the rotating frame.

The Coriolis force is important in describing the motion of a body relative to the rotating Earth. In the Northern Hemisphere, where $\vec{\omega}$ leads upward from the North Pole, the Coriolis force deflects a horizontally moving particle to the right of its path relative to the Earth. In the Southern Hemisphere, a horizontally moving particle is deflected to the left of its original direction of motion (Figure 11.10). Because the magnitude of the horizontal component of the Coriolis force is proportional to the vertical component of $\vec{\omega}$, the horizontal deflection depends upon the latitude, being maximum at the poles and zero at the equator. Because of the Coriolis effect, it can be inferred that, other factors being equal, rivers in the Northern Hemisphere will undercut their right banks more than their left and observation bears this out, with the reverse being true in the Southern Hemisphere.

The Coriolis force also deflects a falling body. The particle's velocity is nearly vertical, so $\vec{\omega}$ lies in the north–south direction, and the force $-2m\vec{\omega} \times \vec{v}'$ is in the east–west direction. A body falling toward the Earth's surface will be deflected to the east. A detailed study of the motion of a particle near the Earth's surface can be found in Section 11.3.

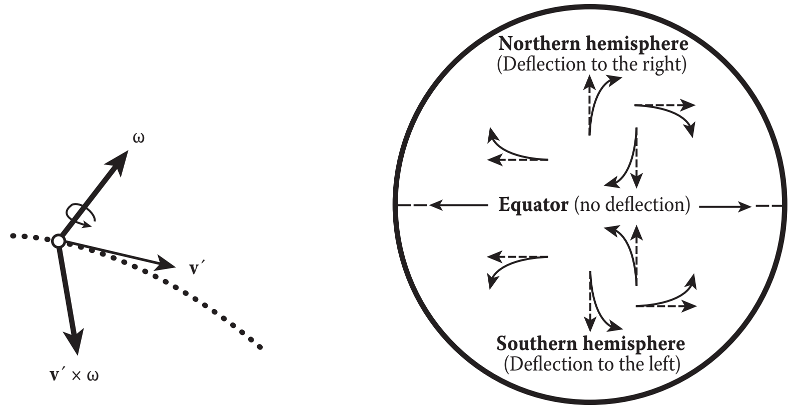


FIGURE 11.10 Coriolis deflection.

11.2.2.1 Trade Winds and Circulation of Ocean Currents

Coriolis effects are also responsible for the circulation of air around high- and low-pressure areas on the Earth's surface. Although the Coriolis force is very small, the time during which it can act on moving air masses is very long. Thus, it can make an appreciable change in the momentum of the air. Without the whirling Coriolis effect, the air masses would flow straight from high-pressure regions to low-pressure regions along the direction of the pressure gradient. The Coriolis forces deflect the air masses to the right of this direction in the Northern Hemisphere and produce counterclockwise cyclonic motion (Figure 11.11).

The overall pattern of the Earth's winds is also influenced by the Coriolis force. Hot air rising all along the equator gets displaced into higher and cooler latitudes, where it sinks to ground level and spreads north and south. Because of the influence of the Coriolis force, the streams that spread away from the equator are deflected to the east as westerlies and Roaring Forties, and those that spread toward the equator are deflected to the west as northeast and southeast trade winds as shown in Figure 11.12a.

The pictures from the *Voyager I* and *II* flybys of Jupiter moons show more clearly than ever before that the Jovian atmosphere is segregated into definite parallel bands of material. The formation of these parallel bands is also a result of Coriolis effects. Jupiter rotates once in about 12 h, so Jovian Coriolis effects are more pronounced than their terrestrial counterparts. Vertical convection

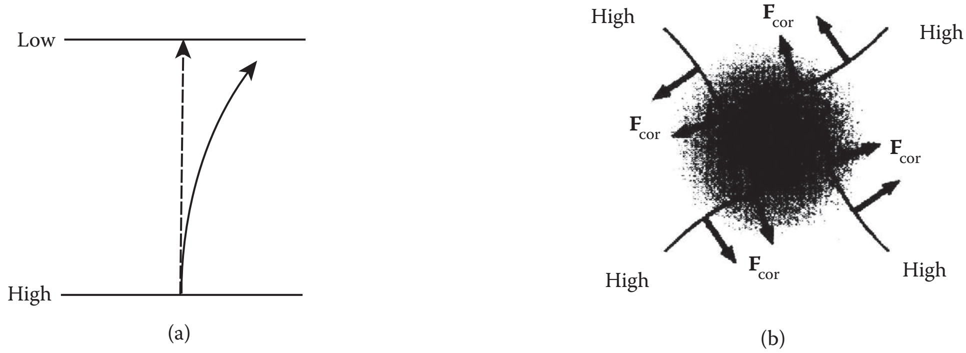


FIGURE 11.11 (a) The Coriolis force deflects air masses to the right and (b) counterclockwise cyclonic motion in the northern hemisphere.

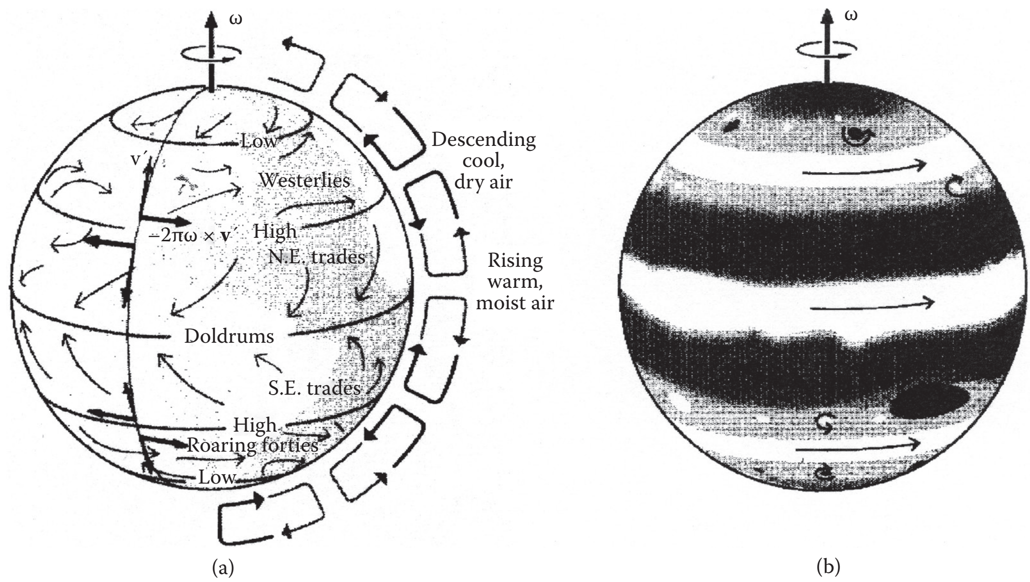


FIGURE 11.12 (a) Overall pattern of the Earth's winds and (b) Sketch of Jupiter.

currents on the planet are Coriolis forces diverted into horizontal (or east–west) trajectories. These jet streams are extreme versions of the less-pronounced east–west trades on Earth (Figure 11.12b).

The Earth's rotation, the Earth's winds, and the sun's radiant energy are the primary factors that control the circulation of ocean currents. This combination produces, predictably, a general pattern of clockwise surface currents in the Northern Hemisphere and counterclockwise surface currents in the Southern Hemisphere (Figure 11.13).

11.2.2.2 Weather Systems

The closed curves in the sketch of Figure 11.14 represent lines of constant pressure or isobars. In the absence of the Coriolis force, winds would blow inward, quickly equalizing the pressure difference. However, the Coriolis force markedly alters the wind pattern. As the air begins to flow inward, it is deflected sideways by the Coriolis force and circulates counterclockwise about the low-pressure area along the isobars. To analyze the motion, we consider the forces on a small parcel of air mass that is rotating about a low-pressure area as shown in Figure 11.14. If the pressure along the inner isobar is p_1 and that along the outer isobar is $p_1 + \Delta p$, then the net pressure force is $(p_1 + \Delta p - p_1)S = (\Delta p)S$ inward, where S is the area of the small air parcel between the two isobars. The Coriolis force is $2mv\omega \sin\lambda$, where m is the mass of the parcel, v is its velocity, λ is its latitude, and ω is the

FIGURE 11.13 General pattern of clockwise surface currents in the northern hemisphere and counterclockwise surface currents in the southern hemisphere.

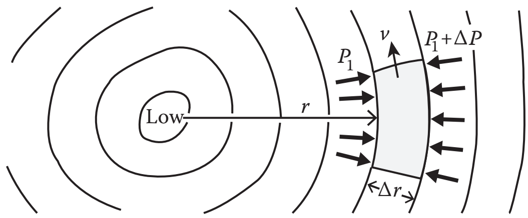


FIGURE 11.14 Forces on a small parcel of air mass that is rotating about a low-pressure area.

angular velocity of the rotating Earth. The air is rotating counterclockwise about the low-pressure area, so the Coriolis force is outward. Hence, the radial equation of motion for steady circular flow around a low pressure area is

$$mv^2/r = (\Delta p)S - 2mv\omega \sin\lambda. \tag{11.40}$$

The mass of the parcel is $\rho(\Delta r)S$, where ρ is the air density, assumed constant. Inserting this into the equation of motion and taking the limit $\Delta r \to 0$ gives

$$\frac{v^2}{r} = \frac{1}{\rho}\frac{\mathrm{d}p}{\mathrm{d}r} - 2v\omega \sin\lambda.$$

Near the center of the low, the pressure gradient dp/dr is large and wind velocities are highest. Far from the center, v^2/r is small and can be ignored. Hence, Equation 11.40 predicts that far from the center, the wind speed is

$$v = \frac{1}{2\rho\omega \sin\lambda}\frac{\mathrm{d}p}{\mathrm{d}r}. \tag{11.41}$$

At sea level, $\rho \cong 1.3$ kg/m^3, and far from low, a typical pressure gradient is about 3 mbar over 120 km: dp/d$r \cong 3 \times 12^{-3}$ (dp/dr can be estimated by looking at a weather map). Equation 11.41 now reduces to

$$v = \frac{80}{\sin\lambda}\,\text{km/h}.$$

This speed is reduced near the ground by friction with the land, but at higher altitudes, Equation 11.41 can be applied with good accuracy.

The pressure gradient at the center of a typhoon or a hurricane can be as high as 30×12^{-3} N/m^3. Winds are so strong that the v^2/r term in Equation 11.40 cannot be ignored. Solving Equation 11.40 for v, we obtain

$$v = \sqrt{(r\omega \sin\lambda)^2 + \frac{r}{p}\frac{\mathrm{d}p}{\mathrm{d}r}} - r\omega \sin\lambda. \tag{11.42}$$

At a distance of 120 km from the eye of a hurricane at latitude 20°, Equation 11.42 gives a wind speed of 160 km/h for a pressure gradient of 30×12^{-3} N/m^3.

Storms such as hurricanes or typhoons are always low-pressure systems. Can a storm be a high-pressure system? The answer is no. In a low, the pressure force is inward, and the Coriolis force is outward, whereas in a high, the directions of the forces are reversed. Equation 11.40 now becomes

$$\frac{v^2}{r} = 2v\omega\sin\lambda - \frac{1}{\rho}\left|\frac{\mathrm{d}p}{\mathrm{d}r}\right|$$

from which we obtain

$$v = r\omega\sin\lambda - \sqrt{(r\omega\sin\lambda)^2 + \frac{r}{p}\frac{\mathrm{d}p}{\mathrm{d}r}}.$$

Thus, if $(1/\rho)\ |\mathrm{d}p/\mathrm{d}r| > r(\omega\ \sin\lambda)^2$, the Coriolis force is too weak to supply the needed centripetal force against the large outward pressure force, and a high-pressure system cannot form.

11.2.2.3 Hurricanes

Hurricanes are intense rotating systems of airflow in the Earth's atmosphere whose direction is always cyclonic, that is, counterclockwise in the Northern Hemisphere and clockwise in the Southern. Why are intense, large-scale storms always cyclonic? Above the Earth's boundary layer, the airflow tends to be parallel to the isobars, the lines of constant barometric pressure. This movement of air, which is called the geostrophic wind, results from a balance between the pressure gradient force $(1/\rho)\nabla p$ and the Coriolis force. In order for these forces to be in balance, the direction of the flow must be cyclonic around a low-pressure center and anticyclonic around a high-pressure center. Frictional forces in the atmospheric boundary layer cause the airflow to be directed inward toward an area of low pressure and outward from a region of high pressure. As the inward-flowing air approaches a low-pressure center, it turns upward, and as it rises to higher altitudes, it tends to cool. However, when the rising air is moist, the energy released by condensation will raise its temperature. The result is a convective updraft with an associated increase in the rate of flow into the system.

11.2.2.4 Bathtub Vortex and Earth Rotation

The draining of a liquid through an opening in the bottom of its container is normally accompanied by the development of vigorous rotational motion, a phenomenon often referred to as the "bathtub vortex." It is a phenomenon that has intrigued scientists for years. In 1908, Ottokar Tumlirz attempted a proof of the Earth's rotation by calculating the directions of streamlines in fluid draining from a tub. The experimental results, however, were inconclusive. In recent years, we have made considerable progress toward explaining the bathtub vortex and similar examples of rotating fluid.

Certain aspects of the vortex are easy to understand. The increase in rotational velocity as the radius of the motion decreases is a familiar consequence of conservation of angular momentum. (An analogy is the increase in the rate of spin of an ice skater who reduces his or her radius of gyration by drawing his or her arms and legs close to his or her body, the axis of rotation.) As the rotational speed of the liquid in a bathtub vortex increases, a dimple appears on the surface that may deepen and break through to the opening in the container as shown in Figure 11.15. A related but less well-known phenomenon occurs at the inlets of jet engines. Here, the flow of the fluid into the compressor leads to the formation of a vortex whose core may bend and reach the ground. Surface debris caught up in such a vortex has led to engine damage.

It is when the direction of the bathtub vortex is discussed that controversy arises. Do the same forces that govern the rotation of hurricanes also control the direction of the bathtub vortex? The answer, as with many questions, can be yes or no.

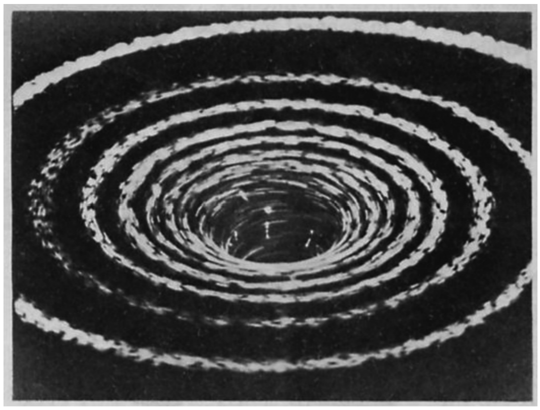


FIGURE 11.15 Bathtub vortex.

Because the bathtub vortex is a manifestation of conservation of angular momentum, its direction is determined by the initial direction of motion of the fluid in an inertial coordinate system, that is, a coordinate system fixed in space. A circular tub fixed to the Earth, however, is itself rotating with an angular velocity $\Omega = \omega \sin\lambda$, where ω is the Earth's rotation, and λ is the local latitude. An element of fluid that is at rest relative to the tub at radius r, therefore, has an absolute velocity $v_0 = r\Omega$. The magnitude of ω is 7.3×10^{-5} rad/s (i.e., one revolution per day), which means that at mid-latitude, where $\lambda \approx 45°$, $\omega \approx 5 \times 10^{-5}$ rad/s, and if the radius is, say, 1 m, the absolute velocity is 5×10^{-5} m/s. The direction of v_0 is counterclockwise in the Northern Hemisphere and clockwise in the Southern. Because the velocity in this case is exceedingly small, it is generally assumed that the direction of a bathtub vortex is an accidental consequence of the residual motion resulting from the method of filling the tub.

If sufficient care were taken to reduce the residual velocities to a value less than v_0, it should be possible to deduce which hemisphere one is located in from the direction of rotation of the bathtub vortex. Experiments of this type were performed in the Northern Hemisphere by Shapiro in 1962 at Cambridge, MA, and then by Trefethen and his colleagues in 1965 in the Southern Hemisphere at Sydney, Australia. In both cases, circular, flat-bottomed tubs 6 ft. in diameter were filled with 6 in. of water. The water inlets were arranged so that the direction of rotation of the water during filling was clockwise in the Northern Hemisphere and counterclockwise in the Southern. Care was taken to minimize thermal gradients to prevent convective motions in the fluid, and the tub was covered to prevent motion resulting from random air currents. After settling times—the interval between filling and drainage—on the order of 24 h, a drain plug 3/8 in. in diameter was removed, allowing the water to flow out over a period of some 20 min. For the first 10 min, no rotational motion was apparent. After this time, however, a small float at the center of the tub began to rotate, and—after all the precautions described above had been taken—the direction of rotation was consistently found to be counterclockwise in the Northern Hemisphere and clockwise in the Southern.

The rotational speed of the float was found to increase with the settling time up to a maximum of one revolution in 3 to 4 s. This result is what one would expect if one applies the equation for conservation of angular momentum. $L_0 = mr\Omega^2$ = constant, where L_0 is the angular momentum of the element of liquid and m is its mass, to an element of liquid whose radius is reduced from 3 ft. to 3/16 in.

11.3 MOTION OF PARTICLE NEAR THE SURFACE OF THE EARTH

The most important application of Equation 11.27 is to a particle moving near the Earth's surface. If we neglect the acceleration of the Earth's center, a set of nonrotating axes with its origin fixed at the

Earth's center can be regarded approximately as an inertial system. A coordinate system attached to the Earth's surface and rotating with the Earth is not an inertial frame (Figure 11.16). In this rotating system, the equation of motion for a particle of mass m moving under gravity and other physical force $\vec{F}$ is given by Equation 11.27, with an extra term $m\vec{g}$ on the right-hand side:

$$m\frac{\delta^2\vec{r}'}{\delta t^2} = m\vec{g} + \vec{F} - m\left[\vec{A} + 2\vec{\omega}\times\dot{\vec{r}}' + \vec{\omega}\times(\vec{\omega}\times\vec{r}')\right]. \tag{11.43}$$

Now,

$$\vec{V} = (\mathrm{d}\vec{R}/\mathrm{d}t)_0 = \vec{\omega}\times\vec{R}, \quad \vec{A} = (\mathrm{d}\vec{V}/\mathrm{d}t)_0 = \vec{\omega}\times(\vec{\omega}\times\vec{R}).$$

Substituting this into Equation 11.43, we obtain

$$\begin{aligned} m\frac{\delta^2\vec{r}'}{\delta t^2} &= m\vec{g} + \vec{F} - m\vec{\omega}\times\left[\vec{\omega}\times(\vec{R}+\vec{r}')\right] - 2m\vec{\omega}\times\dot{\vec{r}}' \\ &= \vec{F} + m\left[\vec{g} - \vec{\omega}\times(\vec{\omega}\times\vec{r})\right] - 2m\vec{\omega}\times\dot{\vec{r}}' \end{aligned} \tag{11.44}$$

where we have grouped the gravitational and centripetal terms in the square brackets. Both are slowly varying functions of position, proportional to the mass of the particle. When we make a measurement of the gravitational acceleration at P, what we actually measure is not $\vec{g}$ but the quantity in the square brackets. Thus, we may define the effective gravitational acceleration $\vec{g}_{eff}$ at the point P by

$$\vec{g}_{eff} = \vec{g} - \vec{\omega}\times(\vec{\omega}\times\vec{r}). \tag{11.45}$$

A plumb line does not point directly toward the Earth's center but is swung outward through a small angle α by the centrifugal force and points in the direction of $\vec{g}_{eff}$ as shown in Figure 11.17. If the motion is near the surface, $\vec{r} = \vec{R} + \vec{r}' \cong \vec{R}$, and $\vec{g}_{eff}$ becomes

$$\vec{g}_{eff} = \vec{g} - \vec{\omega}\times(\omega\times\vec{R}). \tag{11.45a}$$

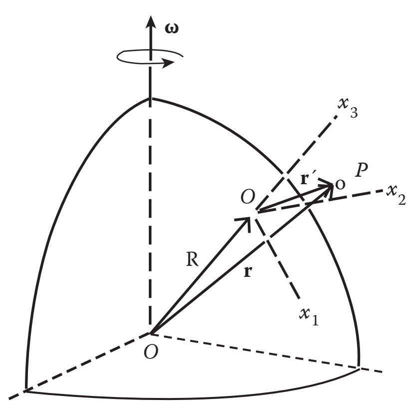


FIGURE 11.16 Rotating frame at the surface of the earth.

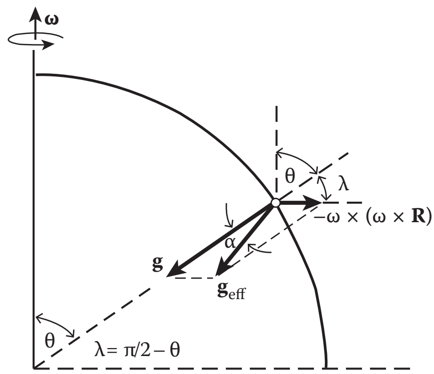


FIGURE 11.17 Effective gravitational acceleration.

The equation of motion now takes the form

$$m\frac{\delta^2\vec{r}'}{\delta t^2} = \vec{F} + m\vec{g}_{eff} - 2m\vec{\omega}\times\dot{\vec{r}}' \qquad (11.46)$$

with $\vec{g}_{eff}$ given by Equation 11.45. The magnitude of the correction term to $\vec{g}$ resulting from centrifugal force is, for $r \ll R$,

$$\left|\vec{\omega}\times(\vec{\omega}\times\vec{R})\right| = \omega^2 R\sin\theta$$

where θ is the co-latitude angle (π/2 – latitude) between $\vec{\omega}$ and $\vec{r} = \vec{R}+\vec{r}' \cong \vec{R}$. Now,

$$\omega = \frac{2\pi}{\tau} = \frac{2\pi}{24\times 3600} = 0.73\times 10^{-4}\ \text{rad/s, and } R = 6{,}371{,}000\ \text{m.}$$

And we find that the correction term is quite small:

$$\omega^2 R \sin\theta = 0.03 \sin\theta\ \text{m/s}^2.$$

The angle α between $\vec{g}$ and $\vec{g}_{eff}$ is given by

$$\alpha \cong \tan\alpha = g_{eff}^h / g_{eff}^v$$

where g_{eff}^h and g_{eff}^v are the horizontal and vertical components, respectively, of $\vec{g}_{eff}$ (using "vertical" to mean toward the Earth's center), and they are given by

$$g_{eff}^h = \omega^2 R\sin\theta\sin\lambda = \omega^2 R\sin\theta\sin(\pi/2-\theta) = \omega^2 R\sin\theta\cos\theta$$

$$g_{eff}^v = g - \omega^2 R\sin\theta\cos(\pi/2-\theta) = g - \omega^2 R\sin^2\theta.$$

Thus,

$$\alpha \cong \tan\alpha = g_{eff}^{h}/g_{eff}^{v} = \omega^2 R\sin\theta\cos\theta/g.$$

The maximum value occurs at $\theta = 45°$ and is about 0.6′.

An object released near the Earth's surface will fall in the direction of $\vec{g}_{eff}$. This is why the Earth has the form of an oblate ellipsoid, flattened at the poles. The degree of flattening is just to make the Earth's surface perpendicular to $\vec{g}_{eff}$ at every point (ignoring local irregularities).

Example 11.5: Projectile Motion over a Rotating Earth

Consider a body moving freely in the Earth's gravitational field, and we observe it from a position on the Earth's surface at latitude λ. We select a coordinate system attached to the earth: The x'-axis is positive to the east; the y'-axis is positive to the north; and the z'-axis is positive upward, as shown in Figure 11.18. The unit vectors associated with the coordinate axes are denoted by $\hat{i}'$, $\hat{j}'$, and $\hat{k}'$.

If the body falls a sufficiently small distance, g remains approximately constant during the process. We also assume that the air resistance is negligible (i.e., $\vec{F} = 0$). The equation of motion of the falling body is then given by

$$\frac{\delta^2\vec{r}'}{\delta t^2} = \vec{g}_{eff} - 2\vec{\omega}\times\dot{\vec{r}}'. \tag{11.47}$$

With the selected coordinate system, we have

$$\vec{\omega} = \omega\cos\lambda\,\hat{j}' + \omega\sin\lambda\hat{k}', \quad \vec{g}_{eff} = -g_{eff}\hat{k}' \tag{11.48}$$

and

$$\vec{\omega}\times\dot{\vec{r}}' = \begin{vmatrix} \hat{i}' & \hat{j}' & \hat{k}' \\ 0 & \omega\cos\lambda & \omega\sin\lambda \\ \dot{x}' & \dot{y}' & \dot{z}' \end{vmatrix}$$
$$= \hat{i}'(\omega\dot{z}'\cos\lambda - \omega\dot{y}'\sin\lambda) + \hat{j}'\omega\dot{x}'\sin\lambda - \hat{k}'\omega\dot{x}'\sin\lambda.$$

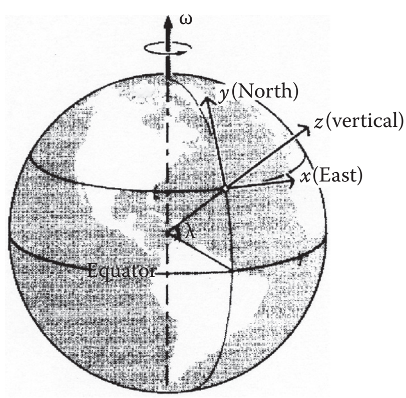


FIGURE 11.18 Rotating frame fixed at the surface of the earth at latitude λ.

The equation of motion becomes, in component forms,

$$\ddot{x}' = -2\omega(\dot{z}'\cos\lambda - \dot{y}'\sin\lambda) \tag{11.49}$$

$$\ddot{y}' = -2\omega\dot{x}'\sin\lambda \tag{11.50}$$

$$\ddot{z}' = -g_{eff} + 2\omega\dot{x}'\cos\lambda. \tag{11.51}$$

The velocity $\vec{v}'$ is, at all times, nearly downward for the falling body. Then, compared with $\dot{z}'$, we can safely neglect terms in $\dot{x}'$ and $\dot{y}'$ from Equations 11.49 through 11.51. But we shall keep them for the moment.

Integrating Equations 11.49 through 11.51 once, we obtain

$$\dot{x}' = -2\omega(z'\cos\lambda - y'\sin\lambda) + \dot{x}'_0 \tag{11.52}$$

$$\dot{y}' = -2\omega x'\sin\lambda + \dot{y}'_0 \tag{11.53}$$

$$\dot{z}' = -g_{eff}t + 2\omega x'\cos\lambda + \dot{z}'_0 \tag{11.54}$$

where the integration constants $\dot{x}'_0$, $\dot{y}'_0$, and $\dot{z}'_0$ are the components of the initial velocity at $t = 0$. Substituting $\dot{y}'$ and $\dot{z}'$ from Equations 11.53 and 11.54 into Equation 11.51, we obtain

$$\ddot{x}' = 2\omega g_{eff}t\cos\lambda - 2\omega(\dot{z}'_0\cos\lambda - \dot{y}'_0\sin\lambda) \tag{11.55}$$

where the terms involving ω^2 have been dropped; with this approximation, $g_{eff} \cong g$.

Integrating Equation 11.55 twice, we obtain, for the initial conditions $x'_0 = y'_0 = z'_0 = 0$ at $t = 0$

$$\dot{x}' = \omega g t^2\cos\lambda - 2\omega t(\dot{z}'_0\cos\lambda - \dot{y}'_0\sin\lambda) + \dot{x}'_0 \tag{11.56}$$

$$x' = \frac{1}{3}\omega g t^3\cos\lambda - \omega t^2(\dot{z}'_0\cos\lambda - \dot{y}'_0\sin\lambda) + \dot{x}'_0 t. \tag{11.57}$$

Similarly, neglecting terms involving ω^2,

$$y' = \dot{y}'_0 t - \omega\dot{x}'_0 t^2\sin\lambda \tag{11.58}$$

$$z' = -\frac{1}{2}gt^2 + \dot{z}'_0 t + \omega\dot{x}'_0 t^2\cos\lambda. \tag{11.59}$$

We now apply our results to different situations:

(1) A body freely falling from rest

In this case, the initial velocity at $t = 0$ is zero, and we have

$$x' = \frac{1}{3}\omega g t^3\cos\lambda \quad \text{and} \quad y' = 0, \quad z' = -\frac{1}{2}gt^2.$$

If the body falls through a vertical distance h, then

$$h \cong z' = \frac{1}{2}gt^2 \quad \text{or} \quad t = \sqrt{2h/g}$$

and

$$x' = \frac{1}{3}\omega\cos\lambda\sqrt{8h^3/g}. \tag{11.60}$$

Thus, the body drifts eastward as a consequence of the rotation of the Earth.

(2) An upward projectile with $\dot{x}'_0 = \dot{y}'_0 = 0$ and $\dot{z}'_0 = v_0$.

In this case, we have

$$x' = \frac{1}{3}\omega g t^3\cos\lambda - \omega t^2 v_0\cos\lambda, \quad y' = 0, \quad z' = -\frac{1}{2}gt^2 + v_0 t. \tag{11.61}$$

Setting $z' = 0$ in the last equation and solving for t, we have $t = 0$ (initial time) or $t = 2v_0/g$ (the projectile strikes the ground).

We can express v_0 in terms of g and the maximum height h that the projectile can reach:

$$\frac{1}{2}mv_0^2 = mgh$$

from which we obtain

$$v_0 = \sqrt{2gh}.$$

Then,

$$t = 2v_0/g = \sqrt{8h/g}.$$

Substituting this into the expression for x', we obtain

$$x' = -\frac{4}{3}\omega\cos\lambda\left(\frac{8h^3}{g}\right)^{1/2}. \tag{11.62}$$

That is, the body drifts westward when it strikes the ground.

(3) A projectile fired with an elevation

If a projectile is fired in an easterly direction, the initial components of velocities are $\dot{x}'_0$, 0, $\dot{z}'_0$, and the displacements are

$$x' = \frac{1}{3}\omega g t^3\cos\lambda - \dot{z}'_0\omega t^2\cos\lambda + \dot{x}'_0 t \tag{11.63a}$$

$$y' = -\omega\dot{x}'_0 t^2\sin\lambda \tag{11.63b}$$

$$z' = -\frac{1}{2}gt^2 + \omega\dot{x}'_0 t^2\cos\lambda + \dot{z}'_0 t. \tag{11.63c}$$

It is apparent from Equation 11.59 that the trajectory is not planar. Although there is no initial y' velocity, because of the rotation of the Earth beneath the projectile, a y' displacement nevertheless occurs. An observer at $(x'y'z')$ observes a drift of the projectile in the negative y' or southerly direction. Looking east, the projectile thus drifts to the right. Similarly, a projectile with initial velocity components 0, $\dot{y}_0'$, $\dot{z}_0'$ will have an x' displacement.

$$x' = \frac{1}{3}\omega g t^3 \cos\lambda + \omega t^2(\dot{y}_0' \sin\lambda - \dot{z}_0' \cos\lambda) \tag{11.64}$$

For large t or small $\dot{z}_0'$, x' is positive; the projectile drifts east (to the right) when fired in a northerly direction.

11.4 FOUCAULT PENDULUM

The Coriolis effect resulting from the rotation of the Earth was dramatically demonstrated by Jean Foucault in 1851, using a long pendulum with a very heavy bob (to reduce the effects of air currents) hung from a support designed to allow the pendulum to swing freely in any direction. His experiments showed that the plane in which the pendulum oscillates rotates slowly with time. The effect is very striking because, unlike previous examples, the motion takes place in a small region of space, and the velocity of the pendulum is not very great. The gravity force is, of course, much more important than the Coriolis force in determining the pendulum's motion. However, the direction of the small Coriolis force is out of the plane of oscillation; thus, despite its smallness, the Coriolis force has a significant effect on the motion of the pendulum.

As shown in Figure 11.19, ϕ is the angle between the line along which the pendulum oscillates and a reference polar axis. Foucault showed that the rate of rotation ϕ of the direction of swing of the bob is related to the latitude λ of the pendulum on the earth and the angular velocity ω of the Earth's rotation by the expression $\dot{\phi} = \omega \sin\lambda$. This formula can be derived as follows:

We first select a Cartesian coordinate system with its origin at the equilibrium position of the pendulum bob and with the z'-axis along the local vertical (Figure 11.20). The x'-axis is positive to the east, and the y'-axis is positive to the north. With this choice of axes, $\vec{g}$ and $\vec{\omega} \times \vec{v}'$ are again given by Equation 11.48:

$$\vec{g}_{eff} = -g_{eff}\hat{k}', \quad \vec{\omega} = \omega\cos\lambda\,\hat{j}' + \omega\sin\lambda\hat{k}'$$

$$\vec{\omega} \times \dot{\vec{r}}' = \hat{i}'(\omega\dot{z}'\cos\lambda - \omega\dot{y}'\sin\lambda) + \hat{j}'\omega\dot{x}'\sin\lambda - \hat{k}'\omega\dot{x}'\sin\lambda.$$

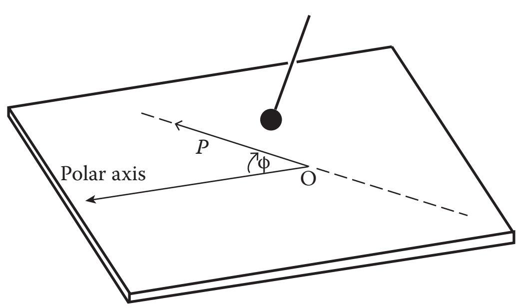


FIGURE 11.19 Point O is the position of equilibrium.

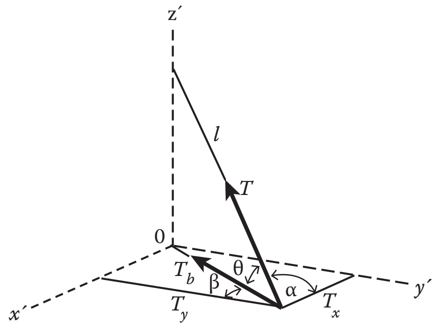


FIGURE 11.20 Foucault pendulum.

The physical force acting on the bob is the vector sum of the tension $\vec{T}$ in the cord and the weight of the bob $m\vec{g}_{eff}$. If we neglect the centrifugal force term, then $\vec{g}_{eff} \cong \vec{g}$, and the equation of motion of the pendulum bob is

$$m\frac{\delta^2\vec{r}'}{\delta t^2} = m\vec{g} + \vec{T} - 2m\vec{\omega}\times\vec{v}'. \tag{11.65}$$

It is clear from Figure 11.19 that the components of $\vec{T}$ are

$$T_b/T = b/l \quad \text{or} \quad T_b = bT/l$$

$$T_x = -xT_b/b = -xT/l,\ T_y = -yT_b/b = -yT/l,\ T_z = (l - z)T/l. \tag{11.66}$$

Equation 11.65 then reduces to

$$\left.\begin{aligned} m\ddot{x}' &= -\frac{x'}{l}T - 2m\omega(\dot{z}'\cos\lambda - \dot{y}'\sin\lambda) \\ m\ddot{y}' &= -\frac{y'}{l}T - 2m\omega\dot{x}'\sin\lambda \\ m\ddot{z}' &= \frac{l-z}{l}T - mg + 2m\omega\dot{x}'\cos\lambda \end{aligned}\right\}. \tag{11.67}$$

A fourth equation that relates x', y', and z' is provided by the constraint equation stating that l is a constant:

$$x'^2 + y'^2 + (1 - z')^2 = l^2. \tag{11.68}$$

We are interested in the case where the amplitude of the bob's swing is small with the horizontal excursions very small compared to the length of the cord. Under this condition, $\dot{z}'$ is small compared to $\dot{x}'$ and $\dot{y}'$ and can be neglected, and tension T is approximately equal to mg. Taking these factors into account, the $x' - y'$ motion is then given by the following differential equations:

$$\ddot{x}' - 2\omega'\dot{y}' + (g/l)x' = 0 \tag{11.69}$$

$$\ddot{y}' + 2\omega'\dot{x}' + (g/l)y' = 0 \tag{11.70}$$

where $\omega' = \omega\sin\lambda = \omega_{z'}$, which is the local vertical component of the Earth's angular velocity.

We are confronted with a set of differential equations that are not in separated form. A convenient method of solving them is to introduce a new variable $u = x' + iy'$. However, the physical behavior of the system can be seen by transforming to a new coordinate system *OXYZ*, which rotates relative to the primed system in such a way as to cancel the vertical component of the Earth's rotation. That is, the system *OXYZ* rotates with angular rate $-\omega'$ about the vertical axis as shown in Figure 11.21. Referring to a point *P*, the equations of transformation are

$$x' = X\cos\omega' t + Y\sin\omega' t, \quad y' = -X\sin\omega' t + Y\cos\omega' t. \tag{11.71}$$

It is obvious that

$$x'^2 + y'^2 = X^2 + Y^2.$$

Substituting x' and y' from Equation 11.71 and their derivatives into Equations 11.69 and 11.70, we obtain the following equation, after collecting terms and dropping terms involving ω'^2:

$$\left(\ddot{X} + \frac{g}{l}X\right)\cos\omega' t + \left(\ddot{Y} + \frac{g}{l}Y\right)\sin\omega' t = 0.$$

If this equation is true for all values of $\omega' t$, the coefficients of the sine and cosine terms must both vanish:

$$\left.\begin{aligned}\ddot{X} + \frac{g}{l}X = 0\\ \ddot{Y} + \frac{g}{l}Y = 0\end{aligned}\right\}. \tag{11.72}$$

These are the differential equations of a two-dimensional harmonic oscillator. Thus, the path, projected on the *XY*-plane, is an ellipse with fixed orientation in the unprimed *OXYZ* system. In the primed system $Ox'y'z'$, the path is an ellipse that undergoes a steady precession with angular velocity ω'. That is, in the primed $Ox'y'z'$ system, the plane of oscillation of a Foucault pendulum rotates clockwise at a rate given by the Foucault formula

$$\dot{\phi} = \omega' = \omega\sin\lambda \tag{11.73}$$

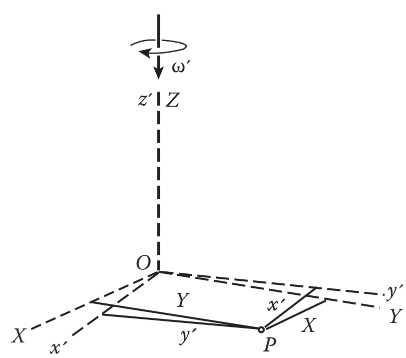


FIGURE 11.21 Rotation of coordinate axes.

in the northern hemisphere and counterclockwise in the southern hemisphere. The period P is

$$P = \frac{2\pi}{\omega'} = \frac{2\pi}{\omega \sin\lambda} = \frac{24}{\sin\lambda}\ \text{h}. \tag{11.74}$$

Thus, at the North Pole ($\sin\lambda = 1$), the pendulum precesses clockwise through a complete revolution every 24 h. However, from the viewpoint of an observer in space, the plane of oscillation of the pendulum at the North Pole remains fixed, but the Earth turns counterclockwise beneath it. At the equator ($\sin\lambda = 0$), the period is infinite; that is, there is no apparent rotation of the plane of oscillation.

In addition to the rotational precession discussed earlier, there is another natural precession of the spherical pendulum that is usually much larger than the rotational precession. However, the natural precession can be reduced to negligibly small if we start the pendulum very carefully. One way to start the pendulum is by drawing it aside with a thread and letting it start from rest by burning the thread. The natural precession of the spherical pendulum will be discussed briefly in Chapter 11.

11.5 LARMOR'S THEOREM

Larmor's theorem concerns the motion of a system of charged particles in a weak uniform magnetic field. Larmor discovered that whenever a system of charged particles (having the same charge-to-mass ratio q/m) moves under the action of a central force field and their mutual (central) forces, the addition of a weak uniform magnetic field produces an additional slow precessional motion of the entire system about the center of mass. This additional slow precessional motion superimposes on the original unperturbed motion. The angular velocity $\vec{\omega}$ of the precession is equal to $-(q/2mc)\vec{H}$, where $\vec{H}$ is the magnetic field (in Gaussian units). This is known as Larmor's theorem.

The proof of Larmor's theorem is straightforward. The equation of motion of the kth particle in an inertial frame and in the absence of magnetic field is

$$m\frac{\mathrm{d}^2\vec{r}_k}{\mathrm{d}t^2} = \vec{F}_k + \vec{F}_k^i, \quad k = 1, 2, \ldots, N \tag{11.75}$$

where N is the total number of charged particles in the system, $\vec{F}_k$ is the central force that depends only on the distance of particle k from the force center that we take as the origin, and $\vec{F}_k^i$ is the sum of the forces on the particle k resulting from the other particles. The forces $\vec{F}_k^i$ depend only on the distances of the particle from one another. When the magnetic field $\vec{H}$ is applied, the equation of motion becomes

$$m\frac{\mathrm{d}^2\vec{r}_k}{\mathrm{d}t^2} = \vec{F}_k + \vec{F}_k^i + \frac{q}{c}\frac{\mathrm{d}\vec{r}_k}{\mathrm{d}t} \times \vec{H} \tag{11.76}$$

where the last term on the right-hand side is the Lorentz force.

Now, in a primed coordinate system that has a common origin with the inertial system and is in rotation about this common origin with an angular velocity $\vec{\omega}$, the equation of motion is

$$\begin{aligned} m\frac{\delta^2\vec{r}_k}{\delta t^2} &= \vec{F}_k + \vec{F}_k^i - m\vec{\omega} \times (\vec{\omega} \times \vec{r}_k') \\ &\quad + \frac{q}{c}(\vec{\omega} \times \vec{r}_k') \times \vec{H} + \frac{\delta\vec{r}_k'}{\delta t} \times (q\vec{H}/c + 2m\vec{\omega}). \end{aligned} \tag{11.77}$$

The last term drops out if we choose $\vec{\omega} = -(q/2mc)\vec{H}$. Equation 11.77 then becomes

$$m\frac{\delta^2 \vec{r}_k}{\delta t^2} = \vec{F}_k + \vec{F}_k^i + \frac{q^2}{4mc^2}\vec{H} \times (\vec{H} \times \vec{r}_k'). \tag{11.78}$$

The forces $\vec{F}_k$ and $\vec{F}_k^i$ depend only on the distances of the particles from the origin and on their distances from one another, and these distances are the same in both the rotating and inertial frames. Therefore, if the last term on the right-hand side of Equation 11.78 is negligibly small, Equation 11.78 has exactly the same form in terms of primed coordinates as Equation 11.75 in unprimed stationary inertial coordinates. The description of the motion in the rotating frame is, thus, to a first approximation, the same as the motion of the system in the stationary inertial frame in the absence of the magnetic field. To a first approximation, the motion of the system in the presence of the magnetic field will be observed to be in the same orbit as in the absence of the magnetic field but precessing with the angular velocity $\vec{\omega} = -(q/2mc)\vec{H}$, commonly known as the Larmor frequency. The corresponding precession is called the Larmor precession. Note that the Larmor frequency is just half the cyclotron frequency. This difference arises from a factor of 2 in the Coriolis force term $-2m\vec{\omega}\vec{v}'$.

The condition that the magnetic field be weak means that the last term in Equation 11.78 must be negligible in comparison with the first two terms. Notice that the term we are neglecting is proportional to H^2, whereas the term in Equation 11.77 that we have dropped is proportional to H. Hence, for sufficiently weak fields, the former may be negligible even though the latter is not. The last term in Equation 11.78 may be written in the form

$$\frac{q^2}{4mc^2}\vec{H} \times (\vec{H} \times \vec{r}_k') = m\vec{\omega} \times (\vec{\omega} \times \vec{r}_k').$$

Another way of formulating the condition for a weak magnetic field is to say that the Larmor frequency $\vec{\omega}(= -q\vec{H}/2mc)$ must be small compared with the frequencies of the motion in the absence of a magnetic field.

The Larmor precession leads to observable changes in the spectra emitted by atoms placed in a magnetic field because the Bohr energy level corresponding to a given angular momentum is slightly shifted by the Larmor precession. The shift in the atomic spectral lines in the presence of a magnetic field was first observed by Dutch physicist Zeeman in 1896 and is now known as the Zeeman effect.

11.6 CLASSICAL ZEEMAN EFFECT

The preceding considerations form the basis for the classical theory of the Zeeman effect. In this theory, it is considered that atoms emit light of frequency ν_0 because electrons within the atoms oscillate with a simple harmonic motion of that frequency. If a weak magnetic field is applied, the line of oscillation of the electrons will rotate about the direction of the field with an angular velocity $\omega = -(q/2mc)H$ (Larmor's theorem).

It is convenient to resolve the assumed simple harmonic motion of an atomic electron into components parallel and perpendicular to the magnetic field. The magnetic field does not affect the frequency of the parallel component of the motion, which continues to emit radiation of frequency ν_0. The perpendicular component can be further resolved into two circular motions of opposite directions, which, in the absence of a field, have the angular velocities $\pm\, 0.2\pi\nu_0$. In the presence of a magnetic field, the angular velocity of one of the circular motions is increased by $\omega = -qH/mc$ while the angular velocity of the opposite circular motion is decreased by the same amount. Consequently, the circular motions produce a radiation of frequencies $\nu_0 \pm \nu_L$, where $\nu_L = \omega/2\pi$ is the Larmor frequency. A weak magnetic field, therefore, causes a spectral line of frequency v_0 to split up into a Lorentz triplet, consisting of one (undeviated) line of frequency v and two lines of frequencies $v \pm v_L$.

Now, let *XYZ* represent a set of coordinate axes fixed in a laboratory, and let the magnetic field *H* be parallel to the *X*-axis (Figure 11.22). Suppose that an electron is executing simple harmonic oscillations of angular frequency ω_0 and amplitude *A* along the line *OP*. Because of the Larmor's effect, the path *OP* revolves around the *X*-axis with Larmor's frequency $\omega = -(q/2mc)H$.

We now resolve the simple harmonic vibration under consideration into three vibrations along the axes *XYZ* as follows:

$$x = A \cos \omega_0 t \cos \theta \tag{11.79}$$

$$\begin{aligned} y &= A\cos\omega_0 t \sin\theta\cos\omega t \\ &= \frac{1}{2}A\sin\theta\left[\cos(\omega_0 - \omega)t + \cos(\omega_0 + \omega)t\right] = y_1 + y_2 \end{aligned} \tag{11.80}$$

$$\begin{aligned} z &= A\cos\omega_0 t \sin\theta\sin\omega t \\ &= \frac{1}{2}A\sin\theta\left[-\sin(\omega_0 - \omega)t + \sin(\omega_0 + \omega)t\right] = z_1 + z_2. \end{aligned} \tag{11.81}$$

But

$$y_1 = \frac{1}{2}A\sin\theta\cos(\omega_0 - \omega)t = \frac{1}{2}A\sin\theta\cos\omega_1 t \tag{11.82}$$

$$z_1 = -\frac{1}{2}A\sin\theta\sin(\omega_0 - \omega)t = -\frac{1}{2}A\sin\theta\sin\omega_1 t \tag{11.83}$$

represent a circular vibration in the *YZ*-plane of angular frequency $\omega_1 = \omega_0 - \omega$ in the clockwise direction, and

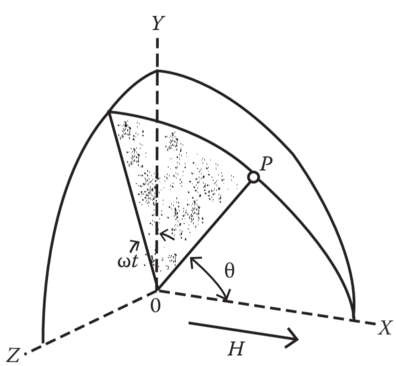


FIGURE 11.22 Electron oscillating along the path *OP*. The magnetic field is parallel to the *x*-axis.

$$y_2 = \frac{1}{2} A \sin\theta \cos(\omega_0 + \omega)t = \frac{1}{2} A \sin\theta \cos\omega_2 t \quad (11.84)$$

$$z_2 = \frac{1}{2} A \sin\theta \sin(\omega_0 + \omega)t = \frac{1}{2} A \sin\theta \sin\omega_2 t \quad (11.85)$$

represent a circular vibration in the YZ-plane of angular frequency $\omega_1 = \omega_0 + \omega$ in the counterclockwise direction. Therefore, we have resolved the original vibration into a linear vibration along the X-axis of angular frequency ω_0 and two circular vibrations in the YZ-plane of angular frequencies:

$$\omega_1 = \omega_0 - \frac{eH}{2mc} \quad \text{and} \quad \omega_2 = \omega_0 + \frac{eH}{2mc} \quad (11.86)$$

respectively, where e is the charge of the electron. Because a linear oscillator emits no radiation along its axis, no radiation of frequency v_0 (= $\omega_0/2\pi$) proceeds in the direction of the magnetic field $\vec{H}$.

The radiating atoms at the origin can be viewed from a point on the X-axis (longitudinal observation, parallel to H) or from one in the YZ-plane (transverse observation, perpendicular to H). For longitudinal observation because the oscillations along the X-axis emit no radiation in this direction, the only lines observed in the spectroscope are circularly polarized lines of frequencies v_1 (= $\omega_1/2\pi$) and v_2 (= $\omega_2/2\pi$). For transverse observation, three lines of frequencies v_0, v_1, v_2 appear. These constitute the normal Zeeman pattern observed only in the simplest spectra (Figure 11.23).

We see that, according to the classical theory, each spectral line of an atomic system should exhibit the normal Zeeman effect. But many spectral lines are observed to have more than three Zeeman components, and the separation between components is frequently different from v_L. This is known as the anomalous Zeeman effect, which involves the electron spin, and so a complete understanding of this anomalous Zeeman effect can be obtained from quantum mechanics.

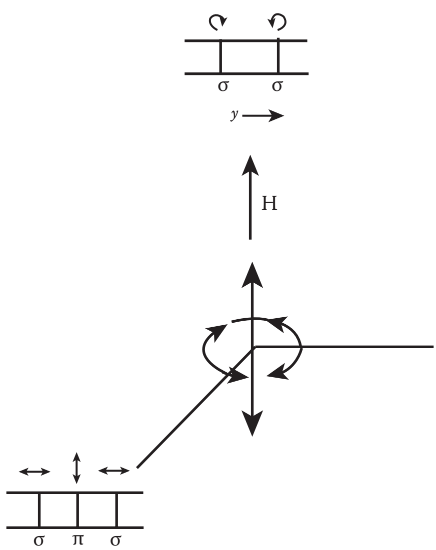


FIGURE 11.23 Normal Zeeman pattern.

11.7 PRINCIPLE OF EQUIVALENCE

The principle of equivalence was introduced in Chapter 2 as the assertion of the equivalence of the initial mass m_I and the gravitational mass m_g of a body.

When gravity acts on a body, it acts on the gravitational mass m_g of the body, and the result of the action is an acceleration of the initial mass m_I of the body. The fact that all bodies fall in a vacuum with the same acceleration indicates that within experimental accuracy, that ratio of inertial to gravitational mass is independent of the body. Newton realized this when he formulated his laws of motion. Einstein generalized this and asserted the equivalence of m_I and m_g.

We now restate the principle of equivalence in a different way. Let us consider a region of space–time in which a constant gravitational field $\vec{g}$ exists. One such region is the neighborhood of any point on the surface of the Earth. If gravity were the only force acting, all bodies in the region would fall with the same acceleration $\vec{a} = \vec{g}$. Hence, by transforming to a frame S' that accelerates with acceleration $\vec{a} = \vec{g}$, we can eliminate the effects of gravitation, and so any object will appear not accelerated unless there is non-gravitational force $\vec{F}$ acting on it. We can show this explicitly: Consider a particle inside a stationary elevator; a gravitational force $m\vec{g}$ and a non-gravitational force $\vec{F}$ act upon it. Newton's second law gives

$$m\vec{a} = m\vec{g} + \vec{F}. \tag{11.87}$$

If the elevator is severed from its supporting cable and falls freely under the action of gravity, in this accelerated frame S', the acceleration of the particle relative to the elevator is $\vec{a}' = \vec{a} - \vec{g}$, and so Equation 11.87 becomes

$$m(\vec{a}' + \vec{g}) = m\vec{g} + \vec{F}$$

from which we obtain

$$m\vec{a}' = \vec{F}. \tag{11.88}$$

We see that gravity has been "transformed away" in the free-fall elevator (frame S'). The elevator and all its contents accelerated downward at the rate $\vec{g}$. If an experimenter inside the elevator releases an object, it will float as if the elevator were motionless in free space. From the point of view of this experimenter, there is no way to determine whether the elevator is in free space or whether it is falling freely in a gravitational field. The freely falling elevator constitutes a local inertial frame. Therefore, the equivalence of gravitational and inertial mass implies the following:

> In an elevator (a small laboratory), falling freely in a gravitational field, mechanical phenomena are the same as those observed in an inertial frame in the absence of a gravitational field.

Einstein, in 1907, generalized this statement as the principle of equivalence by replacing the words "mechanical phenomena" with the laws of physics. This is often called the "strong" equivalence principle, and the equivalence of inertial and gravitational mass is known as the "weak" equivalence principle.

An accelerating coordinate frame can be extended over a large space, so it is non-local. However, real gravitational fields are local, that is, only over a sufficiently small region of the gravitational field can it be considered constant. Hence, the laboratory (the accelerating frame) must be small.

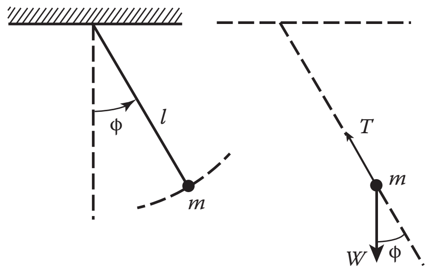


FIGURE 11.24 Experimental investigation of the principle of equivalence.

Only, for small frames, the two are not distinguishable. We should bear in mind that the equivalence principle applies only to local frames.

There have been a number of experimental investigations of the principle of equivalence. Newton did so by studying the period of a pendulum with interchangeable bobs (Figure 11.24). The tangential force is $-W \sin \phi = -m_g g \sin \phi$, and the equation of motion of the pendulum bob is

$$m_I l\ddot{\phi} = -m_g g \sin\phi$$

which reduces to, in the small angle approximation,

$$m_I l\ddot{\phi} = -m_g g\phi.$$

The period of the pendulum is

$$T = \frac{2\pi}{\omega} = 2\pi\sqrt{\frac{l}{g}}\sqrt{\frac{m_I}{m_g}}.$$

Newton looked for a variation in T using bobs of different composition. He found no such change and, from an estimate of the sensitivity of the method, concluded that m_g/m_I is constant to better than one part in a thousand for common materials. Most recent investigations indicated that it is less than 10^{-12}.

The principle of equivalence is generally regarded as a fundamental law of physics. We used it here to discuss the ratio of gravitational to inertial mass. Surprisingly enough, it can be used to show that clocks run at different rates in different gravitational fields. In the following, we present a simple argument that shows how the principle of equivalence forces us to give up the classical notion of time.

11.7.1 Principle of Equivalence and Gravitational Red Shift

A gravitational red shift is the shift to longer wavelength of light passing through a strong gravitational field. What is the connection between this and the equivalence principle? And how does the equivalence principle force us to give up the classical notion of time?

To answer these questions, let us consider radiating atoms in the strong gravitational field of a dense star, such as a white dwarf or neutron star. The radiating atoms emit light at only certain characteristic wavelengths. We can visualize atoms as clocks, which "tick" at characteristic frequencies. The shift toward longer wavelengths corresponds to a slowing of the clocks. So the gravitational red shift implies that clocks in a gravitational field appear to run slow when viewed from outside the

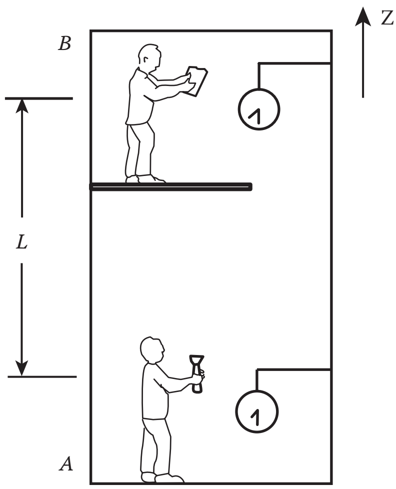


FIGURE 11.25 Two scientists in a closed cabin communicate with light.

field. We shall show that the origin of the effect lies in the nature of space, time, and gravity, not in the trivial effect of gravity on mechanical clocks.

Consider two scientists, *A* and *B*, separated by distance *L* as shown in Figure 11.25. *A* has a clock and a light that he flashes at intervals separated by time T_A. The signals, traveling at the speed of light *c*, are received by *B*, who notes the interval between pulses, T_B, with his own clock. A plot of vertical distance versus time is shown for two light pulses in Figure 11.26. The pulses are delayed by the transit time, *L/c*, but the interval T_B is the same as T_A, where *c* is the speed of light. Hence, if *A* transmits the pulses at, say, 1-s intervals, so that $T_A = 1$ s, then *B*'s clock will read 1 s between the arrival of successive pulses. If observers move upward uniformly with speed *v*, they move equally, and we still have $T_B = T_A$.

Now, we observe the case of both observers accelerating upward at uniform rate *a* as shown in Figure 11.27. *A* and *B* start from rest, and the graph of distance versus time is a parabola. Because *A* and *B* have the same acceleration, the curves are parallel, separated by distance *L* at each instant. It is apparent from Figure 11.27 that $T_B > T_A$ because the second pulse travels farther than the first and has a longer transit time. The effect is purely kinematical. Now, by the principle of equivalence, *A* and *B* cannot distinguish between their upward accelerating system and a system at rest in a

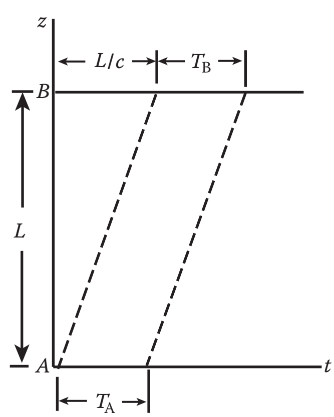


FIGURE 11.26 Time interval observed by the two scientists is the same, $T_B = T_A$.

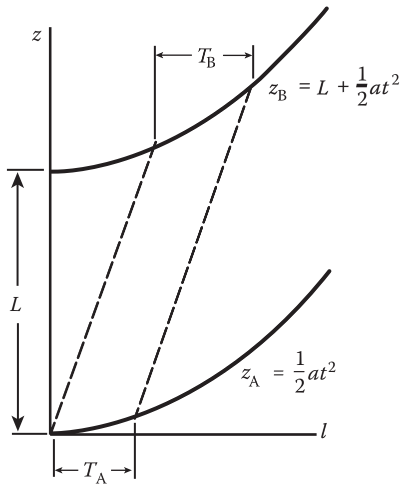


FIGURE 11.27 If the cabin is accelerating upward at uniform rate, $T_B > T_A$.

downward gravitational field with magnitude $g = a$. Thus, if the experiment is repeated in a system at rest in a gravitational field, the equivalence principle requires that $T_B > T_A$ as before. If $T_A = 1$ s, *B* will observe an interval greater than 1 s between successive pulses. *B* will conclude that *A*'s clock is running slow. This is the origin of the gravitational red shift.

By applying the argument quantitatively, the following approximate result is readily obtained:

$$\frac{\Delta T}{T} = \frac{T_B - T_A}{T_A} = \frac{gL}{c^2}$$

where it is assumed that $\Delta T/T \ll 1$.

The gravitational red shift has been measured and confirmed to an accuracy of 1%. The experiment was done by Pound, Rebka, and Snyder at Harvard University. The "clock" was the frequency of a gamma ray, and by using a technique known as the Mossbauer absorption, they were able to measure accurately the gravitational red shift resulting from a vertical displacement of 25 m.

PROBLEMS

1. A uniform sphere is placed on the upper surface of a wedge, and the wedge is accelerated along a horizontal surface with a constant acceleration $\vec{A}$. What must $\vec{A}$ be in order that the sphere will remain at rest with respect to the wedge?
2. A bicycle travels with constant speed around a track of radius ρ. Find the acceleration of the highest point on one of its wheels.
3. A small mass hangs from a string in a car. What is the static angle of the string from the vertical, and what is its tension when the car (a) moves with a constant velocity and (b) moves with a constant acceleration?
4. A uniform cylinder of mass *M* and radius *R* rolls without slipping on a plank, which is accelerated at the rate *A* with respect to a horizontal surface. Find the acceleration of the cylinder with respect to the plank and to the inertial system.
5. A bead slides on a smooth rod. The rod is constrained to rotate uniformly at an angular velocity $\vec{\omega}$ in a plane about an axis passing through one end perpendicular to the rod. Find the position of the bead on the rod as a function of time, and the reaction of the rod on the bead. (Assume the axis is attached to an inertial system.)

6. Find the velocity and acceleration of a point P on the rim of a wheel of radius R, which is rolling down an inclined plane. The inclined plane makes an angle θ with the horizontal.
7. A coordinate system xyz is rotating with respect to an XYZ coordinate system having the same origin and assumed to be fixed in space (i.e., it is an inertial system). The angular velocity of the xyz system relative to the XYZ system is given by

$$\vec{\omega} = 2t\hat{i} - t^2\hat{j} + (2t+4)\hat{k}$$

where t is the time. The position vector of a particle at time t as observed in the xyz system is given by

$$\vec{r} = (t^2+1)\hat{i} - 6t\hat{j} + 4t^3\hat{k}.$$

Find (a) the apparent velocity and true velocity of the particle and (b) the apparent and the true acceleration at time $t = 1$.
8. Find (a) the Coriolis acceleration and (b) the centripetal acceleration of the particle in Problem 10.7 at time $t = 1$.
9. A particle slides over a smooth plane, which is tangential to the Earth's surface at a northern latitude λ. Find the reaction, resulting from Earth's rotation, of the plane on the particle.
10. A bug crawls outward with constant speed v' along the spoke of a wheel, which is rotating with constant angular velocity $\vec{\omega}$ about a vertical axis. Find all the forces acting on the bug. How far can the bug crawl before it starts to slip, given the coefficient of friction μ between the bug and the spoke?
11. If a projectile is fired due east from a point on the Earth's surface at a northern latitude λ with a velocity of magnitude v_0 and at an angle of inclination to the horizontal of α, show that the lateral deflection when the projectile strikes the Earth is

$$d = \frac{4v_0^3}{g^2}\omega \sin\lambda \sin^2\alpha \cos\alpha$$

where ω is the rotation speed of the Earth. If the range of the projectile is R for the case when $\omega = 0$, show that the change of range resulting from the rotation of the Earth is

$$\Delta R = \sqrt{2R^3/g}\,\omega\cos\lambda\left(\cot^{1/2}\alpha - \frac{1}{3}\tan^{3/2}\alpha\right).$$

12. A spherical planet of radius R rotates with a constant angular velocity $\vec{\omega}$. The effective gravitational acceleration $\vec{g}_{eff}$ is some constant, g, at the poles and $0.8g$ at the equator. Find $\vec{g}_{eff}$ as a function of the polar angle θ and g.
13. In Figure 11.28, the O-system, with its z-axis out from the page, is a fixed coordinate system. The O_r-system is a rotating system with x_r and y_r rotating with constant angular velocity $\vec{\omega}$ about the $z = z_r$ axis with respect to the fixed system. An observer sitting at O_r sees a rock on the ground fixed in the fixed system at a distance r along the x-axis. To him, the rock moves in a circular orbit with speed $v' = \omega r$. What force holds the rock in this circular orbit for the rotating observer?
14. A particle of mass m is constrained to move along a smooth rod AOB (Figure 11.29). The rod $A0B$ rotates in a vertical plane about a horizontal axis with constant angular velocity $\vec{\omega}$. The horizontal axis passes through the midpoint O of the rod and is perpendicular to the vertical plane. Determine the motion of the particle P along the rod.

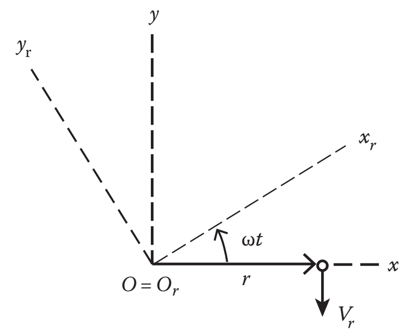


FIGURE 11.28 Two coordinates; one is fixed and one is rotating.

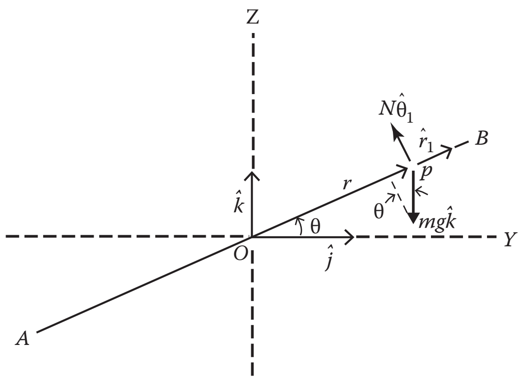


FIGURE 11.29 Particle is constrained to move along a rotating rod.

15. Artificial gravity on spacecraft

One way of overcoming health hazards in the prolonged absence of a gravitational field on board a spaceship is to produce artificial gravity by spinning the ship about an axis. If, for example, the ship were built in the form of a large ring, its outer wall would become the "floor" as the ring spun about its axis. The gravitational field results from centrifugal acceleration.

(a) Estimate the angular velocity of spin required to produce a gravitational field equivalent to that on Earth for a ring-shaped spaceship with an outer radius of 2000 ft. (Figure 11.30).

(b) The circle of radius r_0 in Figure 11.30 represents the floor (the outer wall) of the ring-shaped spaceship. The coordinate system is permanently fixed to the ship with the x_3-axis pointed "up" (i.e., toward the center of the ring). The x_2-axis is tangential to the ring, and the x_1-axis is parallel to the axis of the ring. The angular velocity of rotation is then $\vec{\omega} = \hat{i}\omega$. Find the acceleration $\vec{a}_0$ of the origin of the coordinate system.

(c) If r is the displacement vector of a particle extended from the origin of the accelerated frame, what is the Lagrangian that gives the equations of motion of a particle with respect to the accelerated coordinate system (i.e., with respect to the persons on board the ship)? Show that, with $\vec{\omega}$ and $\vec{a}_0$ as given in (b), this Lagrangian can be reduced to the following form:

$$L = \frac{1}{2}m\left(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2\right) + \frac{1}{2}m\omega^2\left(x_2^2 + x_3^2\right) - mr_0x_3\omega^2 + m\omega(\dot{x}_3x_2 - \dot{x}_2x_3).$$

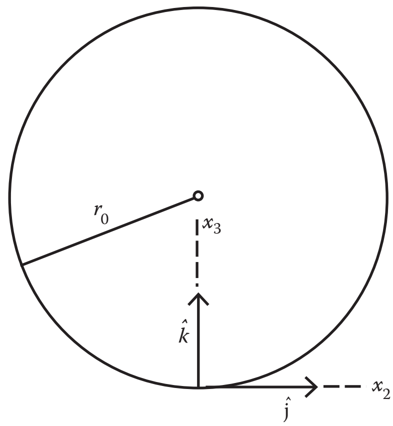


FIGURE 11.30 Artificial gravity in a ring-shaped spacecraft.

(d) Assume that there are no constraints so that the particle has three degrees of freedom. Show that the equations of motion as deduced from the preceding Lagrangian are

$$\ddot{x}_1 = 0$$
$$\ddot{x}_2 - \omega^2 x_2 - 2\omega\ddot{x}_3 = 0$$
$$\ddot{x}_3 + (r_0 - x_3)\omega^2 + 2\omega\dot{x}_2 = 0.$$

The first equation is trivial. Show that the next two equations can be combined into one equation in $z = x_2 + ix_3$, where $i = \sqrt{-1}$

$$\ddot{z} + 2i\omega\dot{z} - \omega^2\dot{z} + i\omega^2 r_0 = 0$$

and that its general solution is of the form

$$z = ir_0 + (c_1 + c_2 t)\exp(-i\omega t).$$

(e) Suppose that a particle is dropped from rest from a height h above the floor. The initial conditions can be taken to be $z_0 = ih$ and $\dot{z}_0 = 0$. Determine the constants c_1 and c_2 in terms of the initial position and velocity, and show that the general solution z is

$$z = ir_0 + (r_0 - h)(\omega t - 1)\exp(-i\omega t).$$

(f) Consider that the motion is over a small distance such that $\omega t \ll 1$; then show that, by expanding the exponent in the solution given in (*e*) to terms of not higher than $(\omega t)^5$, x_2 and x_3 are approximately given by

$$x_2 = -\frac{1}{3}g\omega t^3 + \frac{1}{30}g\omega^3 t^5 + \ldots$$

$$x_3 = h - \frac{1}{2}gt^3 + \frac{1}{8}g\omega^2 t^4 + \ldots$$

where $g = (r_0 - h)\omega^2$. These equations are similar to the equations of motion of a particle dropped from a point above the surface of the Earth. What is the significance of the quantity g? Note that the Coriolis effect is more pronounced in a spinning spaceship. As a numerical example, let $\omega = 0.4\ s^{-1}$, $g = 32$ ft./s^2, and $h = 8.0$ ft. First, solve t from the equation for x_2, and then determine x_3 (the horizontal deflection of the particle). Notice that the deflection is quite large and easily observed compared to the falling of a body near the surface of the Earth.

12 Motion of Rigid Bodies

So far, we have considered the motion of bodies that can be considered as particles. In this chapter, we consider motions in which the finite size of bodies is important. Such bodies will be assumed to be rigid. By a rigid body, we mean that the relative position of the parts of the body remains unchanged during motion. The body thus moves as a whole. In other words, a rigid body can be considered to be a system of particles subject to the set of constraints

$$r_{ij} = c_{ij} \tag{12.1}$$

where r_{ij} is the distance between the ith and jth particles, and c_{ij} are constants at all times. This definition of a rigid body can easily be extended to systems with a continuous mass distribution: the distance between any two infinitesimal volume elements and the mass in every volume element remain constant at all times. This definition of a rigid body is, of course, highly idealized. If both the size and shape of a body vary only slightly (or negligibly little) under the action of external forces, then for the purpose of describing the macroscopic motion of the body as a whole, we can regard the body as a rigid body. We should note that, as one special consequence of the physical speed limit equal to the speed of light, the concept of an ideal rigid body finds no place in special relativity. This is because a rigid body is set in motion instantaneously as a single unit when an external force is applied to it at any point. This means that signals (physical information) can be transmitted instantaneously from one point to another point in a rigid body, which violates the principles of special relativity.

All macroscopic quantities that we introduced in Chapter 2 for systems of N interacting particles are also useful in this chapter:

$$\text{total mass } M = \sum_{i=1}^{N} m_i$$

$$\text{center of mass (CM) } \vec{R} = \frac{1}{M}\sum_{i=1}^{N} m_i\vec{r}_i$$

$$\text{total momentum } \vec{P} = \sum m_i\dot{\vec{r}}_i$$

$$\text{total angular momentum } \vec{L} = \sum_{i=1}^{N} m_i\vec{r}_i \times \dot{\vec{r}}$$

$$\text{total torque } \vec{N}_i = \sum_{i=1}^{N} \vec{r}_i \times \left(\vec{F}^{(e)} + \vec{F}^{(i)}\right)$$

For a rigid body with a continuous mass distribution, we can replace the summation by integration over the volume of the body, for example,

$$M = \sum_{i=1}^{N} m_i \to \iiint_V \rho \, \mathrm{d}m$$

and

$$\vec{R} = \frac{1}{M} \iiint_V \vec{r} \rho \, \mathrm{d}V$$

where dV is the element volume, and ρ is the density dm/dV. The form to be taken by dV and the limits of integration depend on the geometry of the body in consideration.

The equations of motion and the general theorem established for systems of N particles in Chapter 3 can also be applied to rigid bodies. But certain simplifications are apparent because the possible types of motion are restricted. If one point of the body is fixed with respect to the primary inertial system, the only possible motion is that in which every other point moves on the surface of a sphere whose radius is the invariable distance from the moving point to the fixed point. If two points of the body are fixed, then the only possible motion is that in which all points except those on the line joining the two fixed points move in circles about centers located on the line. Now, if three points of the body not on the same straight line are fixed, the body is unable to move at all, and its position is completely determined.

Motion of a rigid body in which one point is kept fixed or two points are kept fixed is known as motion of rotation. This is simpler than the general motion of the body when no point is kept fixed. There is a special case of the latter that is elementary, namely, that in which all points of the body move in the same direction at any instant with the same velocity and acceleration. This is called motion of translation. These two are the most important types of motion of a rigid body because it can be shown that every displacement of a rigid body can be considered to be a combination of rotations and translations (Chasles' theorem from Michel Chasles 1793–1880). The translation of a rigid body will be given by the translation of any point in it; for example, the CM and the mechanics of particle motion will take care of this. In what follows, we shall concentrate on rotation.

We shall not prove Chasles' theorem. The reader who is interested in its proof can find a nice proof in the work of Kleppner and Kolenkow (1973).

12.1 INDEPENDENT COORDINATES OF RIGID BODY

What distinguishes a rigid body from a particle is its finite size; a rigid body must be described by its orientation as well as its location. The position of every particle in a rigid body is determined by the position of any one point of the body (such as the CM) plus the orientation of the body about that point. A total of six coordinates are needed to describe the motion of a rigid body. The position of one particle in the body requires the specification of three coordinates. Two angular coordinates can specify the position of a second particle that lies at a fixed distance from the first. The position of a third particle is determined by only one coordinate because its distance from the first and second particles is fixed. The positions of any other particles in the rigid body are completely fixed by their distances from the first three particles. Thus, a total of six coordinates determine the positions of the particles in a rigid body. Therefore, the motion of a rigid body is controlled by only six equations of motion: three for rotational motion and three for translational motion. For example, referred to an inertial frame of reference, the translational motion of the CM is governed by

$$\vec{F}^{(e)} = \dot{\vec{P}} \tag{12.2}$$

and the rotational motion about the CM or a fixed point is determined by

$$\vec{N}^{(e)} = \frac{\mathrm{d}\vec{L}}{\mathrm{d}t}. \tag{12.3}$$

If the external force vanishes, the CM moves with constant velocity. If the external torque $\vec{N}(e)$ about the CM vanishes, the angular momentum about the CM (and thereby the rotational motion) does not change, and the rigid body is said to be in equilibrium. For static equilibrium, the CM must be initially at rest and the total angular momentum about the CM must initially be zero.

12.2 EULERIAN ANGLES

A total of six independent coordinates are needed to describe the motion of a rigid body. How shall these six coordinates be assigned? First, we select a set of Cartesian axes x, y, z that has its origin at a fixed point within the rigid body and that we shall call the body axes. The problem is solved when these body coordinates are related to an external inertial reference frame, which is also Cartesian (X', Y', Z'). Three of the six generalized coordinates serve to locate the origin of the body axes with respect to the origin of the external axes. The second set of three generalized coordinates gives the orientation of x, y, z relative to the external axes (or relative to three fixed axes x', y', z' that have the same origin as the body axes but that are parallel to the external axes X', Y', Z' as shown in Figure 12.1).

To relate the body axes x, y, z to the fixed axes x', y', z' (or the external axes), we employ three angles. A number of sets of angles have been described in the literature, but the most common and useful are the Eulerian angles (L. Euler, 1707–1783). The definition of the Eulerian angles varies little from one text to another. We adopt Goldstein's (1980) definition, that is, the Eulerian angles are generated in the following three successive angles of rotation:

Step 1: Rotate the x'-, y'-, z'-axes counterclockwise by an angle φ about the z'-axis, forming the intermediate set of axes x'', y'', z'' of Figure 12.2 (z'' will be the same as z'). The $x''y''$-plane and the $x'y'$-plane are in the same plane; they are produced in Figure 12.3. We now express x', y', x'', and y'' in the plane polar coordinates r and α:

$$\left.\begin{aligned} x' &= r\cos\alpha \\ y' &= r\sin\alpha \end{aligned}\right\} \tag{12.4}$$

$$\left.\begin{aligned} x'' &= r\cos(\alpha-\phi) = r(\cos\alpha\cos\phi + \sin\alpha\sin\phi) \\ y'' &= r\sin(\alpha-\phi) = r(\sin\alpha\cos\phi - \cos\alpha\sin\phi) \end{aligned}\right\}. \tag{12.5}$$

Combining Equations 12.4 and 12.5, with $z'' = z'$, gives the set of transformation for the rotation

$$\left.\begin{aligned} x'' &= x'\cos\phi + y'\sin\phi \\ y'' &= -x'\sin\phi + y'\cos\phi \end{aligned}\right\}. \tag{12.6}$$

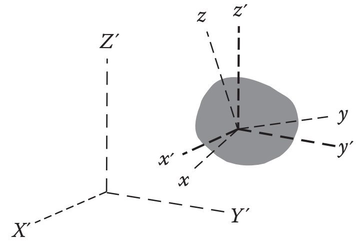


FIGURE 12.1 Primed axes represent an external frame of reference; the unprimed axes are fixed in the rigid body (the body axes).

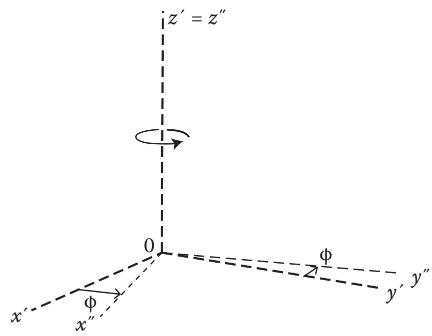


FIGURE 12.2 Rotation defining the angle ϕ.

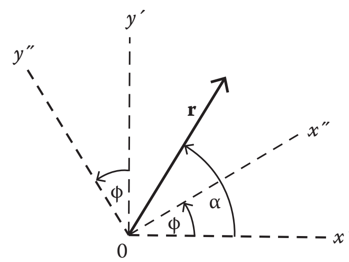


FIGURE 12.3 $x''y''$ plane and $x'y'$ plane of Figure 12.2.

Equation 12.6 can be expressed in matrix notation as

$$\tilde{X}'' = \tilde{D}\tilde{X}' \tag{12.7}$$

where $\tilde{X}''$, $\tilde{X}'$ are column matrices and $\tilde{D}$ is the transformation matrix:

$$\tilde{X}'' = \begin{pmatrix} x'' \\ y'' \\ z'' \end{pmatrix}, \quad \tilde{X}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}, \quad \tilde{D} = \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12.8}$$

Step 2: Rotate the x''-, y''-, z''-axes counterclockwise by an angle θ about the x''-axis, producing the X-, Y-, Z-axes of Figure 12.4. The x''- (X''-) axis is the line of intersection of the $x''y''$- and XY-plane; it is called the line of nodes.

Applying a procedure similar to that used in Step 1 to obtain Equation 12.6, we now arrive at

$$\left.\begin{aligned} X &= x'' \\ Y &= y''\cos\theta + z''\sin\theta \\ Z &= -y''\sin\theta + z''\cos\theta \end{aligned}\right\} \tag{12.9}$$

or, in matrix notation,

$$\tilde{X} = \tilde{C}\tilde{x}'' \tag{12.10}$$

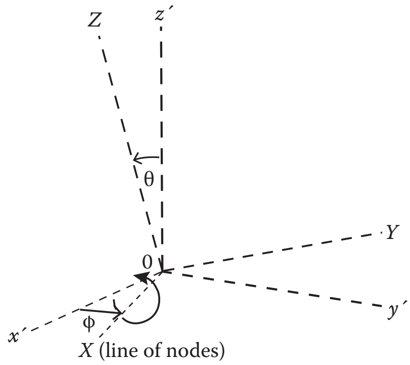


FIGURE 12.4 Rotation defining the angle θ.

where

$$\tilde{C} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta & \sin\theta \\ 0 & -\sin\theta & \cos\theta \end{pmatrix}. \tag{12.11}$$

Step 3: Rotate the X-, Y-, Z-axes counterclockwise by an angle ψ about Z, producing the desired x-, y-, z-axes of Figure 12.5. We now have

$$\left.\begin{aligned} x &= X\cos\psi + Y\sin\psi \\ y &= -X\sin\psi + Y\cos\psi \\ z &= Z \end{aligned}\right\} \tag{12.12}$$

or, in matrix notation,

$$\tilde{x} = \tilde{B}\tilde{X} \tag{12.13}$$

where the transformation matrix $\tilde{B}$ has the same form as $\tilde{D}$

$$\tilde{B} = \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{12.14}$$

Each of the three linear transformations of Equations 12.6, 12.9, and 12.12 is orthogonal, and the rotation matrices $\tilde{B}$, $\tilde{C}$, and $\tilde{D}$ are orthogonal matrices.

Substitution of Equations 12.10 and 12.7 into Equation 12.13 gives the complete transformation from the x'-, y'-, z'-axes to the body coordinates, the x-, y-, z-axes:

$$\tilde{x} = \tilde{B}\tilde{X} = \tilde{B}\tilde{C}\tilde{x}'' = \tilde{B}\tilde{C}\tilde{D}\tilde{x}' = \tilde{A}\tilde{x}' \tag{12.15}$$

where

$$\tilde{A} = \tilde{B}\tilde{C}\tilde{D} \tag{12.16}$$

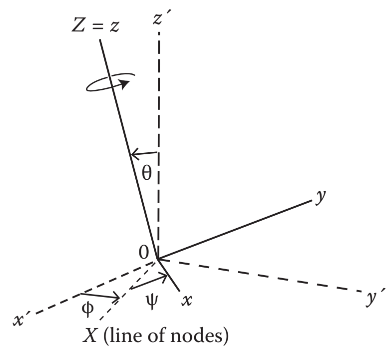


FIGURE 12.5 Rotation defining the angle ψ.

is the product matrix (or the full rotation matrix) and has the following matrix elements:

$$A_{ij} = \sum_{k,l} B_{ik} C_{kl} D_{lj}$$

$$\begin{aligned}
A_{11} &= \cos\psi\cos\phi - \cos\theta\sin\phi\sin\psi & A_{12} &= \cos\psi\sin\phi + \cos\theta\cos\phi\sin\psi & A_{13} &= \sin\psi\cos\theta \\
A_{21} &= -\sin\psi\cos\phi - \cos\theta\sin\phi\cos\psi & A_{22} &= -\sin\psi\sin\phi + \cos\theta\cos\phi\cos\psi & A_{23} &= \cos\psi\sin\theta. \\
A_{31} &= \sin\theta\sin\phi & A_{32} &= -\sin\theta\cos\phi & A_{33} &= \cos\theta
\end{aligned} \quad (12.17)$$

The inverse transformation from the body axes x, y, z to the x'-, y'-, z'-axes

$$\tilde{x}' = \tilde{A}^{-1}\tilde{x} \quad (12.18)$$

is given by the transported matrix $\tilde{A}^t$, and for orthogonal matrices, the transposed matrix is equal to the inverse matrix $\tilde{A}^{-1}$.

12.3 RATE OF CHANGE OF VECTOR

There is no interesting new physics in the translational motion of a rigid body. So we shall concentrate on rotation of a rigid body about some axis. The axis of rotation may be chosen to pass through any point. As shown in Chapter 2, the energy and the angular momentum of a body (many-particle systems) could be expressed as the sum of two contributions from the motion of the CM and the motion about the CM. We can choose the axis of rotation to pass through the CM. This choice will result in complete separation of translation and rotational motion. In what follows, we will always place the origin of the body system at CM, through which passes the axis of rotation. Unless otherwise stated, this will always be understood to be the case.

We begin the study of the rotational motion of rigid bodies with an examination of the way vectors behave under rotation. Consider a vector $\vec{V}$ fixed in a rotating body; the vector will appear to change as the body is observed to rotate. How is the time rate of change of vector $\vec{V}$ in the rotating body coordinate system related to its time rate of change in the fixed nonrotating coordinate axes x,

y, z? If $\vec{V}$ is varying in the rotating body frame, and if we denote its time rate of change measured in the rotating body frame by $\partial\vec{V}/\partial t$, then the time rate of change of $\vec{V}$ as seen from the fixed, nonrotating coordinate axes is given by

$$\left(\mathrm{d}\vec{V}/\mathrm{d}t\right)_{fix} = \left(\partial\bar{V}/\partial t\right)_{rot} + \vec{\omega}\times\vec{V} \tag{12.19}$$

where $\vec{\omega}$ is the angular velocity vector of the rotating body, and it lies along the axis of rotation. Equation 12.19 is a statement of the transformation of the time derivatives between the two coordinate systems. This result was derived in Chapter 11 (Equation 11.23) and will be used frequently in the following sections. No conditions were imposed upon the vector $\vec{V}$ in the course of derivation.

The angular velocity $\vec{\omega}$ is often expressed in terms of the Euler angles and their time derivatives. The general infinitesimal rotation associated with $\vec{\omega}$ can be considered as consisting of three successive infinitesimal rotations with angular velocities: $\omega_\theta = \dot{\theta}$, $\omega_\phi = \dot{\phi}$, $\omega_\psi = \dot{\psi}$. An angular velocity is a vector pointing along the axis of rotation in the direction of the advance of a right-hand screw turning in the sense of the angular velocity, so, accordingly, in Figure 12.5, $\dot{\phi}$ is a vector pointing along Oz', $\dot{\theta}$ a vector along OX, and $\dot{\psi}$ a vector along Oz. It remains to find the components of these vectors along the x, y, and z body axes. The body axes are the most useful for discussing the equations of motion of rigid bodies. By inspection of Figure 12.5, we find the cosines of angles between these components as given in Table 12.1.

Thus, the required relations are

$$\begin{aligned}
\omega_1 &= \omega_x = \dot{\phi}\sin\theta\sin\psi + \dot{\theta}\cos\psi \\
\omega_2 &= \omega_y = \dot{\phi}\sin\theta\cos\psi - \dot{\theta}\sin\psi \\
\omega_3 &= \omega_z = \dot{\phi}\cos\theta.
\end{aligned} \tag{12.20}$$

Alternatively, the transformation matrices $\tilde{A}$, $\tilde{B}$, $\tilde{C}$, and $\tilde{D}$ may be used to furnish the desired relations. In the $Ox'y'z'$ system, the components of $\dot{\phi}$ are $\dot{\phi}_x = 0$, $\dot{\phi}_y = 0$, and $\dot{\phi}_z = \dot{\phi}$. What are the components of $\dot{\phi}$ in the $Oxyz$ body axes? These are obtained by applying the full rotation matrix $\tilde{A}$ to $\dot{\phi}$:

$$\begin{pmatrix} \dot{\phi}_x \\ \dot{\phi}_y \\ \dot{\phi}_z \end{pmatrix} = \tilde{A}\begin{pmatrix} 0 \\ 0 \\ \dot{\phi} \end{pmatrix};$$

the results are

$$\dot{\phi}_x = \dot{\phi}\sin\psi\cos\theta,\ \dot{\phi}_y = \dot{\phi}\cos\psi\sin\theta,\ \dot{\phi}_z = \dot{\phi}\cos\theta. \tag{12.21a}$$

The velocity $\dot{\theta}$ is in the direction of the line of nodes. To obtain the components of $\dot{\theta}$ in the $Oxyz$ body axes, we need the transformation matrix $\tilde{B}$. Application of $\tilde{B}$ to $\dot{\theta}$

TABLE 12.1
Cosines of Angles between Angular Velocities and the Time Derivatives of the Euler Angles

	ω_x	ω_y	ω_z
$\dot{\theta}$	cos ψ	−sin ψ	0
$\dot{\phi}$	sin θ ψ	sin θ cos θ	cos θ
$\dot{\psi}$	0	0	1

$$\begin{pmatrix} \dot{\theta}_x \\ \dot{\theta}_y \\ \dot{\theta}_z \end{pmatrix} = \tilde{B} \begin{pmatrix} \dot{\theta} \\ 0 \\ 0 \end{pmatrix}$$

gives

$$\dot{\theta}_x = \dot{\theta}\cos\psi, \quad \dot{\theta}_y = -\dot{\theta}\sin\psi, \quad \dot{\theta} = 0. \tag{12.21b}$$

Because $\dot{\psi}$ is parallel to the z (body)-axis, no transformation is needed:

$$\dot{\psi}_x = \dot{\psi}, \quad \dot{\psi}_y = 0, \quad \dot{\psi}_z = 0. \tag{12.21c}$$

Collecting the individual components of $\vec{\omega}$, we finally obtain

$$\begin{aligned} \omega_1 &= \dot{\phi}_x + \dot{\theta}_x + \dot{\psi}_x = \dot{\phi}\sin\theta\sin\psi + \dot{\theta}\cos\psi \\ \omega_2 &= \dot{\phi}_y + \dot{\theta}_y + \dot{\psi}_y = \dot{\phi}\sin\theta\cos\psi - \dot{\theta}\sin\psi \\ \omega_3 &= \dot{\phi}_z + \dot{\theta}_z + \dot{\psi}_z = \dot{\phi}\cos\theta + \dot{\psi} \end{aligned} \tag{12.22}$$

which is the same result obtained earlier, Equation 12.20.

12.4 ROTATIONAL KINETIC ENERGY AND ANGULAR MOMENTUM

We regard a rigid body as an assembly of n particles m_α, $\alpha = 1, 2, ..., n$. If the body rotates with an instantaneous angular velocity $\vec{\omega}$ about some axis passing through the CM (Figure 12.6), then the instantaneous velocity of the particle in the fixed coordinate axes can be obtained from Equation 12.19. But we are considering a rigid body, so $(\partial\vec{r}/\partial t)_{rot} = 0$, and Equation 12.19 gives

$$\vec{v} = \vec{\omega} \times \vec{r}.$$

We now see that all velocities are measured in the fixed coordinate axes; all velocities relative to the rotating body's axes must vanish because the body is rigid. The rotational kinetic energy T_{rot} is given by

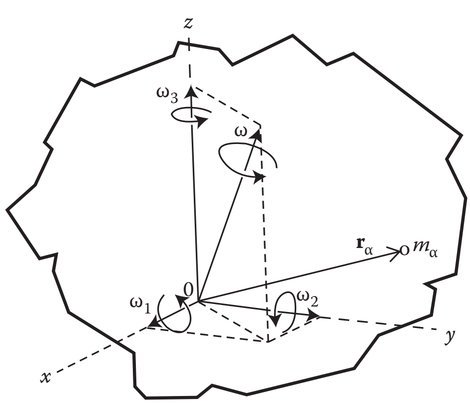


FIGURE 12.6 Rotating rigid body.

$$T_{rot} = \frac{1}{2}\sum_{\alpha=1}^{n} m_\alpha v_\alpha^2 = \frac{1}{2}\sum_{\alpha=1}^{n} m_\alpha \left(\vec{\omega}\times\vec{r}_\alpha\right)\cdot\left(\vec{\omega}\times\vec{r}_\alpha\right)$$
$$= \frac{1}{2}\sum_{\alpha=1}^{n} m_\alpha \left[\omega^2 r_\alpha^2 - \left(\vec{\omega}\cdot\vec{r}_\alpha\right)^2\right] \tag{12.23}$$

where, in the last step, we employed the vector identity

$$\left(\vec{A}\times\vec{B}\right)\cdot\left(\vec{A}\times\vec{B}\right) = A^2B^2 - \left(\vec{A}\cdot\vec{B}\right)^2.$$

We now express T_{rot} in terms of the components of $\vec{r}_\alpha$ and $\vec{\omega}$: $\vec{r}_\alpha = (x_{\alpha,1}, x_{\alpha,2}, x_{\alpha,3})$ and $\vec{\omega} = (\omega_1, \omega_2, \omega_3)$:

$$T_{rot} = \frac{1}{2}\sum_{\alpha=1}^{n} m_\alpha \left[\left(\sum_i \omega_i^2\right)\left(\sum_k x_{\alpha,k}^2\right) - \left(\sum_i \omega_i x_{\alpha,i}\right)\left(\sum_j \omega_j x_{\alpha,j}\right)\right].$$

This can be simplified as follows:

$$T_{rot} = \frac{1}{2}\sum_{ij}\sum_{\alpha} m_\alpha \left[\omega_i\omega_j\delta_{ij}\left(\sum_k x_{\alpha,k}^2\right) - \omega_i\omega_j x_{\alpha,i}x_{\alpha,j}\right]$$
$$= \frac{1}{2}\sum_{ij}\omega_i\omega_j\sum_{\alpha} m_\alpha\left[\delta_{ij}\sum_k x_{\alpha,k}^2 - x_{\alpha,i}x_{\alpha,j}\right] \tag{12.24}$$
$$= \frac{1}{2}\sum_{ij} I_{ij}\omega_i\omega_j$$

where

$$I_{ij} = \sum_{\alpha} m_\alpha\left[\delta_{ij}\sum_k x_{\alpha,k}^2 - x_{\alpha,i}x_{\alpha,j}\right]. \tag{12.25}$$

In matrix notation, we have

$$T_{rot} = \frac{1}{2}\tilde{\omega}^t \cdot \ddot{I} \cdot \tilde{\omega} \tag{12.26}$$

with

$$\tilde{I} = \begin{pmatrix} I_{11} & I_{12} & I_{13} \\ I_{21} & I_{22} & I_{23} \\ I_{31} & I_{32} & I_{33} \end{pmatrix}, \quad \tilde{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix}, \quad \tilde{\omega}^t = (\omega_1, \omega_2, \omega_3). \tag{12.27}$$

The object represented by $\tilde{I}$ is known as the inertia tensor (the matrix representation of the inertia tensor). They are constants that depend on the properties of the rigid body and upon the location of the origin and orientation of the body axes. The general properties of the inertia tensor are discussed in the next section.

The angular momentum of a rigid body can likewise be expressed in terms of the inertial tensor. The angular momentum $\tilde{L}$ about the origin O is

$$\tilde{L} = \sum_{\alpha} \vec{r}_{\alpha} \times \vec{p}_{\alpha} = \sum_{\alpha} m_{\alpha} \vec{r}_{\alpha} \times \vec{v}_{\alpha} = \sum_{\alpha} m_{\alpha} \vec{r}_{\alpha} \times \left(\vec{\omega} \times \vec{r}_{\alpha} \right). \tag{12.28}$$

Expanding the vector triple by using the vector identity

$$\vec{A} \times \left(\vec{B} \times \vec{C} \right) = \vec{B} \left(\vec{A} \cdot \vec{C} \right) - \vec{C} \left(\vec{A} \cdot \vec{B} \right),$$

we finally obtain

$$\vec{L} = \sum_{\alpha} m_{\alpha} \left[r_{\alpha}^{2} \vec{\omega} - \vec{r}_{\alpha} \left(\vec{r}_{\alpha} \cdot \vec{\omega} \right) \right]. \tag{12.29}$$

Similarly, we can express the angular momentum vector $\vec{L}$ in terms of the components of the vectors $\vec{r}_{\alpha}$ and $\vec{\omega}$. The ith component of $\vec{L}$, L_i, is

$$\begin{aligned} L_i &= \sum_{\alpha} m_{\alpha} \left[\omega_i \sum_{k} x_{\alpha,k}^{2} - x_{\alpha,j} \sum_{j} x_{\alpha,j} \omega_j \right] \\ &= \sum_{\alpha} m_{\alpha} \sum_{j} \left[\omega_j \delta ij \sum_{k} x_{\alpha,k}^{2} - x_{\alpha,i} x_{\alpha,j} \omega_j \right] \\ &= \sum_{j} \omega_j \sum_{\alpha} m_{\alpha} \left[\delta ij \sum_{k} x_{\alpha,k}^{2} - x_{\alpha,i} x_{\alpha,j} \right] \\ &= \sum_{l} I_{ij} \omega_j. \end{aligned} \tag{12.30}$$

In matrix notation, we would have

$$\tilde{L} = \tilde{I} \cdot \tilde{\omega}. \tag{12.31}$$

In general, $\tilde{L}$ is not necessarily parallel to $\vec{\omega}$. Comparing Equations 12.31 and 12.26, we find that the rotational kinetic energy can be expressed as

$$T_{rot} = \frac{1}{2} \tilde{\omega} \cdot \tilde{L}. \tag{12.32}$$

The equation of motion for the general rotation of a rigid body about its CM or about a fixed point in space is

$$\vec{N}^{(e)} = \frac{\mathrm{d}\vec{L}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t} \left(\tilde{I} \cdot \tilde{\omega} \right) \quad \text{or} \quad N_j^{(e)} = \frac{\mathrm{d}L_j}{\mathrm{d}t} = \sum_{k} \frac{\mathrm{d}}{\mathrm{d}t} (I_{jk} \omega_k) \tag{12.33}$$

which simplifies considerably for the case of rotation about a single fixed axis. For definiteness, we choose the z-axis as the axis of rotation, $\vec{\omega} = \omega\hat{k}$. The components of the angular momentum in Equation 12.32 are then

$$L_x = I_{xz}\omega_z,\ L_y = I_{yz}\omega_z,\ L_z = I_{zz}\omega_z, \tag{12.34}$$

and the equations of motion (Equation 12.33) reduce to

$$N_x^{(e)} = \frac{\mathrm{d}}{\mathrm{d}t}(I_{xz}\omega_z), \quad N_y^{(e)} = \frac{\mathrm{d}}{\mathrm{d}t}(I_{yz}\omega_z), \quad N_z^{(e)} = \frac{\mathrm{d}}{\mathrm{d}t}(I_{zz}\omega_z). \tag{12.35}$$

If the rigid body is symmetrical about the z-axis, we have from Equation 12.25 that

$$I_{xz} = I_{yz} = 0, \quad I_{zz} = \sum_\alpha m_\alpha R_\alpha^2$$

where R_α is the perpendicular distance from the axis of rotation to the mass element α, and it is sometimes called the lever arm (or the moment arm) of the element relative to the axis of rotation. The torques $N_x^{(e)}$ and $N_y^{(e)}$ in Equation 12.35 are then zero. If I_{xz} or I_{yz} is nonzero, the body is unbalanced and the bearing must provide the torque $N_x^{(e)}$ or $N_y^{(e)}$ to keep the axis of rotation from moving.

Because I_{zz} is time-independent, the equation of motion for the z-component in Equations 12.35 reduces to

$$N_z^{(e)} = \mathrm{d}L_z/\mathrm{d}t = I_{zz}\dot{\omega}_z$$

which has the same mathematical structure as the equation for linear motion in one dimension $F_z = M\dot{v}_z$. However, for the general case, we have to go back to Equation 12.33 as $\vec{L}$ and $\vec{\omega}$ are not necessarily parallel.

The analogy between the physical quantities of linear motion and angular motion about a fixed axis is shown below.

Angular Motion	**Linear Motion**
Moment of inertia I_{zz}	Mass, m
Angular position, θ	Linear position, z
Angular velocity, $\omega_z = \mathrm{d}\theta/\mathrm{d}t$	Linear velocity, $v_z = \mathrm{d}z/\mathrm{d}t$
Angular acceleration, $\dot{\omega}_z$	Linear acceleration, $a_z = \dot{v}_z$
Angular momentum, $L_z = I_{zz}\omega_z$	Linear momentum, $p_z = mv_z$
Kinetic energy, $T_{rot} = 1/2 I_{zz}\omega_z^2$	Kinetic energy, $T = 1/2\ mv_z^2$
Torque	Force

Example 12.1: Physical Pendulum

The oscillations resulting from gravity on a body suspended from a fixed horizontal axis provide a good example of the motion of a rigid body with one point fixed. This is called a physical or compound pendulum (Figure 12.7). First, we need to calculate the total torque about the axis of suspension through point 0, which is at a distance b from the CM C. Because gravity is acting

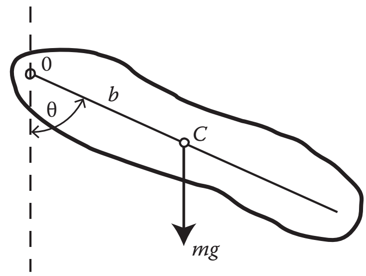


FIGURE 12.7 Physical pendulum.

downward, there is a force mg acting at point C. The resultant torque about the axis for any displacement, therefore, has the magnitude

$$N = mgb \sin \theta. \tag{12.36}$$

And the equation of motion takes the simple form

$$I\ddot{\theta} = -mgb\sin\theta \tag{12.37}$$

in which the minus sign occurs because the torque tends to produce rotation in the direction of decreasing θ, and I is the moment of inertia of the body about the axis of rotation passing through 0. This equation is identical in form to the equation of motion for a simple pendulum.

For small displacement, $\sin \theta \approx \theta$, Equation 12.37 becomes

$$\ddot{\theta} + \frac{mgb}{I}\theta = 0 \tag{12.38}$$

and its solution is

$$\theta = A\sin\left(\sqrt{mgb/I}t + B\right) \tag{12.39}$$

in which A and B are arbitrary constants that can be determined by the initial conditions of the motion. Equation 12.39 represents an oscillation of frequency f_0 and period T_0, where

$$f_0 = \frac{1}{2\pi}\sqrt{mgb/I}, T_0 = \frac{1}{f_0} = 2\pi\sqrt{I/mgb}.$$

Recall a simple pendulum of length l; its frequency is $f_0 = (1/2\pi)(g/l)^{1/2}$. Thus, we see that a physical pendulum is equivalent (in frequency or period) to a simple pendulum of length l defined by

$$l = I/mgb,$$

and l is sometimes called the length of the equivalent simple pendulum.

Example 12.2: Plane (Laminar) Motion of a Rigid Body

Next to rotation about a fixed axis, the simplest case of motion of a rigid body is that in which all its particles move in planes parallel to a fixed plane. This, in general, involves both translation and rotation. As a special case, consider a round object (ball, cylinder, etc.) of mass m and radius

b rolling under gravity down a perfect, rough inclined plane on which no slipping can take place (Figure 12.8). The forces acting on the object are the weight mg at the CM 0, the reaction $\vec{R}$ of the plane normal to the plane at contact point A, and the force of friction $\vec{F}_f$ directed up the plane if the object rolls down. As the object rolls, the CM 0 translates, while the particles composing the object rotate about 0. If we denote the displacement of the CM parallel to the plane by x, its equation of motion becomes

$$m\ddot{x}_{cm} = mg\sin\theta - F_f. \tag{12.40}$$

The resultant torque about 0 is wholly a result of F_f, and we, therefore, have for the rotational motion about 0,

$$I_{cm}\dot{\omega} = F_f b \tag{12.41}$$

where I_{cm} is the moment of inertia about the axis of the rotation, and $\dot{\omega}$ is the angular acceleration about this axis.

If no slipping can occur, then $x_{cm} = b\varphi$, and from this, we have

$$\dot{x}_{cm} = b\dot{\phi} = b\omega \text{ and } \ddot{x}_{cm} = b\ddot{\phi} = b\dot{\omega} \tag{12.42}$$

where φ is the angle of rotation. Substituting Equation 12.42 into Equation 12.40, we obtain

$$\ddot{x} = \frac{mg\sin\theta}{m + I_{cm}/b^2}. \tag{12.43}$$

The angular acceleration $\dot{\omega}$ is, from Equation 12.42,

$$\dot{\omega} = \frac{\ddot{x}_{cm}}{b} = \frac{mg\sin\theta}{b(m + I_{cm}/b^2)}. \tag{12.44}$$

The object therefore rolls down the inclined plane with constant linear acceleration and also with constant angular acceleration. For a homogeneous sphere, $I_{cm} = 2mb^2/5$, and for a homogeneous right circular cylinder, $I_{cm} = mb^2/2$.

The preceding results can also be obtained from energy consideration. Because no mechanical energy is lost to heat in pure rolling, energy is conserved:

$$E = T_{rot} + T_{tran} + V = \text{constant}.$$

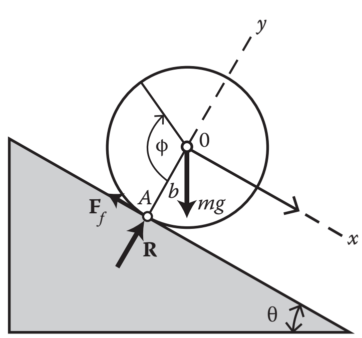


FIGURE 12.8 Round object rolling down an inclined plane.

Now,

$$T_{rot} = \frac{1}{2} I_{cm}\omega^2, \quad T_{tran} = \frac{1}{2} m\dot{x}_{cm}^2, \quad \text{and} \quad V = mgz_{cm}$$

where z_{cm} is the vertical distance of the CM and is given by $z_{cm} = -x_{cm} \sin\theta$; we also have $\omega = \dot{x}_{cm}/b$. Substituting these into the energy conservation equation yields

$$\frac{1}{2}\frac{I_{cm}\dot{x}_{cm}^2}{b^2} + \frac{1}{2} m\dot{x}_{cm}^2 - mgx_{cm}\sin\theta = E. \tag{12.45}$$

Differentiating this equation with respect to t and then solving for x_{cm}, we find the same result, Equation 12.43.

Example 12.3: Angular Momentum of a Rotating Dumbbell

Consider a dumbbell formed by two point masses m at the ends of a massless rod of length $2b$. The dumbbell rotates at a fixed inclination α with constant angular velocity $\vec{\omega}$ about a pivot at the center of the rod as shown in Figure 12.9. Find the angular momentum of the system.

Solution:

The most direct method is to calculate the angular momentum from the definition $\vec{L} = \sum_i \vec{r}_i \times \vec{p}_i$. Here, each mass moves in a circle of radius $b\cos\alpha$ with angular speed ω.

We, however, choose to calculate $\vec{L}$ from Equation 12.31. First, we resolve $\vec{\omega} = \omega\hat{k}$ into components $\omega_\perp$ and $\omega_\parallel$, which are perpendicular and parallel to the rod. From Figure 12.10, we see that

$$\omega_\perp = \omega \cos\alpha, \text{ and } \omega_\parallel = \omega \sin\alpha.$$

Because the masses are point particles, $\omega_\parallel$ produces no angular momentum. Hence, the angular momentum is a result entirely of $\omega_\perp$. The momentum of inertia about the direction of $\omega_\perp$ is $2mb^2$, and the magnitude of the angular momentum is

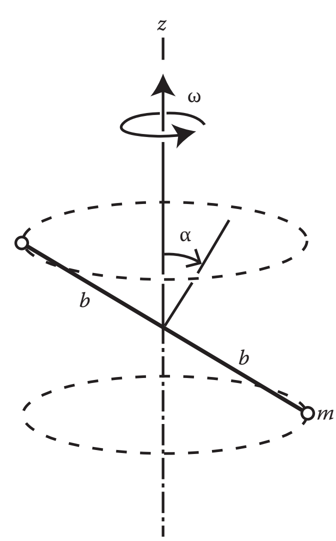


FIGURE 12.9 Rotating dumbbell.

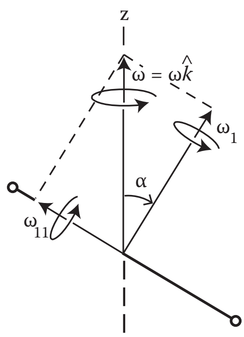


FIGURE 12.10 Resolving $\vec{\omega}$ into components $\omega_\perp$ and $\omega_\parallel$.

$$L = I\omega_\perp = 2\ mb^2\ \omega_\perp = 2mb^2\ \omega \cos \alpha$$

where $\vec{L}$ points along the direction of $\omega_\perp$. Hence, $\vec{L}$ swings around the rod, and the tip of $\vec{L}$ traces out a circle about the z-axis.

Note that the angular momentum is constant in magnitude but changes in direction ($\vec{L}$ rotates in space with the rod). Now the time rate of change of angular momentum is equal to the external torque:

$$\frac{d\vec{L}}{dt} = \vec{N}^{(e)}.$$

This equation refers to a fixed (inertial) frame. We can find $d\vec{L}/dt$ easily by decomposing $\vec{L}$ into L_z and L_h as shown in Figure 12.11. The component $L_z = L \cos \alpha$, parallel to the z-axis, is constant. Hence, there is no torque in the z-direction. The horizontal component $L_h = L \sin \alpha$ swings with the rod.

Example 12.4: Simple Gyroscopic Precession

Gyroscope precession is a fascinating phenomenon, and it is also relatively hard to understand. We now try to understand it by using the basic concepts of angular momentum, torque, and the time derivative of a vector.

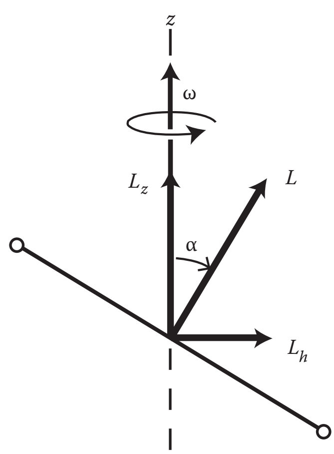


FIGURE 12.11 Decomposing $\vec{L}$ into L_z and L_h.

A gyroscope consists primarily of a small heavy flywheel and an axle that can be set into rapid spinning. The end of the axle rests on a support (a free pivot), allowing the axis to take various orientations without constraint. If the spinning wheel is released horizontally with one end supported by the pivot, it wobbles horizontally and then settles down to uniform precession, in which the axle slowly rotates about the vertical with a constant slant angular velocity, $\vec{\Omega}$, as shown in Figure 12.12.

If the gyroscope is released with the flywheel spinning rapidly, it will precess. In this case, the large spin angular momentum of the flywheel $\vec{L}_s$ dominates the dynamics of the system. Nearly all of the angular momentum lies in $\vec{L}_s$, which is directed along the axle and has magnitude $L_s = I_0\omega_s$, where I_0 is the moment of inertia of the wheel about its axle, and ω_s is the wheel's spin angular velocity. If the spin is counterclockwise, $\vec{L}_s$ points along the positive x-axis. There is torque on the wheel. If we take the pivot as the origin, the torque is a result of the weight of the wheel acting at the end of the axle. The magnitude of the torque is

$$N^{(e)} = mgd \tag{12.46}$$

where d is the distance of the CM of the flywheel from the pivot point 0. The direction of $\vec{N}^{(e)}$ is given by the right-handed corkscrew rule, that is, it is in the y-direction, and it causes $\vec{L}_s$ to change. According to Equation 12.33, the change in $\vec{L}_s$ in an infinitesimal time interval dt is

$$\mathrm{d}\vec{L}_s = \vec{N}^{(e)}\mathrm{d}t. \tag{12.47}$$

The increment $\mathrm{d}\vec{L}_s$ is parallel to $\vec{N}^{(e)}$ and perpendicular to $\vec{L}_s$ as shown in Figure 12.13. And the angular momentum $\vec{L}_s$ rotates counterclockwise in the xy-plane. From Figure 12.13, we have

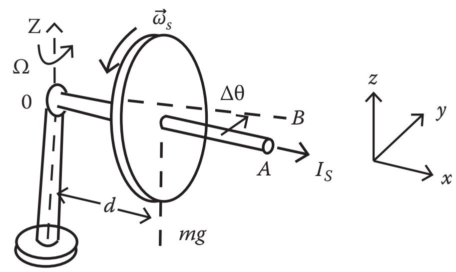


FIGURE 12.12 Sketch of gyroscope.

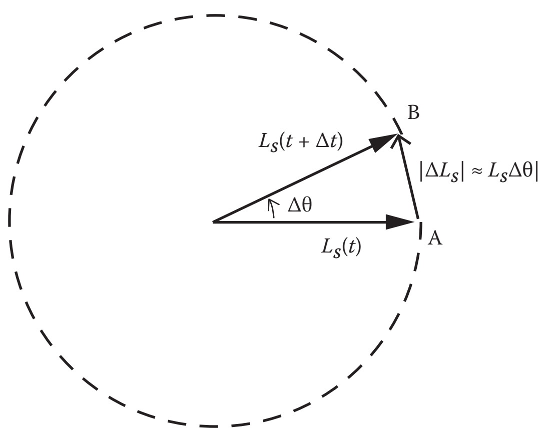


FIGURE 12.13 Increment $\mathrm{d}\vec{L}_s$ is parallel to $\vec{N}^{(e)}$ and perpendicular to $\vec{L}_s$.

$$\left|\Delta\vec{L}_s\right| = \left|\vec{L}_s\right|\Delta\theta$$

from which it follows that

$$\left|\frac{\mathrm{d}\vec{L}_s}{\mathrm{d}t}\right| = L_s\frac{\mathrm{d}\theta}{\mathrm{d}t} = L_s\Omega \tag{12.48}$$

where Ω is the angular velocity of precession. The direction of $\mathrm{d}\vec{L}_s/\mathrm{d}t$ is tangential to the horizontal circle swept out by $\vec{L}_s$, parallel to $\vec{N}^{(e)}$, as we expected.

Solving Equations 12.46 through 12.48 yields the angular velocity of the precession:

$$\Omega = \frac{mgd}{I_0\omega_s}. \tag{12.49}$$

This result is a good approximation when the applied torque is small or the spin is high.

The angular momentum associated with the precession motion was ignored in our discussion. When the wheel precesses about the z-axis, it has a small orbital angular momentum in the z-direction. This orbital angular momentum is constant in magnitude and direction as long as the precession is uniform and, therefore, plays no dynamic role.

Equation 12.49 indicates that Ω increases as the flywheel slows. (This effect is easy to see with a toy gyroscope.) When Ω becomes so large, we cannot neglect small changes in the angular momentum about the vertical axis resulting from frictional torque. Thus, eventually uniform precession gives way to a violent and erratic motion.

We have considered the case that the axle of the gyroscope is horizontal. Actually, the rate of uniform precession is independent of the angle of elevation as shown below.

Example 12.5

Consider a gyroscope in uniform precession with its axle at angle φ with the vertical (Figure 12.14). Show that the angular velocity of precession Ω is independent of the angle of elevation φ.

Solution:

The component of $\vec{L}_s$ in the xy-plane varies as the gyroscope precesses, and the component parallel to the z-axis remains constant. The horizontal component of $\vec{L}_s$ is $L_s\sin\varphi$. Hence,

$$\left|\mathrm{d}\vec{L}_s/\mathrm{d}t\right| = \Omega l\sin\phi.$$

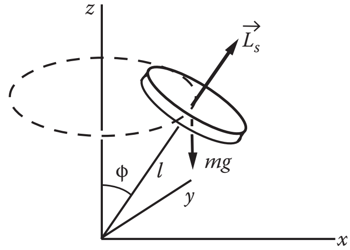


FIGURE 12.14 Gyroscope in uniform precession with its axle at angle ϕ with the vertical.

The torque resulting from the weight *mg* is horizontal and has magnitude *mgl*sin φ. We have

$$\Omega L_s \sin\phi = mgl \sin\phi$$

from which we find

$$\Omega = \frac{mgl}{I_0\omega_s}.$$

We see that the angular velocity of precession Ω is independent of the angle of elevation φ.

12.5 INERTIA TENSOR

Various types of quantities are encountered in the discussion of physical phenomena. Two familiar types that we have used frequently in this book are scalars and vectors. The quantity *I* that appeared in the previous section is a new type of quantity, a tensor. The general theory of tensor analysis deals with tensors referred to generalized curvilinear coordinates. We do not need such a general theory, and so it will not be treated. We actually deal with tensors referred to a right-handed Cartesian coordinate system. Such tensors are known as Cartesian tensors.

All the quantities that we are concerned with can be considered to be tensor quantities of one kind or another. The different kinds of tensors are classified according to the order (or rank and valence). The number of free (unsummed) indices present in the term gives the order of a tensor. If there are no free indices, the quantity concerned is a tensor of order zero, which is a scalar quantity. If only one index is free, the quantity is a tensor of first order that is a vector quantity. If two indices are free, the order is two, and so on.

A zero-order tensor (a scalar) represents just a number, which may be real or complex. A first-order Cartesian tensor represents three numbers that are the magnitudes of the Cartesian components of a vector quantity. We write this array of numbers as $\vec{A} = (A_1, A_2, A_3)$ or as a row matrix $\tilde{A} = (A_1, A_2, A_3)$ or as a column matrix. A second-order tensor represents an array of nine numbers because each of the two unsummed indices may take on the values 1, 2, or 3. We write these nine scalar quantities in a matrix array:

$$\tilde{A} = \begin{pmatrix} A_{11} & A_{12} & A_{13} \\ A_{21} & A_{22} & A_{23} \\ A_{31} & A_{32} & A_{33} \end{pmatrix}$$

and it is the matrix representation of the second-order tensor *A*.

In what follows, we shall use the word “tensor” to mean a Cartesian tensor of order two. Tensors occur most frequently when one vector, say, $\vec{B}$, is given as a linear function of another vector, say, $\vec{A}$:

$$B_x = T_{xx}A_x + T_{xy}A_y + T_{xz}A_z$$

$$B_y = T_{yx}A_x + T_{yy}A_y + T_{yz}A_z$$

$$B_z = T_{zx}A_x + T_{zy}A_y + T_{zz}A_z$$

These three equations can be written compactly in terms of a single expression:

$$B_i = \sum_k T_{ik}A_k,\ i = x, y, z. \tag{12.50}$$

Assuming that $i = x$ and performing a summation with the subscript k sequentially with the values x, y, z, we get the first of the equations preceding Equation 12.50, while assuming that $i = y$, we get the second, and so forth. In matrix notation, we have

$$\begin{pmatrix} B_x \\ B_y \\ B_z \end{pmatrix} = \begin{pmatrix} T_{xx} & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} \end{pmatrix} \begin{pmatrix} A_x \\ A_y \\ A_z \end{pmatrix}. \tag{12.51}$$

The nine quantities T_{ij} are the components of a tensor, which we shall denote by $\tilde{T}$. An example is the relation 12.30 between the angular momentum vector $\vec{L}$ and the angular velocity vector $\vec{\omega}$ of a rigid body in the most general case. The quotient of $\vec{L}$ and $\vec{\omega}$ is the inertia tensor $\tilde{I}$.

It should be noted that the quotient of two vectors could not be defined consistently within the class of vectors. Just like the quotient of two integers is, in general, not an integer, but rather a rational number. Therefore, it should not be surprising to find that $\tilde{I}$ is a new type of quantity, a Cartesian tensor of the second order. By "inertia tensor," we mean the matrix representing it. The truth of our claim that $\tilde{I}$ is a tensor depends on the establishment of its behavior under a coordinate transformation. Under a coordinate transformation, its nine components should transform according to the following rule:

$$I'_{ij} = \sum_{k,m} \lambda_{ik}\lambda_{jm} I_{km} \tag{12.52}$$

where λ_{ik} are the components of the transformation matrix $\tilde{\lambda}$. To prove this, let us return to Equation 12.30:

$$L_k = \sum_{ki} I_{ki}\omega_i. \tag{12.30}$$

Because it is a vector relationship, an entire analogous expression should exist in any coordinate system rotated with respect to the system in which Equation 12.30 applies:

$$L'_k = \sum_{ki} I'_{ki}\omega'_i. \tag{12.30a}$$

Here, the primed quantities refer to the rotated system. Now, the corresponding vectors in the two systems obey an orthogonal transformation (Chapter 2); thus, we write

$$L_k = \sum_m \lambda_{mk} L'_m, \quad \omega_i = \sum_j \lambda_{ji}\omega'_j.$$

Substituting these into Equation 12.30 gives

$$\sum_m \lambda_{mk} L'_m = \sum_i I_{ki} \sum_j \lambda_{ji}\omega'_j.$$

Multiply both sides of this equation by λ_{ik} and sum over k. The result is, after appropriate rearrangement,

$$\sum_m \left(\sum_k \lambda_{ik} \lambda mk \right) L'_m = \sum_j \left(\sum_{kj} \lambda_{ik} \lambda_{jl} I_{kl} \right) \omega'_j$$

which becomes

$$\sum_m \delta_{im} L'_m = \sum_j \left(\sum_{kj} \lambda_{ik} \lambda_{jl} I_{kl} \right) \omega'_j.$$

Summing over m gives

$$L'_i = \sum_j \left(\sum_{kj} \lambda_{ik} \lambda_{jl} I_{kl} \right) \omega'_j.$$

Comparing this with Equation 12.30a, we obtain the desired transformation rule, Equation 12.52:

$$I'_{ij} = \sum_{k,m} \lambda_{ik} \lambda_{jm} I_{km}.$$

Although this transformation rule is derived for the inertia tensor, it holds generally for all second-order tensors T. In matrix form, Equation 12.52 becomes

$$\tilde{T}' = \tilde{\lambda} \tilde{T} \tilde{\lambda}^t$$

where we have transposed the second orthogonal matrix, thereby inverting the order of its multiplication with $\tilde{T}$. The transposed tensor $\tilde{T}^t$ is obtained from $\tilde{T}$ by reflecting in the leading diagonal or, equivalently, interchanging the row and column indices. Because, for orthogonal matrices, the transposition of the matrix is equal to its inverse, we also have

$$\tilde{T}' = \tilde{\lambda} \tilde{T} \tilde{\lambda}^{-1}.$$

This type of transformation is called a similarity transformation.

The tensor $\tilde{T}$ is called symmetric if $\tilde{T}^t = \tilde{T}$ and anti-symmetric (or skew-symmetric) if $\tilde{T}^t = -\tilde{T}$. The special tensor

$$\tilde{1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

is called the unit (or identity) tensor with the property $\tilde{1}\tilde{T} = \tilde{T}$ for all tensor $\tilde{T}$.

12.5.1 Diagonalization of a Symmetric Tensor

Any symmetric tensor can be diagonalized by a suitable choice of axes. To see this, let $\tilde{T}$ be a symmetric tensor (actually its matrix representation) and $\vec{A}$ an arbitrary vector. $\vec{A}$ is called an eigenvector of $\tilde{T}$ with eigenvalue λ if

$$\tilde{T}\vec{A} = \lambda\vec{A}$$

or, equivalently,

$$\left(\tilde{T} - \lambda\tilde{1}\right)\vec{A} = 0.$$

This equation has a nontrivial solution if the determinant of the coefficients vanishes:

$$\det\left(\tilde{T} - \lambda\tilde{1}\right) = \begin{pmatrix} T_{xx} - \lambda & T_{xy} & T_{xz} \\ T_{yx} & T_{yy} - \lambda & T_{yz} \\ T_{zx} & T_{zy} & T_{zz} - \lambda \end{pmatrix} = 0.$$

This cubic equation for *A*- has three roots that are either all real or one real and one complex conjugate pair. But for a symmetric tensor, the latter possibility can be ruled out. To see this, suppose that λ is complex, and the corresponding eigenvector $\vec{A}$ may also be complex. Taking the complex conjugate of $\tilde{T}\vec{A} = \lambda\vec{A}$, we have $\tilde{T}\vec{A}^* = \lambda\vec{A}^*$. Multiplying these equations by $\vec{A}^*$ and $\vec{A}$, respectively, we obtain

$$\vec{A}^*\tilde{T}\vec{A} = \lambda\vec{A}^*\vec{A},\ \vec{A}\tilde{T}\vec{A}^* = \lambda^*\vec{A}\vec{A}^*.$$

Because $\tilde{T}$ is symmetric, the left-hand sides of these equations must be equal. Hence, the right-hand sides must be equal to $\lambda\vec{A}^*\vec{A} = \lambda^*\vec{A}\vec{A}^*$. But $\vec{A}^*\vec{A} = \vec{A}\vec{A}^* = \left|A_x\right|^2 + \left|A_y\right|^2 + \left|A_z\right|^2 > 0$; thus, $\lambda = \lambda^*$, that is, λ is real. Therefore, for a symmetric tensor, there are three real eigenvalues λ_1, λ_2, λ_3, and three corresponding eigenvectors $\vec{A}_1$, $\vec{A}_2$, $\vec{A}_3$. We can show that $\vec{A}_1$, $\vec{A}_2$, and $\vec{A}_3$ are orthogonal if the three eigenvalues λ_1, λ_2, and λ_3 are distinct. We leave this proof as homework. Next, it is clear that if $\vec{A}_1$ is an eigenvalue, then so is any multiple of $\vec{A}_1$. This means that we can choose to normalize it: $\hat{e}_1 = \vec{A}_1/A_1$. Then, the three eigenvectors form an orthonormal triad $\hat{e}_1$, $\hat{e}_2$, $\hat{e}_3$. If we choose these as our axes, then $\tilde{T}$ must take the diagonal form

$$\tilde{T} = \begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}.$$

Applying this to the inertia tensor, we shall see, in the following section, that the eigenvectors are the principal axes, and the corresponding eigenvalues are the principal moments of inertia.

12.5.2 Moments and Products of Inertia

We now return to the inertial tensor. The *ij*th components of this inertia tensor are defined in Equation 12.25:

$$I_{ij} = \sum_\alpha m_\alpha \left[\delta_{ij}\sum_k x_{\alpha k}^2 - x_{\alpha,i}x_{\alpha,j}\right].$$

Clearly, the inertia tensor is symmetric $I_{ij} = I_{ji}$. Therefore, there are only six independent components; they are (with $x_1 = x$, $x_2 = y$, $x_3 = z$)

$$
\begin{aligned}
I_{11} &= \sum_{\alpha} m_{\alpha}\left(r_{\alpha}^{2} - x_{\alpha}^{2}\right) \\
I_{22} &= \sum_{\alpha} m_{\alpha}\left(r_{\alpha}^{2} - y_{\alpha}^{2}\right) \\
I_{33} &= \sum_{\alpha} m_{\alpha}\left(r_{\alpha}^{2} - z_{\alpha}^{2}\right)
\end{aligned} \tag{12.53}
$$

and

$$
\begin{aligned}
I_{12} = I_{21} &= -\sum_{\alpha} m_{\alpha} x_{\alpha} y_{\alpha} \\
I_{13} = I_{31} &= -\sum_{\alpha} m_{\alpha} x_{\alpha} z_{\alpha} \\
I_{23} = I_{32} &= -\sum_{\alpha} m_{\alpha} y_{\alpha} z_{\alpha}
\end{aligned} \tag{12.54}
$$

where $I_{11} = I_{xx}$, $I_{12} = I_{xy}$, and so forth. The diagonal elements I_{11}, I_{22}, and I_{33} are called the moments of inertia about the corresponding axes. They play a role in rotational motion similar to that of mass in translational motion. The off-diagonal components I_{12}, I_{13}, and so forth are known as the products of inertia. These components are collectively referred to as the inertia parameters of the system.

For a continuous body with mass density $\rho = \rho(r)$, the sum in Equation 12.54 becomes an integral over the volume of the body:

$$
I_{ij} = \int_{V} \rho(r)\left[\delta_{ij}\sum_{k} x_{k}^{2} - x_{i}x_{j}\right]\mathrm{d}V \tag{12.55}
$$

where $\mathrm{d}V = \mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3$ is the element of volume at the position defined by the vector $\vec{r}$, and V is the volume of the body.

12.5.3 Parallel-Axis Theorem

From the preceding discussion, it is obvious that the moments and products of inertia are functions of the point through which the axis of rotation passes and of the orientation of the body with respect to the fixed (or equivalently, the external) coordinate axes. We also saw that in order to separate rotational motion from translational motion, it is necessary to place the origin of the body's coordinate axes at the CM through which the axis of rotation passes. However, for certain geometrical shapes, it may not always be convenient to compute the moments and products of inertia using such a body coordinate system. We now show how the products and moments of inertia about any point of the body are related to the products and moments of inertia with respect to a parallel coordinate system located at the CM. For this purpose, let us consider two parallel coordinate systems O and P as shown in Figure 12.15. The ijth component of the inertia tensor in the P system may be written as

$$
J_{ij} = \sum_{\alpha} m_{\alpha}\left[\delta_{ij}\sum_{k} X_{\alpha,k}^{2} - X_{\alpha,i}X_{\alpha,j}\right]. \tag{12.56}
$$

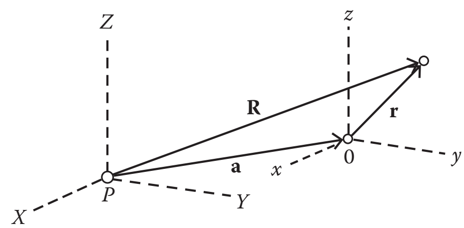


FIGURE 12.15 Two parallel coordinate systems O and P.

But

$$\vec{R} = \vec{r} + \vec{a} \quad \text{or} \quad X_i = x_i + a_i.$$

Substituting these into Equation 12.56), J_{ij} becomes

$$J_{ij} = \sum_\alpha m_\alpha \left[\delta_{ij} \sum_k x_{\alpha,k}^2 - x_{\alpha,i} x_{\alpha,j} \right] + \sum_\alpha m_\alpha \left[\delta_{ij} \sum_k \left(2x_{\alpha,k} a_k + a_k^2 \right) - \left(a_i x_{\alpha,j} + a_j x_{\alpha,j} + a_i a_j \right) \right]. \tag{12.57}$$

The first term is the ijth component I_{ij} in the O system. Equation 12.57 becomes, after regrouping,

$$J_{ij} = I_{ij} + \sum_\alpha m_\alpha \left[\delta_{ij} \sum_k a_k^2 - a_i a_j \right] + \sum_\alpha m_\alpha \left[2\delta_{ij} \sum_k x_{\alpha,k} a_k - a_j x_{\alpha,j} \right]. \tag{12.58}$$

If the origin O is located at the CM, then the last summation vanishes because each term in the last summation involves a sum of the form $\sum m_a x_a$, which is zero if O is the CM, and Equation 12.58 reduces to

$$J_{ij} = I_{ij} + \sum_\alpha m_\alpha \left[\delta_{ij} \sum_k a_k^2 - a_i a_j \right] = I_{ij} + M\left(a^2 \delta_{ij} - a_i a_j \right) \tag{12.59}$$

where

$$M = \sum m_a \text{ and } a^2 = \sum a_k^2.$$

The second term on the right-hand side is the inertia tensor of the body with respect to the CM. The inertia tensor (J) about a point is expressible as the sum of the inertia tensor (I) of the CM and that with respect to the CM. This is known as Steiner's parallel-axis theorem for the inertia tensor (Jacob Steiner, 1796–1863).

Example 12.6

Find the moments and products of inertia for a homogeneous rectangular parallelepiped with edges a, b, and c, of density ρ.

Solution:

First, we select two coordinate systems. The coordinate axes of the first system lie along the three edges with the origin located at one corner. The second system is parallel to the first system, and its origin is at the CM (Figure 12.16). We first calculate the inertia tensor in the CM system. According to Equation 12.55, we have

$$I_{11} = \rho\int_{-a/2}^{a/2}\int_{-b/2}^{b/2}\int_{-c/2}^{c/2}\left(x_2^2 + x_3^2\right)\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = \frac{1}{2}M(b^2 + c^2)$$

where $M = \rho abc$, and from symmetry considerations, we find

$$I_{22} = \frac{1}{12}M(a^2 + c^2) \quad \text{and} \quad I_{33} = \frac{1}{12}M(a^2 + b^2).$$

Similarly, from Equation 12.55, we have

$$I_{12} = \rho\int_{-a/2}^{a/2}\int_{-b/2}^{b/2}\int_{-c/2}^{c/2}(-x_1x_2)\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = 0$$

$$I_{13} = \rho\int_{-a/2}^{a/2}\int_{-b/2}^{b/2}\int_{-c/2}^{c/2}(-x_1x_3)\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = 0$$

$$I_{23} = \rho\int_{-a/2}^{a/2}\int_{-b/2}^{b/2}\int_{-c/2}^{c/2}(-x_2x_3)\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = 0.$$

We next calculate the inertia tensor referred to as the second coordinate system with the origin at one corner. Instead of direct computation, we apply the parallel-axis theorem. In keeping with the notation of this example, we rewrite Equation 12.59 as

$$J_{ij} = I_{ij} + \sum_\alpha m_\alpha\left[\delta_{ij}\sum_k s_k^2 - s_is_j\right]$$

where

$$s_1 = a/2,\ s_2 = b/2,\ s_3 = c/2,\ s^2 = (a^2 + b^2 + c^2)/4.$$

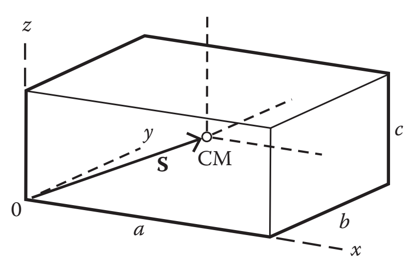


FIGURE 12.16 Homogeneous rectangular parallelepiped.

From this equation, we have

$$J_{11} = I_{11} + M(s^2 - s_1^2) = \frac{1}{2}M(b^2 + c^2),\ J_{22} = \frac{1}{2}M(a^2 + c^2),\ J_{33} = \frac{1}{2}M(b^2 + c^2)$$

$$J_{12} = I_{12} + M(-s_1 s_2) = -\frac{1}{4}Mab,\ J_{13} = -\frac{1}{4}Mac,\ J_{23} = -\frac{1}{4}Mbc.$$

12.5.4 Moments of Inertia about an Arbitrary Axis

Equation 12.25 defines moments and products of inertia about the coordinate axes, I_{ij}. Now, if the rigid body is rotating about an arbitrary axis $\hat{\lambda}$, Equation 12.25 cannot be used to compute the moment of inertia about $\hat{\lambda}$. The calculations are usually still quite straightforward if the I_{ij} are known.

Let α, β, and γ be the direction cosines of $\hat{\lambda}$, which passes through the origin of the fixed axes xyz, and let r_j be the position vector of a mass point m_j and d_j the perpendicular distance from m_j to the line $\hat{\lambda}$ as shown in Figure 12.17. The moment of inertia about the line $\hat{\lambda}$ is

$$I_\lambda = \sum_j m_j d_j^2. \tag{12.60}$$

Now, x_j/r_j, y_j/r_j, and z_j/r_j are the direction cosines of the line OM; then the cosine of the angle θ between the two lines $\hat{\lambda}$ and OM is given by

$$\cos\theta = \alpha\frac{x_j}{r_j} + \beta\frac{y_j}{r_j} + \gamma\frac{z_j}{r_j}$$

so

$$ON = OM\cos\theta = \alpha x_j + \beta y_j + \gamma z_j.$$

We also have

$$(OM)^2 = x_j^2 + y_j^2 + z_j^2,$$

and hence,

$$d_j^2 = (OM)^2 - (ON)^2 = x_j^2 + y_j^2 + z_j^2 + (\alpha x_j + \beta y_j + \gamma z_j)^2.$$

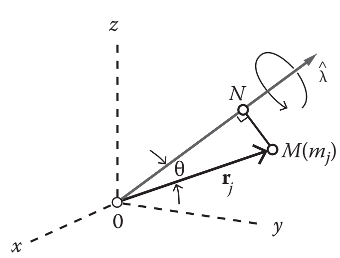


FIGURE 12.17 Moment of inertia about an arbitrary axis.

Grouping coefficients of α^2, β^2, γ^2, and so forth, and using the definitions of moments and products of inertia, we obtain

$$I_\lambda = \alpha^2 I_{xx} + \beta^2 I_{yy} + \gamma^2 I_{zz} - 2\alpha\beta I_{xy} - 2\alpha\gamma I_{xz} - 2\beta\gamma I_{yz}. \tag{12.61}$$

From Equation 12.61, we can calculate the moment of inertia about any axis through O if the moments and products of inertia with respect to the coordinate axes are known. This, in turn, can be related, with the aid of the parallel-axis theorem, to the moment of inertia about a parallel axis passing through the CM.

Equation 12.61 has a geometrical interpretation. As we allow α, β, and γ to vary, we find moments of inertia about all lines through the origin O. Let us write for one of these lines $\alpha = x/R$, $\beta = y/R$, $\gamma = z/R$, and substitute them into Equation 12.61. Here, x, y, and z are the coordinates of a point P on the line, and R is P's distance from the origin. Substitution yields

$$x^2 I_{xx} + y^2 I_{yy} + z^2 I_{zz} - 2xyI_{xy} - 2xzI_{xz} - 2yzI_{yz} = R^2 I. \tag{12.62}$$

This equation represents a quadratic surface provided we set $R^2I = 1$. For a given set of moments and products of inertia I_{xx} and so forth, we can construct a quadratic surface with its center at the origin such that the distance from O to the point P is $R = 1/(I_\lambda)^{1/2}$. I_λ cannot be zero and none of the values of R can be infinite; hence, the surface must be a closed surface, that is, an ellipsoid. This ellipsoid is referred to as Poinsot's ellipsoid of the inertia of the body at point O. It is fixed in the body with its center at the origin O. In general, there is a different one for each point of the body.

For some purposes, it is convenient to express the moment of inertia about an axis in terms of the product of the mass of the body and the square of a quantity having the dimension of length

$$I_\lambda = Mk_\lambda^2$$

or

$$k_\lambda = \sqrt{I_\lambda/M}. \tag{12.63}$$

The quantity k_λ is referred to as the radius of gyration of the body about the axis λ.

Example 12.7

Find the radius of gyration of a homogeneous rod of constant linear density ρ and of length b about an axis perpendicular to the rod and through one end.

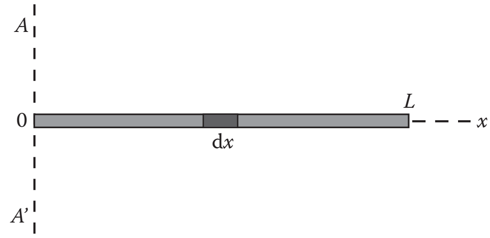


FIGURE 12.18 Homogeneous rod.

Solution:

In Figure 12.18, *OL* represents the rod, and the mass element d*m* is d*m* = ρd*x*. Then,

$$I = \int_0^b x^2\,\mathrm{d}m = \int_0^b \rho x^2\,\mathrm{d}x = mb^2/3. \tag{12.64}$$

The radius of gyration follows immediately:

$$k = \sqrt{I/m} = b/\sqrt{3}. \tag{12.65}$$

Let us make one observation about the result embodied in Equation 12.64. The rotation of the rod about axis *AOA′* can be studied by replacing the rod by a particle of mass *m*/3 whose distance from the axis is *b*, or by a particle of mass *m* whose distance from the axis is $b/\sqrt{3}$. Ordinarily, greater significance is attributed to the latter type of replacement, and the distance $b/\sqrt{3}$ is the radius of gyration of the rod with respect to the axis of rotation. We can immediately generalize this by setting, in all cases,

$$k = \sqrt{I/M}$$

as the radius of gyration of the rigid body of mass *M*, whose moment of inertia about the axis of rotation in question is *I*.

12.5.5 Principal Axes of Inertia

The expressions for the rotational kinetic energy and the angular momentum will be considerably simplified if we employ a coordinate system so that the products of inertia all vanish; the matrix representing the inertia tensor would, therefore, be diagonal

$$I_{ij} = I_i\delta_{ij} \tag{12.66}$$

where we have employed the following notation:

$$I_1 = I_{11},\ I_2 = I_{22},\ I_{33} = I_{33}. \tag{12.67}$$

The axes of this coordinate system are called the principal axes for the body at the point *O* (which is the origin of the coordinate system in question); and the three moments of inertia I_1, I_2, and I_3 are known as the principal moments of inertia of the body. We will see later that when a body is rotating about a principal axis through its CM there is no resultant force or couple on the axis.

For principal axes, the general formulas for angular momentum, rotational kinetic energy, and moment of inertia, Equations 12.30, 12.24, and 12.61, respectively, reduce to

$$T_{rot} = \frac{1}{2}\left(I_1\omega_1^2 + I_2\omega_2^2 + I_3\omega_3^2\right) \tag{12.68}$$

$$\vec{L} = \hat{e}_1\omega_1 + \hat{e}_2\omega_2 + \hat{e}_3\omega_3 \tag{12.69}$$

$$I = \alpha^2 I_1 + \beta^2 I_2 + \gamma^2 I_3 \tag{12.70}$$

where $\hat{e}_i$, i = 1, 2, 3, are the three unit vectors along the principal axes. Because T_{rot} cannot be negative, we see that the diagonal matrix representing the moment of inertia *I* is positive definite with $I_i > 0$. The equation for rotational motion can be compared with the corresponding one for the translational motion with velocity $\vec{v}$; $T_{rot} = mv^2/2 = \vec{p}\cdot\vec{v}/2$.

The principal difference is that the mass of the body has no directional properties, so that the coefficients of v_1^2, v_2^2, and v_3^2 are all equal.

Alternatively, from Equations 12.68 through 12.70, we have

$$T_{rot} = \frac{L_1^2}{2I_1} + \frac{L_2^2}{2I_2} + \frac{L_3^2}{2I_3}. \tag{12.71}$$

The principal axes system does, in fact, exist for any rigid body and for any point taken as the origin. The problem of finding the directions of the principal axes in a given situation is one of rotating the coordinate system so as to eliminate the products of inertia. Suppose that this has been done, and the body is rotating about one of its principal axes, say, the z-axis. Selecting the principal axes as the coordinate ones, we have $\omega_x = \omega_y = 0$, $\omega_z = \omega$. Because the inertia tensor has the form of Equation 12.66, when the coordinate axes are chosen in this way, Equation 12.30 gives the following values of the components of the angular momentum of a body:

$$L_x = L_y = 0,\ L_z = I_z\omega.$$

Consequently, the vector $\vec{L}$ has the same direction as $\vec{\omega}$. The same result is obtained for rotation of a body about other principal axes. In all these cases, we arrive at the relationship $\vec{L} = I\vec{\omega}$, where I is the corresponding principal moment of inertia of the body.

Now, assume that we have arbitrarily selected a set of xyz-axes and have computed the moments and products of inertia with respect to them. Then, in order that angular momentum $\vec{L}$ and angular velocity $\vec{\omega}$ have the same direction, we must have, from Equation 12.30,

$$\begin{aligned} L_1 &= I_{11}\omega_1 + I_{12}\omega_2 + I_{13}\omega_3 = I\omega_1 \\ L_2 &= I_{21}\omega_1 + I_{22}\omega_2 + I_{23}\omega_3 = I\omega_2 \\ L_3 &= I_{31}\omega_1 + I_{32}\omega_2 + I_{33}\omega_3 = I\omega_1 \end{aligned}$$

or

$$\begin{aligned} (I_{11} - I)\omega_1 + I_{12}\omega_2 + I_{13}\omega_3 &= 0 \\ I_{21}\omega_1 + (I_{22} - I)\omega_2 + I_{23}\omega_3 &= 0 \\ I_{31}\omega_1 + I_{32}\omega_2 + (I_{33} - I)\omega_3 &= 0 \end{aligned} \tag{12.72}$$

If these equations have a nontrivial solution for the ratios $\omega_1{:}\omega_2{:}\omega_3$, then the determinant of the coefficients vanishes:

$$\begin{vmatrix} I_{11} - I & I_{12} & I_{13} \\ I_{21} & I_{22} - I & I_{23} \\ I_{31} & I_{32} & I_{33} - I \end{vmatrix} = 0. \tag{12.73}$$

The expansion of this determinant leads to a cubic secular equation in I, whose three roots we shall denote by I_1, I_2, and I_3. These are the principal moments of inertia.

The direction of the principal axis can now be found from Equation 12.72. The components ω_1, ω_2, and ω_3 define a direction of the angular velocity with respect to the coordinate axes. Because, for rotation about the principal axis, the angular velocity coincides with this axis, the set of numbers ω_1, ω_2, and ω_3 that satisfies Equation 12.72 comprises the direction numbers of the axis. For example, if the body rotates about an axis corresponding to the principal moment I_1, we can determine the direction of this principal axis by substituting I_1 for I in Equation 12.72:

$$(I_{11}-I_1)\omega_{11}+I_{12}\omega_{21}+I_{13}\omega_{31}=0$$
$$I_{21}\omega_{11}+(I_{22}-I_1)\omega_{21}+I_{23}\omega_{31}=0$$
$$I_{31}\omega_{11}+I_{32}\omega_{21}+(I_{33}-I_1)\omega_{31}=0$$

where the second subscript 1 on the ω_{i1} indicates that we are considering the principal axis associated with I_1. These equations can be solved for the ratios ω_1:ω_2:ω_3, which are the direction numbers of the principal axis corresponding to I_1. This principal axis is thereby defined relative to the original coordinate system. The directions corresponding to I_2 and I_3 can be found in a similar way.

Example 12.8: Principal Axes for a Uniform Cube

In Example 12.5, we calculated the moments and products of inertia for a uniform parallelepiped in coordinate axes that lie along the three edges of the parallelepiped with the origin at one corner. For simplicity, let us consider a uniform cube of edge a. Letting $b = c = a$ in Example 12.5, we found the moments and products of inertia for a uniform cube:

$$I_{11} = I_{22} = I_{33} = 2Ma^2/3 \text{ and } I_{12} = I_{13} = I_{23} = -Ma^2/4.$$

Evidently, the coordinate axes chosen for the calculation are not principal axes. In order to find the principal axes, we first solve the following determinant:

$$\begin{vmatrix} 2Ma^2/3-I & -Ma^2/4 & -Ma^2/4 \\ -Ma^2/4 & 2Ma^2/3-I & -Ma^2/4 \\ -Ma^2/4 & -Ma^2/4 & 2Ma^2/3-I \end{vmatrix} = 0.$$

Subtracting the first row from the second, we obtain

$$\begin{vmatrix} 2Ma^2/3-I & -Ma^2/4 & -Ma^2/4 \\ -11Ma^2/12+I & 11Ma^2/12-I & 0 \\ -Ma^2/4 & -Ma^2/4 & 2Ma^2/3-I \end{vmatrix} = 0.$$

Factoring out $(11Ma^2/12 - I)$ from the second row and then expanding the determinant, we obtain

$$\left[\frac{11}{12}Ma^2-1\right]\left[\left(\frac{2}{3}Ma^2-1\right)^2-\frac{1}{8}(Ma^2)^2-\frac{1}{4}Ma^2\left(\frac{2}{3}Ma^2-1\right)\right]=0$$

from which we obtain the following principal moments of inertia for the cube

$$I_1 = I_2 = 11Ma^2/12,\ I_3 = Ma^2/6.$$

Because two of the principal moments of inertia are equal ($I_1 = I_2$), the principal axis associated with I_3 must be an axis of symmetry. Let us find the direction of this principal axis. For this purpose, we write Equation 12.72 in the form

$$\left(\frac{2Ma^2}{3} - \frac{Ma^2}{6}\right)\omega_{13} = \frac{Ma^2}{4}\omega_{23} - \frac{Ma^2}{4}\omega_{33} = 0$$

$$-\frac{Ma^2}{4}\omega_{13} + \left(\frac{2Ma^2}{3} - \frac{Ma^2}{6}\right)\omega_{23} - \frac{Ma^2}{4}\omega_{33} = 0$$

$$-\frac{Ma^2}{4}\omega_{13} - \frac{Ma^2}{4}\omega_{23} + \left(\frac{2Ma^2}{3} - \frac{Ma^2}{6}\right)\omega_{33} = 0$$

where the second subscript 3 on ω_{i3} indicates that we are considering the principal axis associated with I_3. These equations can be simplified to

$$2\frac{\omega_{13}}{\omega_{33}} - \frac{\omega_{23}}{\omega_{33}} = 1, \quad -\frac{\omega_{13}}{\omega_{33}} + 2\frac{\omega_{23}}{\omega_{33}} = 1, \quad \frac{\omega_{13}}{\omega_{33}} + \frac{\omega_{23}}{\omega_{33}} = 1.$$

Subtracting the second equation from the first, we obtain $\omega_{13} = \omega_{23}$. Using this result in the first equation, we find $\omega_{13} = \omega_{33}$, and the direction ratios

$$\omega_1:\omega_2:\omega_3 = 1:1:1.$$

Thus, when the cube is rotating about this principal axis (the principal axis associated with I_3), the projections of ω on the three coordinate axes are all equal. Therefore, the principal axis associated with I_3 coincides with the main diagonal of the cube. On the other hand, the orientation of the principal axes associated with I_1 and I_2 is arbitrary, so they need only lie in a plane normal to the diagonal of the cube.

Determination of the principal axes of inertia is greatly simplified if the body is symmetrical for it is clear that the position of the body's CM and the directions of the principal axes must have the same symmetry as the body. For example, consider a body possessing a plane of symmetry. Calling this the *xy*-plane, the *z*-axis is a principal axis, and then the contribution to the products of inertia I_{xz} and I_{yz} from any element mass m_i at (x, y, z) is exactly canceled by that from the element mass m_j at $(x, y, -z)$. The other two of the principal axes of inertia lie in a plane that can be determined as follows:

Suppose the body is rotating about one of these two principal axes. The angular momentum and the angular velocity then coincide, and we can write

$$\vec{L} = I\vec{\omega}. \tag{12.74}$$

Then, from Equation 12.30, we have

$$I_{11}\omega_1 + I_{12}\omega_2 = I\omega_1,\ I_{12}\omega_1 + I_{22}\omega_2 = I\omega_2. \tag{12.75}$$

If θ is the angle between the *x*-axis and the particular principal axis about which the body is rotating, then

$$\omega_2/\omega_1 = \tan\theta$$

and Equation 12.75 becomes, upon dividing by ω_1,

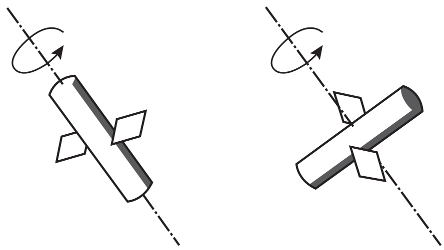


FIGURE 12.19 Rotating cylindrical satellite.

$$I_{11} + I_{12} \tan\theta = I,\ I_{12} + I_{22} \tan\theta = I \tan\theta.$$

Eliminating I from these two equations, we find

$$(I_{22} - I_{11})\tan\theta = I_{12}(\tan^2\theta - 1)$$

from which we obtain

$$\tan 2\theta = \frac{2I_{12}}{I_{11} - I_{22}}$$

where we have used the identity $\tan 2\theta = 2\tan\theta/(1 - \tan^2\theta)$.

In the interval 0° to 180°, there are two values of θ, differing by 90°, that satisfy Equation 12.76. These two values of θ give the directions of the two principal axes in the *xy*-plane.

A body whose three principal moments of inertia are all different is called an asymmetrical top. If two are equal, it is termed a symmetrical top. In this case, the direction of one of the principal axes in the *jy*-plane can be chosen arbitrarily. If all three principal moments of inertia are equal, the body is called a spherical top, and the three principal axes of inertia can be chosen arbitrarily as any three mutually perpendicular axes. If two of the principal moments of inertia are equal and the third is zero, the body is called a rotator (or rotor), or as an example, two point masses connected by a weightless shaft or a diatomic molecule.

Example 12.9: Stability of a Rotating Cylindrical Satellite

One of the early space satellites was cylindrical in shape and was put into orbit spinning around its long axis. Even though the satellite was torque-free, it began to wobble as soon as it was put into orbit, ending by spinning around the transverse axis (Figure 12.19). Find an explanation for this.

The reason is that although $\vec{L}$ is conserved for torque-free motion, the kinetic energy of rotation can change if the body is not absolutely rigid. Now, if the satellite is rotating slightly off the symmetry axis, each part of the body experiences a time-varying centripetal force; the body warps and bends, and energy is dissipated by internal friction in the structure. So the rotational kinetic energy of the satellite must decrease. For a body rotating about a single principal axis, Equation 12.71 gives $T_{rot} = L^2/2I$; T_{rot} is the minimum for the axis with the greatest momentum of inertia, and the motion is stable around that axis. For the cylindrical satellite, the initial axis of rotation (along its long axis) had a minimum momentum of inertia. T_{rot} was the maximum, and the motion was therefore not stable.

12.6 EULER'S EQUATIONS OF MOTION

The preceding sections have dealt with the inertial properties of a rigid body. We come now to what we might call the essential physics of the mechanics of rigid bodies, namely, the dynamics of its rotational motion about a fixed point under the action of external forces. In Section 12.1, we saw that the rotation of a rigid body is governed by the rotational analogue of Newton's second law: the rate of change of total angular momentum is equal to the applied torque:

$$\vec{N}^{(9e)} = \frac{\mathrm{d}\vec{L}}{\mathrm{d}t} \tag{12.76}$$

in which $\vec{N}^{(9e)}$ is the net external torque, and $\vec{L}$ is the angular momentum. If the rigid body also has a translational motion, Newton's second law governs the motion of its CM:

$$\vec{F}^{(e)} = \frac{\mathrm{d}\vec{P}}{\mathrm{d}t} \tag{12.77}$$

where the derivatives must be evaluated in an inertial frame. These two sets of equations of motion are equivalent to Lagrange's equations for the coordinates of the CM.

Because the angular momentum $\vec{L}$ involves the inertial parameters, a computation of $\mathrm{d}\vec{L}/\mathrm{d}t$ involves the time rate of change of the inertial parameters as well as the rate of change of $\vec{\omega}$. However, if we choose the principal axes of the body as the reference coordinate system, the moments of inertia with respect to these axes are fixed, which simplifies the computation of $\mathrm{d}\vec{L}/\mathrm{d}t$. Thus, in general, we employ a coordinate system that is fixed in the body and rotates with it. This body system is not an inertial one. Let us now rewrite Equation 12.76 with respect to the body axes. The time rate of change of the angular momentum vector in a fixed (inertial) system versus a rotating system is given by Equation 12.19:

$$\frac{\mathrm{d}\vec{L}}{\mathrm{d}t} = \frac{\partial\vec{L}}{\partial t} + \vec{\omega}\times\vec{L}, \tag{12.78}$$

where $\partial/\partial t$ is the derivative in the rotating (body) frame. Thus, the equation of motion in the body system is

$$\vec{N} = \frac{\partial\vec{L}}{\partial t} + \vec{\omega}\times\vec{L} \tag{12.79}$$

or

$$\frac{\partial\vec{L}}{\partial t} = -\vec{\omega}\times\vec{L} + \vec{N}. \tag{12.80}$$

It is seen here that the term $-\vec{\omega}\times\vec{L}$ on the right-hand side of Equation 12.80 appears as fictitious torque arising because of the non-inertial nature of the body axes.

Because I is constant relative to the body axes, we can now substitute Equation 12.30 to obtain, with the help of the fact that $\partial\vec{\omega}/\partial t = \mathrm{d}\vec{\omega}/\mathrm{d}t$,

$$I\frac{\mathrm{d}\vec{\omega}}{\mathrm{d}t} + \vec{\omega}\times(I\vec{\omega}) = \vec{N}. \tag{12.81}$$

A considerable simplification of these equations occurs if we choose as body axes the principal axes of the body; then, by Equation 12.69, Equation 12.81 becomes

$$\begin{aligned} I_1\dot{\omega}_1 &= \omega_2\omega_3(I_2 - I_3) + N_1 \\ I_2\dot{\omega}_2 &= \omega_3\omega_1(I_3 - I_1) + N_2 \\ I_3\dot{\omega}_3 &= \omega_1\omega_2(I_1 - I_2) + N_3. \end{aligned} \tag{12.82}$$

These are Euler's equations of motion for a rigid body in a force field. There is a restriction on their use: if one point in the body is held fixed, that point is to be taken as the origin for the body axes, and the moments of inertia I_1, I_2, I_3 and the torques N_1, N_2, N_3 are relative to that point. If the body is unconstrained, the CM of the body is to be taken as the origin for the body axes. Here ω_1, ω_2, ω_3 are not angular velocity components in the ordinary sense of being derivatives of some spatial coordinates that describe the position of the body at time t. One more point should also be noted: as the principal axes are fixed in the body, any externally applied force must be resolved along these axes in the direction the body is moving.

Euler's equations are nonlinear equations and are, therefore, difficult to solve in the general case. In the absence of external torques, Euler's equations reduce to simpler forms:

$$\begin{aligned} I_1\dot{\omega}_1 &= \omega_2\omega_3(I_2 - I_3) \\ I_2\dot{\omega}_2 &= \omega_3\omega_1(I_3 - I_1) \\ I_3\dot{\omega}_3 &= \omega_1\omega_2(I_1 - I_2). \end{aligned} \tag{12.83}$$

As an example, the external torques acting on the Earth are so weak that the rotational motion can be considered to a first approximation as relatively torque-free.

Equation 12.83 can be solved, in principle, in terms of elliptical integrals. The solution can then be written in terms of the constants of the motion, angular momentum, and kinetic energy. But this is too complicated to be considered here. Instead, in the following sections, we shall consider some illuminating cases where axial symmetry allows direct integration of Equation 12.83.

In general, the angular momentum is not proportional and parallel to the angular velocity. Thus, even though the angular momentum may be a constant of the motion, the angular velocity is not. In general, the angular velocity vector will precess around the angular momentum vector, and the angle between them varies in time. This is known as nutation.

12.7 MOTION OF A TORQUE-FREE SYMMETRICAL TOP

As an application of Euler's equations of motion, we consider the motion of a torque-free symmetrical top: the motion of a heavy symmetrical top spinning freely about its axis of symmetry under the influence of a torque produced by its own weight.

A symmetrical top in a rigid body possesses an axis of symmetry, so that two of its three principal moments of inertia are equal. We assume that the axis of symmetry is the z-axis and designate by I_3 the moment of inertia about the symmetry axis; $I_1 = I_2 = I$ is the moment of inertia about the x- or y-axis. For force-free motion, the body's CM is either at rest or in uniform motion relative to the fixed, external inertial system. We can, therefore, place the origin of the coordinate system at the CM of the body. Equation 12.83 now becomes

$$\begin{aligned} I_1\dot{\omega}_1 &= \omega_2\omega_3(I_2 - I_3) \\ I_2\dot{\omega}_2 &= \omega_3\omega_1(I_3 - I_1) \\ I_3\dot{\omega}_3 &= 0. \end{aligned} \tag{12.84}$$

It should be noted that $\vec{\omega}$ does not, in general, coincide with the axis of mass symmetry because otherwise, the problem is trivial; the motion is a steady spin about the axis of symmetry.

The third equation of Equation 12.84 tells us that ω_3 is a constant in time:

$$\omega_3(t) = \text{constant}.$$

Defining a constant k by

$$k = \frac{\omega_3(I_3 - I)}{I} \tag{12.85}$$

the first two equations of Equation 12.84 can be written in the form

$$\dot{\omega}_1 = -k\omega_2 \text{ and } \dot{\omega}_2 = k\omega_1. \tag{12.86}$$

Differentiating the first and substituting the second, we obtain

$$\ddot{\omega}_1 = -k^2\omega_1. \tag{12.87}$$

Because k^2 is a positive quantity, Equation 12.87 is just the equation of harmonic motion; its solution is

$$\omega_1(t) = A\cos(kt + \phi_0) \tag{12.88}$$

where A, a constant, is the amplitude of ω_1 and φ_0 is a phase constant. Also from Equation 12.86:

$$\omega_2(t) = -\dot{\omega}/k = A\sin(kt + \phi_0). \tag{12.89}$$

Equations 12.88 and 12.89 are the parametric equations for a circle of radius A in the plane of the xy-axes. The projection of $\vec{\omega}$ vector onto the xy-plane rotates about the z-axis with a constant angular frequency k. It is clear that

$$\sqrt{\omega_1^2 + \omega_2^2 + \omega_3^2} = \sqrt{A^2 + \omega_3^2}$$

is constant during the motion. Figure 12.20 illustrates the geometrical relationship of ω_1 to ω_2 at a particular instant. Equations 12.88 and 12.89 show that, if the phase constant φ_0 is chosen as zero, at $t = 0$, $\omega_1 = A$ and $\omega_2 = 0$; as t increases, ω_1 decreases while ω_2 increases. This shows that the axis of rotation precesses and describes a right circular cone (called the body cone) about the z-axis (the axis of symmetry) counterclockwise with angular velocity k. The frequency f_p and the period T_p of the precession are given by

$$f_p = \frac{k}{2}, T_p = \frac{2\pi}{k} = \frac{2\pi I}{(I_3 - I)\omega_3}. \tag{12.90}$$

The rotation of the Earth can be considered, to a first approximation, as the rigid body motion of the type just discussed. Its axis of symmetry departs slightly from its axis of rotation. The angular velocity of the earth is 7.29×10^{-5} rad/s. The quantity $(I_3 - I)/I$ was determined astronomically to be 0.00329. Substituting into Equation 12.90, we find that $T_p = 305$ days. This periodic motion of $\vec{\omega}$ about the Earth's axis of symmetry is known observationally as the variation of latitude. It is very difficult to measure T_p because the diameter of the circle formed by the intersection of the body cone with the Earth's surface is only a few feet. The most recent observed period of precession of the Earth's axis of rotation about the pole is about 433 days. The difference from the value just calculated is attributed to the fact that the Earth is not a perfectly rigid body.

It is clear that in this case the angular momentum vector of axial symmetry becomes

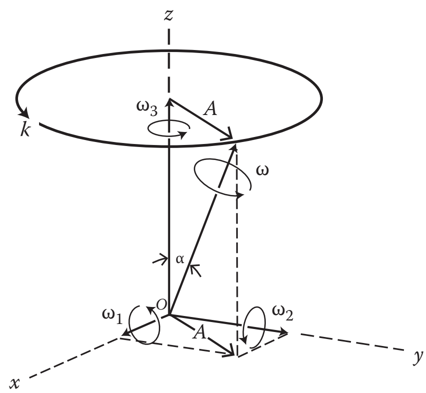


FIGURE 12.20 Angular velocity in free rotation.

$$\vec{L} = I\left(\omega_1\hat{i} + \omega_2\hat{j}\right) + I_3\omega_3\hat{k}. \tag{12.91}$$

Hence, it is in the plane of ω_3 and $\vec{\omega}$. Its position relative to $\vec{\omega}$ depends on the values of I and I_3. In Figure 12.20, we denote the angle between ω_3 and $\vec{\omega}$ by α, and that between ω_3 and $\vec{L}$ by θ, and remembering that the projection of $\vec{\omega}$ is $\vec{A}$ as shown in Figure 12.19, we get

$$\tan\theta = \frac{\vec{L}\cdot\hat{A}}{L_3} = \frac{LA}{I_3\omega_3} = \frac{I}{I_3}\tan\alpha \tag{12.92}$$

where $\hat{A} = \vec{A}/A$, a unit vector in the direction of $\vec{A}$, and L_3 is the component of $\vec{L}$ along ω_3. It is seen that $\theta < \alpha$, if $I < I_3$, and $\vec{L}$ lies between ω_3 and $\vec{\omega}$ (Figure 12.21a). On the other hand, if $I > I_3$, then $\theta > \alpha$, and $\vec{L}$ lies farther from ω_3 than does $\vec{\omega}$ (Figure 12.21b).

So far, we are discussing a processional motion of $\vec{\omega}$ about an axis fixed in the body. In order to describe the motion relative to an observer outside the body, we use an external axis fixed in space. In such an external inertial frame, we have

$$\left.\frac{d\vec{L}}{dt}\right|_{fix} = \vec{N}$$

where the subscript "fix" indicates that this rotational analogue of Newton's second law is only valid in the external system. Thus, for torque-free motion, $\vec{L}$ is stationary in the external frame and is

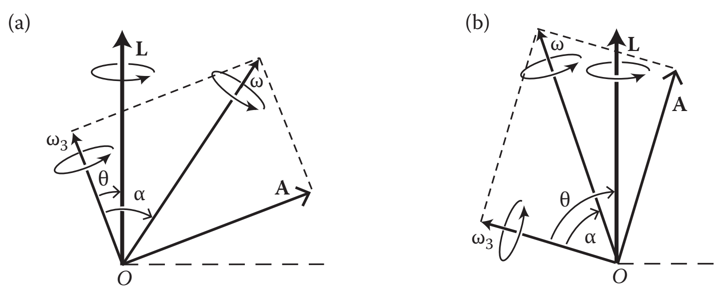


FIGURE 12.21 (a) Oblate spheroid. (b) Prolate spheroid.

constant in time. The line defined by this constant vector $\vec{L}$ is called the invariable line. The kinetic energy T_{rot} is a second constant of the motion for the torque-free motion. Multiply the first of the Euler Equations 12.82 by ω_1, the second by ω_2, and the third by ω_3, and add, we get

$$\frac{d}{dt}\left[\frac{1}{2}\left(I_1\omega_1^2 + I_2\omega_2^2 + I_3\omega_3^2\right)\right] = N_1\omega_1 + N_2\omega_2 + N_3\omega_3. \tag{12.93}$$

The left-hand side is the time rate of change of the kinetic energy; the right-hand side is the rate at which the applied torque works. Thus, in the absence of external torque, the kinetic energy is constant.

The rotational kinetic energy T_{rot} can be expressed as $\vec{L}\cdot\vec{\omega}/2$. Because T_{rot} and $\vec{L}$ are constants, this means that during the motion, $\vec{\omega}$ must change in such a way that its projection on the stationary $\vec{L}$ is constant. Thus, $\vec{\omega}$ precesses around and makes a constant angle with the angular momentum vector $\vec{L}$ (space cone). Motion of the type discussed here may be represented geometrically by the rolling of one cone on another as depicted in Figure 12.22a and b. The cone with its axis along $\vec{L}$ and semi-vertical angle θ is fixed in space. This cone is often called the space cone. The second cone with its axis along ω_3 and the semi-vertical angle α is fixed in the body and is called the body cone. To an observer in the body system, $\vec{\omega}$ would appear to trace out the body cone, and to an observer in an external initial system, $\vec{\omega}$ would appear to trace out the space cone. In either event, ω_3, $\vec{\omega}$, and $\vec{L}$ remain co-planar as one cone rolls on the other. The rate of precession of $\vec{\omega}$ about the invariable line (direction of constant momentum vector $\vec{L}$) can be calculated as follows.

In Figure 12.23, the $Ox'y'z'$ system of axes has its orientation fixed in space, and we choose the z'-axis to be the direction of $\vec{L}$. The $Oxyz$ system is the principal body axis, where the axis of symmetry of the body is the z-axis. Line ON is the line of node (the intersection of the xy-plane with the $x'y'$-plane). And θ, φ, and ψ are the Eulerian angles. Construct line OM perpendicular to ON. Then, angle MOy equals ψ in magnitude. It is clear that ON, OM, and the z-axis form a third system of axes ξ, η, ζ, with the ζ-axis coinciding with the z-axis. The ξ-, η-, and ζ-axes are also principal axes. Furthermore, the principal moments are the same: $I_1 = I_2 = I_{\xi\xi} = I_{\eta\eta} = I$ and $I_3 = I_{\zeta\zeta}$. Now consider the rotation of the η,ζ-plane. We see that the angular speed of rotation of the η,ζ-plane about the z'-axis is equal to the time rate of the angle φ. Thus, $\dot{\phi}$ is the angular rate of precession of the axis of rotation and also of the symmetry axis about the invariable line (direction of constant angular momentum vector) as viewed from outside the body. This precession appears as a wobble like that seen in an imperfectly thrown football. To compute $\dot{\phi}$, we first express the components of $\vec{L}$ and $\vec{\omega}$ in the $O\xi\eta\zeta$-axes. Because $\dot{\phi}$ (directed along the z-axis) and $\dot{\psi}$ (directed along the z'-axis) are both perpendicular to the ξ-axis, they will not contribute to ω_ξ. But θ lies along ON, so its contribution to ω_ξ is 6. Next, because $\dot{\phi}$ lies along the z-axis, its contribution to ω_η and ω_ξ are $\dot{\phi}\sin\theta$ and $\dot{\phi}\cos\theta$, respectively.

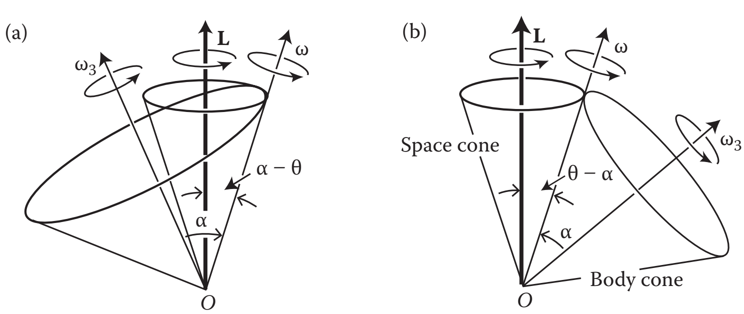


FIGURE 12.22 (a) Space and body cone, $I_3 > I$. (b) Space and body cone, $I_3 < I$.

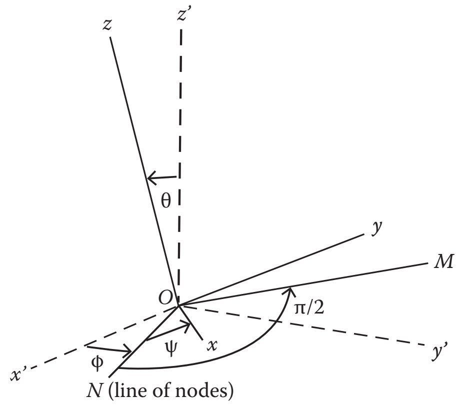


FIGURE 12.23 System *Ox′y′z′* is an external set of axes; system *Oxyz* is a body set of axes.

Similarly, because $\dot{\psi}$ lies along the *z*- (or ζ-) axis, its contribution to ω_ζ is $\dot{\psi}$, but it will have no contributions to ω_ξ and ω_η. Adding these together, we obtain

$$\omega_\xi = 0, \quad \omega_\eta = \dot{\phi}\sin\theta, \quad \omega_\zeta = \dot{\phi}\cos\theta + \dot{\psi}\,. \tag{12.94}$$

Similarly,

$$L_\xi = 0, L_\eta = L\sin\theta, L_\zeta = L\cos\theta. \tag{12.95}$$

It was shown in Equation 12.91 that $\vec{L}$, $\vec{\omega}$, and ω_3 lie in the same plane. But ω_3 lies along the *z*- (or ζ-) axis. So $\vec{L}$, $\vec{\omega}$, and ω_3 lie in the ηζ-plane. Again, letting α denote the angle between the angular velocity vector $\vec{\omega}$ and ω_3 (the *z*-axis), we have the following relationships:

$$\omega_\eta = \omega\sin\alpha, \omega_\zeta = \omega\cos\alpha, L_\eta = I\,\omega\sin\alpha, L_\zeta = I\,\omega\cos\alpha \tag{12.96}$$

from which it follows that

$$\frac{L_\eta}{L_\zeta} = \tan\theta = \frac{I}{I_3}\tan\alpha. \tag{12.97}$$

Comparing Figures 12.21 and 12.23, we see that angle θ in Figure 12.21 corresponds to the Eulerian angle θ in Figure 12.23. We also note that Equation 12.97 is identical to Equation 12.92.

From the second of Equation 12.94 and the first of Equation 12.96, we have

$$\dot{\phi}\sin\theta = \omega\sin\alpha \quad \text{or} \quad \dot{\phi} = (\omega\sin\alpha)/\sin\theta \tag{12.98}$$

for the rate of precession. This equation can be put into a more useful form. To this purpose, let us first write Equation 12.97 as

$$\frac{\sin\theta}{\sqrt{1-\sin^2\theta}} = \frac{I}{I_3}\frac{\sin\alpha}{\sqrt{1-\sin^2\alpha}}$$

from which we obtain, after a little algebra,

$$\frac{\sin\alpha}{\sin\theta}=\left[1+\left(\frac{I_3^2}{I^2}-1\right)\cos^2\alpha\right]^{1/2}$$

and Equation 12.98 now becomes

$$\dot{\phi}=\omega\left[1+\left(\frac{I_3^2}{I^2}-1\right)\cos^2\alpha\right]^{1/2}. \tag{12.99}$$

The free precession of the Earth provides a good illustrative example of our above discussion. It is known that the axis of rotation is very slightly inclined with respect to the geographic pole defining the axis of symmetry. The angle is approximately 0.2 s of arc. The ratio of the moments of inertia I_3/I is about 1.00329. Then Equation 12.99 gives

$$\dot{\phi}=1.00329$$

and the associated period of the Earth's wobble is approximately

$$2\pi/\dot{\phi}=0.997 \text{ day.}$$

This free precession of the Earth's axis in space is superimposed upon a much longer gyroscopic precession of about 26,000 years, the latter resulting from a torque exerted on the Earth by the sun and moon.

Finally, we emphasize that the reader should not be confused by the three basic angular rates: (1) the magnitude of the angular velocity vector $\vec{\omega}$; (2) the precession of the angular rate k of the axis of rotation (the direction of the angular velocity vector $\vec{\omega}$) about the axis of symmetry of the body; and (3) the precession of angular rate $\dot{\phi}$ of the axis of symmetry about the invariable line (direction of constant angular momentum vector $\vec{L}$).

12.8 MOTION OF HEAVY SYMMETRICAL TOP WITH ONE POINT FIXED

In this section, we shall consider the motion of a heavy symmetrical top spinning freely about its axis of symmetry under the influence of a torque produced by its own weight. Some point O on the axis of spin is fixed. We assume this to be different from the CM of the body. A wide variety of physical systems, from a simple toy top to complicated gyroscopic navigational instruments, are approximated by such a heavy symmetrical top. The theory of spinning tops even finds applications in astronomy and nuclear physics.

The notation for our coordinate axes is shown in Figure 12.24. The symmetry axis of the body is one of the principle axes, and we choose it as the z-axis of the body coordinate system. The principal moment of inertia h must now be evaluated relative to the point O, which is at a distance h from the CM. $I_1 = I_2 = I$ remains unchanged. At any given instant, the orientation of the top is completely specified by the three Euler's angles as shown in Figure 12.24. ON is the line of nodes; θ gives the inclination of the z-axis from the vertical; φ measures the azimuth of the top about the vertical; and ψ is the rotation angle of the top about its own z-axis. The motion of the top in φ corresponds to precession, and variation with θ is known as nutation.

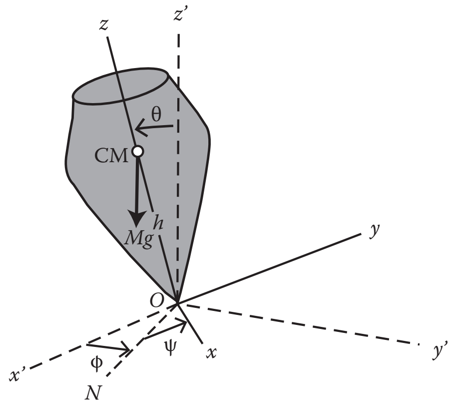


FIGURE 12.24 Heavy spinning top under gravity.

The kinetic energy T is

$$T = \frac{1}{2}\sum_{i=1}^{3} I_i \omega_i^2 = \frac{1}{2}\sum_{i=1}^{3} I\left(\omega_1^2 + \omega_2^2\right) + \frac{1}{2} I_3 \omega_3^2.$$

Using Equation 12.22, we can rewrite kinetic energy in terms of Euler's angles:

$$T = \frac{1}{2} I\left(\dot{\phi}^2 \sin^2\theta + \dot{\theta}^2\right) + \frac{1}{2} I_3 \left(\dot{\phi}\cos\theta + \dot{\psi}\right)^2. \tag{12.100}$$

The potential energy V is

$$V = Mgh\cos\theta. \tag{12.100a}$$

The Lagrangian L is

$$\begin{aligned} L &= T - V \\ &= \frac{1}{2} I\left(\dot{\phi}^2 \sin^2\theta + \dot{\theta}^2\right) + \frac{1}{2} I_3 \left(\dot{\phi}\cos\theta + \dot{\psi}\right)^2 - Mgh\cos\theta. \end{aligned} \tag{12.101}$$

It is too cumbersome to solve the three Lagrange's equations of motion for the Euler's angles. But we can get all the information by examining the constants of the motion. We first notice that both φ and ψ are cyclic coordinates in the Lagrangian L. The generalized momenta conjugate to these coordinates are therefore constants of motion:

$$p_\phi = \partial L/\partial\phi = (I\sin^2\theta + I_3\cos^2\theta)\dot{\phi} + I_3\dot{\psi}\cos\theta = Ia \tag{12.102}$$

$$p_\psi = \partial L/\partial\psi = I_3(\dot{\psi} + \dot{\phi}\cos\theta) = Ib \tag{12.103}$$

where a and b represent the two constants of motion. It is clear that p_φ is the angular momentum along the z-axis for which φ is the rotation angle, and p_ψ is the angular momentum along the z'-axis (the vertical axis) for which ψ is the rotation angle. We can also arrive at this result from torque consideration. The torque of gravity is along the line of nodes, and because both the z- and z'-axes are perpendicular to the line of nodes, there is no component of the torque along these axes: $N_\theta = Mgh \sin\theta$, $N_\varphi = 0$, and $N_\psi = 0$.

Equations 12.102 and 12.103 can be solved for $\dot{\phi}$ and $\dot{\psi}$ in terms of θ. First, we rewrite Equation 12.103 as

$$\dot{\psi} = \frac{Ib - I_3\dot{\phi}\cos\theta}{I_3}. \tag{12.104}$$

Substituting this result into Equation 12.102 to eliminate ψ:

$$I\dot{\phi}\sin^2\theta + Ib\cos\theta = Ib$$

or

$$\dot{\phi} = \frac{a - b\cos\theta}{\sin^2\theta}. \tag{12.105}$$

This furnishes the precession $\dot{\phi}$ if θ is known. Using this expression for $\dot{\phi}$ back in Equation 12.104, we obtain

$$\dot{\psi} = \frac{Ib}{I_3} - \cos\theta\frac{a - b\cos\theta}{\sin^2\theta}. \tag{12.106}$$

If θ were known as a function of time, Equations 12.105 and 12.106 could be integrated to furnish the dependence on time of φ and ψ.

There is one further constant of the motion available. If there are no frictional forces acting on the top to dissipate energy, then the total energy remains constant:

$$E = \frac{1}{2}\left(\dot{\theta}^2 + \dot{\phi}^2\sin^2\theta\right) + \frac{1}{2}I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)^2 + Mgh\cos\theta = \text{constant}. \tag{12.107}$$

Now, in view of Equation 12.103, it is not only clear that E is a constant of the motion, but that the quantity

$$\frac{1}{2}I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right)^2 + Mgh\cos\theta\mathrm{t}$$

is also a constant. Combining this with Equation 12.107) we find

$$E' = \frac{1}{2}\left(\dot{\theta}^2 + \dot{\phi}^2\sin^2\theta\right) + Mgh\cos\theta = \text{constant}. \tag{12.108}$$

Equations 12.105 and 12.106 can be used to eliminate φ and ψ from this energy equation, resulting in a differential equation involving θ alone. Then, from the value of θ, we found that Equations 12.105 and 12.106 yield angles φ and ψ, respectively, as functions of time. Substituting Equation 12.105 into Equation 12.108, we get

$$E' = \frac{1}{2}I\dot{\theta}^2 + \frac{I(a - b\cos\theta)^2}{2\sin^2\theta} + Mgh\cos\theta = \frac{1}{2}I\dot{\theta}^2 + V'(\theta) \tag{12.109}$$

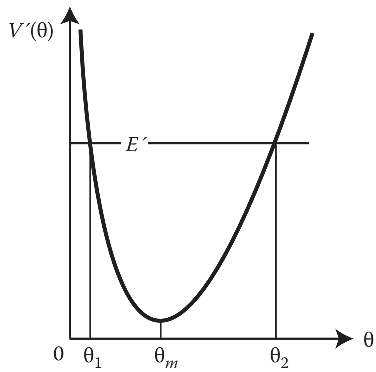


FIGURE 12.25 Energy diagram for the top.

where

$$V'(\theta) = \frac{I(a - b\cos\theta)^2}{2\sin^2\theta} + Mgh\cos\theta \tag{12.110}$$

can be considered as an "effective potential." Equation 12.109 can be solved to yield $t(\theta)$:

$$t(\theta) = \int \frac{d\theta}{\sqrt{2[E' - V'(\theta)]/I}} \tag{12.111}$$

This integral can be inverted to give $\theta(t)$. But detailed integration involves elliptic integrals. Physics will then be obscured in the profusion of mathematics. Fortunately, some qualitative features of the motion can be obtained by other means as outlined below.

If we plot $V'(\theta)$ against θ (Figure 12.25) in the physically limited range $0 \le \theta < \pi$, we have an energy diagram of the kind discussed in Chapter 6 on central force motion. For a given total and constant energy E, motion will be limited for the case $E' > V'$, that is, motion will be limited to the range in θ for which θ is real. We denote the two extreme values of θ by θ_1 and θ_2. The symmetry axis Oz of the top will be bobbing back and forth between the two right circular cones of half-angles θ_1 and θ_2, while processing with the angular velocity $\dot{\phi}$ about Oz'. Such a bobbing motion is termed a nutation. A steady or pure precession (precession without nutation) is also possible. We shall discuss this latter case first.

12.8.1 Precession without Nutation

The energy diagram exhibits a single minimum at $\theta = \theta_m$. The motion at such a fixed angle of inclination is a steady precession. The value of θ_m may be found by setting

$$\left.\frac{\partial V'}{\partial \theta}\right|_{\theta=\theta_m} = 0$$

and we find θ_m to be given by

$$I\cos\theta_m B^2 - Ib\sin^2\theta_m B - Mgh\sin^4\theta_m = 0 \tag{12.112}$$

where $B = a - b\cos\theta$. This is a quadratic equation in B; we can solve for B in terms of θ_m and obtain

$$B = \frac{b\sin^2\theta_m}{2\cos\theta_m}\left[1 \pm \sqrt{1 - \frac{4Mgh\cos\theta_m}{Ib^2}}\right]. \tag{12.113}$$

For a physical solution, B must be real. Thus, the quantity under the radical sign cannot be negative. If $\theta_m < 90^0$, we have

$$Ib^2 > 4\,Mgh\cos\theta_m \text{ or } (Ib)^2 > 4\,IMgh\cos\theta_m. \tag{12.114}$$

Now, from Equation 12.103,

$$p_\psi = I_3\left(\dot{\psi} + \dot{\phi}\cos\theta\right) = I_3\omega_3 = Ib$$

we find

$$\omega_3 = Ib/I_3.$$

Combining this with Equation 12.114 yields

$$\omega_3 > \frac{2\sqrt{MghI\cos\theta_m}}{I_3}. \tag{12.115}$$

Only if the top has at least this minimum value of the angular velocity of spin is steady precession possible. For $\theta = \theta_m$, Equation 12.105 gives

$$\phi'_m = \frac{B}{\sin^2\theta_m} \tag{12.116}$$

which means that we have two possible values of the precessional angular velocity: one for each of the values of B given by Equation 12.113. For a rapidly spinning top, ω_3 (or Ib) is much greater than the minimum value just given. Thus, the second term in the radical of Equation 12.113 is small, and we can expand the radical. Retaining only the first non-vanishing term in each case, we find, depending on whether we choose the plus or the minus sign,

$$\text{fast precession: } \dot{\phi}_{m(+)} \cong \frac{I_3\omega_3}{I\cos\theta_m} \tag{12.117}$$

$$\text{slow precession: } \dot{\phi}_{m(+)} \cong \frac{Mgh}{I_3\omega_3}. \tag{12.118}$$

A slow precession is usually observed. It should be emphasized that steady precession (motion with $\dot{\theta} = 0$ and $\dot{\phi} = \text{constant}$) can occur only for a very particular set of initial conditions. The apparent absence of the θ motion in practical cases is a result of the fact that the spin is usually very large. In this event, it can be shown that the range of variation of θ will be very small during the motion and, thus, is often overlooked.

It is commonly known that if a top is started spinning sufficiently fast and with its axis vertical, the axis of the top will remain steady in the upright position for a while, a condition called sleeping.

This corresponds to the constant value of zero for θ in Equation 12.115. Thus, the criterion for stability of the sleeping top is given by

$$\omega_3 > \frac{2\sqrt{MghI}}{I_3}. \tag{12.119}$$

Friction gradually slows the top down so that the condition just given no longer holds. The top then begins to fall and topples over eventually.

12.8.2 Precession with Nutation

For the general case, in which $\theta_1 < \theta < \theta_2$, the symmetry axis of the top will be bobbing back and forth between the two right circular cones of half-angles θ_1 and θ_2 while processing with the angular velocity $\dot{\phi}$ about the Oz' axis. If $\dot{\phi}$ does not change sign as θ varies between its limits, the path described by the projection of the symmetry axis of the body on a unit sphere with center at the origin is depicted in Figure 12.26a. The horizontal circles are the intersections of the cones of half-angles θ_1 and θ_2 with the unit sphere. If $\dot{\phi}$ does change sign between the limiting values of θ, the path described by the nutational–precessional symmetry axis is a loop as shown in Figure 12.26b. If $\dot{\phi}$ vanishes at one of the limiting values of θ, say, θ_1, it gives rise to the cusp-like motion depicted in Figure 12.26c.

We now calculate these motions for a frequently observed case, namely, a rapidly spinning top that has slow precession and small nutation. An approximate solution for the motion with this condition is not difficult to find. We first rewrite Equation 12.102 as

$$I\sin^2\theta\dot{\phi} + I_3\omega_3\cos\theta = Ia$$

and then, differentiating it with respect to time t, we obtain

$$I\sin\theta\ddot{\phi} = I_3\omega_3\dot{\theta}. \tag{12.120}$$

Next, from Lagrange's equation for θ, we have

$$I\ddot{\theta} = \left(Mgh - I_3\omega_3\dot{\phi}\right)\sin\theta. \tag{12.121}$$

In arriving at Equations 12.120 and 12.121, the quadratic terms in $\dot{\phi}$ and $\dot{\theta}$ have been neglected. If we define

$$\omega_P = \frac{Mgh}{I_3\omega_3}, \quad \omega_l = \frac{I_3\omega_3}{I} \tag{12.122}$$

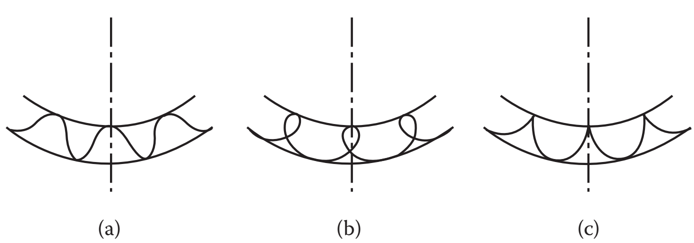


FIGURE 12.26 (a) $\dot{\phi}$ never vanishes; (b) $\dot{\phi}$ changes sign at $\phi > \phi_1$; and (c) $\dot{\phi}$ vanishes $\phi = \phi_1$.

then, in terms of ω_p and ω_l, Equations 12.120 and 12.121 become

$$\ddot{\phi}\sin\theta = \omega_l\dot{\theta}, \quad \ddot{\theta} = \omega_l\left(\omega_p - \dot{\phi}\right)\sin\theta. \tag{12.123}$$

If we differentiate the first equation with respect to time and then substitute the second, we obtain

$$\frac{d^2\dot{\phi}}{dt^2} + \omega_l^2\dot{\phi} = \omega_l^2\omega_p \tag{12.124}$$

where a quadratic term of order $\ddot{\phi}\dot{\theta}$ has been dropped. The solution of this equation is of the form

$$\dot{\phi}(t) = \omega_p + A\cos(\omega_l t + \alpha).$$

For the initial conditions $\dot{\phi} = \omega_0, \phi = 0, \dot{\theta} = 0$, and $\theta = \theta_0$ at $t = 0$, we find $A = \omega_0 - \omega_p$, and $\alpha = 0$. Accordingly, we have

$$\dot{\phi}(t) = \omega_p - (\omega_p - \omega_0)\cos\omega_l t, \quad \phi(t) = \omega_p t - \frac{\omega_p - \omega_0}{\omega_l}\sin\omega_l t\,. \tag{12.125}$$

To find θ, we substitute Equation 12.125 into the first equation of Equation 12.123. This gives

$$\theta = (\omega_p - \omega_0)\sin\omega_l\, t\,\sin\theta_0. \tag{12.126}$$

In arriving at this result, we have made the approximation $\theta \cong \theta_0$ because of the small size of ω_p and ω_0. The integration of the preceding equation gives the solution for θ:

$$\theta(t) = \theta_0 + \frac{\omega_p - \omega_0}{\omega_l}\sin\theta_0(1 - \cos\omega_l t) \tag{12.127}$$

which exhibits nutation of the top between the limits θ_0 and $\theta_0 + 2[(\omega_p - \omega_0)/\omega_l]\sin\theta_0$. The sign of $(\omega_p - \omega_0)$ determines which is the θ_2 (the upper bound of θ) and which is the θ_1 (the lower bound). A precession $\varphi(t)$ that has a sinusoidal motion superimposed upon it is the nutation.

12.9 STABILITY OF ROTATIONAL MOTION

In this section, we examine the equation of the stability of force-free rotation about the principal axes of the rigid body. This has important applications; for example, the stability of a spinning satellite is of importance. Stability means that if a small perturbation is applied to the system, the motion will either return to its original mode or will execute small oscillations about it. We shall see that stable rotation occurs around the principal axes corresponding to the greatest and smallest moments of inertia, while rotation is unstable about the median axis.

We consider a rigid body for which all of the principal moments of inertia are distinct: $I_3 \neq I_2 \neq I_1$, and choose the body axes that coincide with the principal axes. If the body is rotating around the principal axis corresponding to I_1, we examine the effect of the perturbation producing a small rotation about the other axes. The angular velocity vector now has the form

$$\vec{\omega} = \omega_1\hat{e}_1 + \omega_2\hat{e}_2 + \omega_3\hat{e}_3 \tag{12.128}$$

where ω_2 and ω_3 are small perturbations, whose product can be neglected compared to all other quantities of interest to the discussion. Accordingly, the force-free Euler's Equation 12.82 becomes

$$\dot{\omega}_1 = 0, \quad \dot{\omega}_2 = \omega_3\omega_1(I_3 - I_1)/I_2, \quad \dot{\omega}_3 = \omega_2\omega_1(I_1 - I_2)/I_3. \tag{12.129}$$

The first of these equations indicates ω_1 = constant. We next differentiate the second equation and substitute the third equation and obtain the approximate equation of motion for ω_2:

$$\ddot{\omega}_2 + K\omega_2 = 0 \tag{12.130}$$

where

$$K = (I_1 - I_3)(I_1 - I_2)\frac{\omega_1^2}{I_2 I_3}.$$

The behavior of ω_2 depends mainly on the sign of K. For positive K, ω_2 will execute a simple harmonic motion of constant amplitude, and negative K implies exponential growth of ω_2 (i.e., it is unstable). The condition for stable motion can then be written as

$$(I_1 - I_3)(I_1 - I_2) > 0$$

which is equivalent to

$$I_1 < I_2 \text{ and } I_1 < I_3; \text{ or } I_1 > I_2 \text{ and } I_1 > I_3. \tag{12.131}$$

Similarly, we can examine $\omega_3(t)$ with the same result. Thus, we conclude that stable rotation occurs about the principal axes corresponding to the greatest and the smallest moment of inertia, and rotation about the principal axis corresponding to the intermediate moment of inertia is unstable.

Friction is not included in our discussion. In the presence of friction, analysis will show that only rotation about the principal axis corresponding to the greatest moment of inertia is stable. That is why a person falling off a mountain always ends up falling head over heels. We shall not examine this dissipate case here because it is too complex.

When two of the moments of inertia are equal, say, $I_1 = I_2 = I$, there is stability only about the principal axis corresponding to the moment of inertia I_3, independent of whether I_3 is greater or less than I.

PROBLEMS

1. Find the center of mass of each of the following:
   (a) A solid homogeneous hemisphere of radius a
   (b) A semicircular, homogeneous flat plate of radius a
2. Find the center of mass of the quadrant of an elliptical plate of constant thickness enclosed by the two semi-axes. The density varies in such a way that at any point, it is directly proportional to the distance from the point to the major axis.
3. Find the moment of inertia and the radius of gyration of each of the following:
   (a) A thin circular hoop or cylindrical shell about its axis
   (b) A right circular cone of height h and radius a about its axis
4. Use the parallel axis theorem to find the moment of inertia of a solid circular cylinder about a line on the surface of the cylinder and parallel to the axis of the cylinder.

5. Consider a rigid body that is in the form of a plane lamina of any shape. If the lamina is in the xy-plane, and I_x, I_y, and I_z denote the moments of inertia about the x-, y-, and z-axes, respectively, prove that

$$I_z = I_x + I_y.$$

This is known as the perpendicular axis theorem for a plane lamina.

6. (a) Find the inertia tensor for a uniform square plate of side a and mass m about a diagonal in a coordinate system $Oxyz$, where O is at one corner and the x- and y-axes are along two edges.
   (b) Find the angular momentum about the origin for the plate just mentioned in (a) when it is rotating with an angular speed ω about (1) the x-axis and (2) about the diagonal through the origin.
7. (a) Find the principal moments of inertia of a square plate about a corner. (b) Find the directions of the principal axes for part (a).
8. Find (a) the principal moments of inertia at the center of a uniform rectangular plate of sides a and b and (b) the torque needed to rotate such a plate about a diagonal with constant angular velocity ω.
9. A solid cylinder of radius a and mass M rolls down an inclined plane of angle α. The coefficient of friction between the cylinder and inclined plane is μ. Discuss the motion for various values of μ.
10. A cube of edge s and mass M is suspended vertically from one of its edges (Figure 12.27).
    (a) Show that the period of small oscillations is $P = 2\pi\sqrt[4]{2}\sqrt{s/3g}$.
    (b) What is the length of the equivalent pendulum?
11. A dumbbell (two equal masses connected by a massless rigid rod of length $2a$) rotates at a fixed inclination α with a constant angular velocity ω about a pivot point at the center of the rod (Figure 12.28). Find the angular momentum and torque of the system.
12. Let $\vec{T}$ be a symmetric tensor, and λ_1, λ_2, and λ_3 are its three real eigenvalues with the three corresponding eigenvectors A_1, A_2, and A_3. Show that the three eigenvectors A_1, A_2, and A_3 are orthogonal if the three eigenvalues λ_1, λ_2, and λ_3 are distinct.
13. An automobile wheel has its axle (axis of rotation) slightly bent relative to the symmetry axis of the wheel. To remedy the situation, counterbalance weights can be suitably located on the rim so as to make the axle a principal axis for the total system (weights and wheel). For simplicity, we consider the wheel as a thin, uniform circular disk of radius a and mass

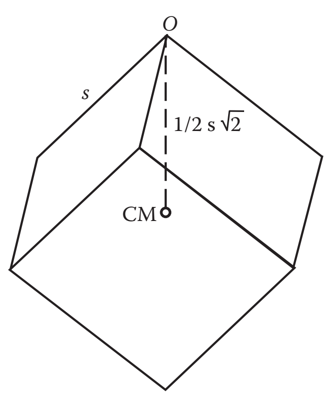


FIGURE 12.27 Cube suspended vertically from one of its edges at O.

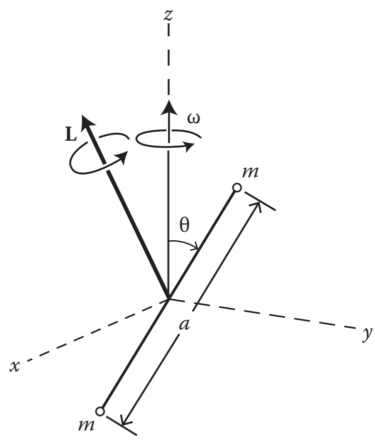


FIGURE 12.28 Dumbbell rotates at a fixed inclination θ with a constant angular velocity ω.

M as shown in Figure 12.29. The axle inclines by a small angle θ relative to the x-axis (symmetry axis of the disk). Two balancing weights (m') lie in the xy-plane. Show that if

$$bm' = \frac{\theta\, aM}{8}$$

the wheel will be dynamically balanced.

14. A coin is initially set spinning rapidly on its edge on a table. As the spin decreases as a result of frictional forces, a point is reached where the coin begins to precess (wobble). With further loss of energy, the coin will gradually lie down with an increase of the precession rate, and at the same time, the actual turning rate, as seen above, decreases. Explain the curious behavior of a spinning coin just described with your calculations in detail. (Assume frictional torque is negligible in comparison to the gravitational torque.)

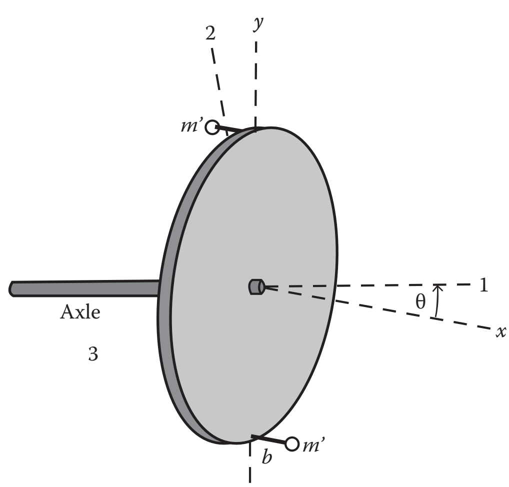


FIGURE 12.29 Automobile wheel has its axle (axis of rotation) slightly bent relative to the symmetry axis of the wheel.

15. A gyroscopic compass is used for navigation on ships and airplanes. In its simplest form, such a compass consists of a rapidly rotating circular disk whose axis is free to turn in a horizontal plane. Show that because of the Earth's motion, the system is in equilibrium when the axis of the rotating disk lies along a geographical north–south line or along a meridian.
16. Consider a uniform bar of cylindrical cross section, mounted as shown in Figure 12.30, rotating about a horizontal shaft with constant angular velocity $\vec{\omega}$. The supports for the shaft are mounted on a turntable that revolves about a vertical axis at constant angular velocity $\vec{\Omega}$. Discuss the motion of this rotating bar by using Euler's equations.
17. Consider the motion of a cylindrical symmetry space vehicle (Figure 12.31) in its coasting phases (i.e., all external forces on the vehicle are negligibly different from zero). Show that its rate of precession is proportional to its rate of spinning. Is the precession always parallel to the spin?
18. A uniform thin circular disk is constrained to spin with an angular velocity ω about an axis passing through the center of the disk but making an angle α with the axis of symmetry of the disk. It is suddenly released. Find the half-angle of the cone in space described by the symmetry axis and the time in which this cone is described.

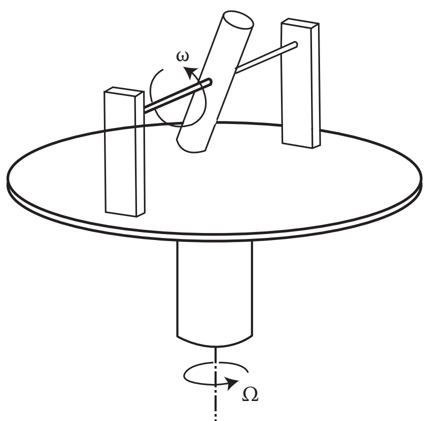


FIGURE 12.30 Rotating turntable and a rotating cylindrical bar.

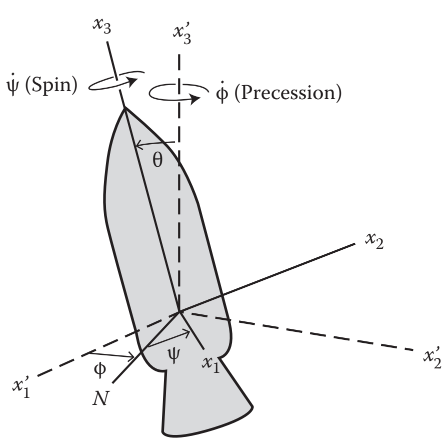


FIGURE 12.31 Cylindrical symmetry space vehicle.

19. An electron of mass m is revolving around the nucleus in an atom along a circular orbit of radius r. An external magnetic field B acts in a direction that makes an angle θ with the normal to the electron's orbital plane. The magnetic field produces a torque

$$\vec{N} = -\left[\frac{1}{2}e(\vec{r}\times\vec{v})\right]\times\vec{B}$$

where e is the electronic charge, $\vec{v}$ is the velocity of the electron, and the quantity within the brackets is the magnetic moment of the effective circular current loop. Show that the normal to the electron's orbital plane precesses with an angular velocity $e\vec{B}/2m$ about the direction of the field (known as the Larmor precession).

20. For forces that act only for a very short time, Equation 12.76 can be rewritten in integral form:

$$\mathrm{d}\vec{L} = \int \vec{N}^{(e)}\,\mathrm{d}t.$$

The time integral on the right-hand side is called the angular (rotational) impulse. For rotation about a fixed z-axis, the preceding equation becomes

$$\mathrm{d}L_z = I_{z\,z}\Delta\omega_z = \int N_z^{(e)}\,\mathrm{d}t.$$

We now apply this result to the dynamics of billiard shots. Consider a stationary billiard ball of radius a on a pool table. If the cue stick hits the ball in its vertical median plane in the horizontal direction at a vertical distance h above the table, show that the condition for the ball to roll without slipping is $h = 7a/5$.

REFERENCES

Goldstein, H. *Classical Mechanics*, Addison Wesley, Readings, Massachusetts, 1980.
Kleppner, D., and Kolenkow, R.J. *An Introduction to Mechanics*, McGraw-Hill, New York, 1973.

13 Theory of Special Relativity

We now examine the modification of the structure of classical mechanics brought about by the special theory of relativity. We do not intend to present a comprehensive discussion of the special theory of relativity; only its essential parts are outlined in the following section. Our main interest is to see how to incorporate the special theory of relativity into the framework of classical mechanics.

13.1 HISTORICAL ORIGIN OF SPECIAL THEORY OF RELATIVITY

Before Einstein, the concept of space and time were those described by Galileo and Newton. In any unaccelerated frame of reference, called the inertial reference frame, Newton's laws of motion are valid, especially the first law, which states that free objects maintain a state of uniform motion. Time was assumed to have an absolute or universal nature in the sense that any two inertial observers who have synchronized their clocks will always agree on the time of any event. An event is any happening that can be given space and time coordinates.

The Galileo transformation asserts that any one inertial frame is as good as any other one describing the laws of classical mechanics. However, physicists of the 19th century were not able to grant the same freedom to electromagnetic theory, which did not seem to be Galilean invariant. It is worthwhile to spend some time examining this inconsistency of electromagnetism and Galilean relativity. Classical electromagnetism is summarized in Maxwell's four differential equations, which have the form (in Gaussian units in which the electric and magnetic field vectors have the same dimensions)

$$\nabla\times\vec{E} = -\frac{1}{c}\frac{\partial\vec{B}}{\partial t}, \quad \nabla\cdot\vec{B} = 0,$$
$$\nabla\times\vec{H} = \frac{1}{c}\left[\frac{\partial\vec{D}}{\partial t} + 4\pi\vec{j}\right], \quad \nabla\cdot\vec{D} = 4\pi\rho$$

where c is the ratio of electromagnetic and electrostatic units of charge. If we consider only empty space, we have $\vec{D} = \vec{E}$, $\bar{B} = \vec{H}$, and in the absence of charges and currents, we find

$$\left.\begin{aligned} \nabla\times\vec{E} &= -\frac{1}{c}\frac{\partial\vec{H}}{\partial t}, \quad \nabla\cdot\vec{H} = 0, \\ \nabla\times\vec{H} &= \frac{1}{c}\frac{\partial\vec{E}}{\partial t}, \quad \nabla\cdot\vec{E} = 0 \end{aligned}\right\}. \tag{13.1}$$

Hence,

$$\nabla\times(\nabla\times\vec{H}) = \nabla(\nabla\cdot\vec{H}) - \nabla^2\vec{H} = \frac{1}{c}\frac{\partial}{\partial t}\nabla\times\vec{E} = -\frac{1}{c^2}\frac{\partial^2\vec{H}}{\partial t^2}$$

so that the auxiliary magnetic vector $\vec{H}$ satisfies the wave equation

$$\nabla^2\vec{H} = \frac{1}{c^2}\frac{\partial^2\vec{H}}{\partial t^2} \tag{13.2}$$

with a similar equation for $\vec{E}$. It follows that disturbances in the fields propagate with velocity c. This suggests the identification of light and electromagnetic radiation, and such identification gives a very satisfactory explanation of optical phenomena. But the wave equation 13.2 contains no reference to the velocity of the source of the light, and this naturally suggests that the velocity of the light must be independent of the velocity of its source. This is in agreement with observations in astronomy. For example, there exist certain binary stars consisting of two stars moving in orbits about their common center of mass. At one point in the orbit, one star will be traveling toward the Earth, and another, directly away. If the center of mass is at a distance d from the Earth, the light will reach the Earth at a time of order d/c after it has been emitted, where c is the velocity of light. We will show, in a moment, that for any small change v in c, we have a change in the time of arrival by an amount given by

$$\Delta t = -\mathrm{d}v/c^2.$$

This change would produce apparent irregularities in the motion of such stars. No such irregularities have been observed, and we are, therefore, to conclude that the velocity of light is independent of the velocity of the source.

To derive the simple result of change in arrival time, let us assume, for the sake of simplicity, that there are two stars, A and B, of equal mass moving on opposite sides of a circular orbit around their common center of mass C. Let us consider an observer P, at a very great distance d, from the center of the orbit of the stars, so that the angle θ subtended by their orbits is always very small (Figure 13.1). We consider only the light rays that are emitted in such a way as eventually to reach the point P (the speed of light always being c relative to the emitting star). Let us begin with those rays that reach P after being emitted at the time t_1 when both stars are in the line of sight, that is, at A and B. The time taken for light from the nearer star, A, to reach P will be

$$t_{1A} = \frac{d-a}{c}$$

where a is the radius of the orbit, and that light from the farther star, B, will be

$$t_{1B} = \frac{d+a}{c}.$$

At the time t_2 when the two stars are A' and B' (the diameter of the orbit is perpendicular to the line of sight), the light from A', which is receding from P, will have a velocity of $c - v$, and that from B, which is approaching P, will have a velocity of $c + v$ (where v is the speed of rotation of the orbit). The times taken for this light to reach P will be

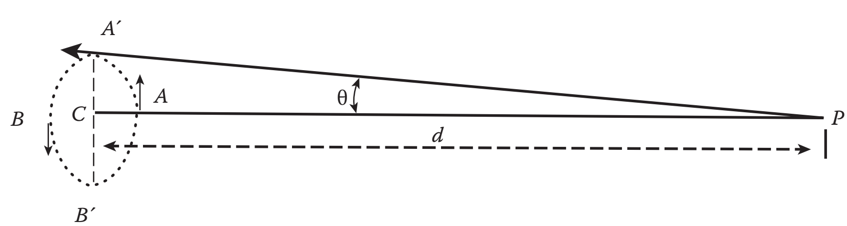


FIGURE 13.1 Motion of a binary star supports the fact that the velocity of light is independent of the velocity of the source.

$$t_{2A} = \frac{d}{c-v} \quad t_{2B} = \frac{d}{c+v}\,.$$

We now compute the time difference for the same star at two different positions, say, star *A*.

$$\Delta t = t_{2A} - t_{1A} = \frac{d}{c-v} - \frac{d-a}{c} \cong \frac{a}{c} - \frac{d}{c}\frac{v}{c} \approx -\frac{dv}{c^2}$$

because *d* is an astronomical order of distance, and the radius of the orbit *a* is negligibly small in comparison with *d*.

Let us return to the finding of the independence of velocity of light and the velocity of the source. Now, the independence of velocity of light and the velocity of the source pose the problem of the frame of reference with respect to which *c* is to be measured. In the theory of sound, a similar problem arises, but there, it is easily resolved because the speed is to be measured relative to the still air. In the 19th century, it seemed reasonable to give a similar answer in the case of light. Because of the works of Young and Fresnel, light was viewed as a mechanical wave, so its propagation required a physical medium that was called the ether. Because light can travel through space, it was assumed that ether fills all of space, and the velocity of light must be measured with respect to the stationary unobserved ether. This would have the great advantage of linking the hitherto separated theories of mechanics and electromagnetism. A difficulty, at once, arises. As mechanics holds in every one of the inertial frames, the Maxwell equations should then hold in every one of the inertial frames. It is easy to see that this cannot be so. Let us apply the Lorentz transformation

$$x' = x - vt,\ y' = y,\ z' = z,\ t' = t$$

to the Maxwell's equations. For example, the first component of the equation

$$\nabla \times \vec{H} = \frac{1}{c}\frac{\partial \vec{E}}{\partial t}$$

becomes

$$\frac{\partial H_{z'}}{\partial y'} - \frac{\partial H_{y'}}{\partial z'} = \frac{1}{c}\left(\frac{\partial E_{x'}}{\partial t'} - v\frac{\partial E'}{\partial x'}\right). \qquad (13.3)$$

However, there is no other equation in the set linking $\partial E_{x'}/\partial t'$ and $\partial E_{x'}/\partial x'$, so it is impossible to transform the right-hand side, by transformation of the field vectors, in such a way that the transformed equation would read

$$\frac{\partial H_{z'}}{\partial y'} - \frac{\partial H_{y'}}{\partial z'} = \frac{1}{c}\frac{\partial E_{x'}}{\partial t'}\,.$$

Alternatively, we can use a much simpler argument. Consider that a source at rest in an inertial frame *S* emits a light wave, which travels out as a spherical wave at a constant speed. But, observed in a frame *S*′ that moves uniformly with respect to *S*, the light wave is no longer spherical, and the speed of light is also different.

Therefore, for electromagnetic phenomena, inertial frames of reference are not equivalent under the Galilean transformation. A number of attempts were proposed to resolve this conflict. These include the following:

1. The Maxwell equations are wrong and need to be modified to obey the Galilean transformation.
2. There is a preferred frame of reference: that of stationary unobserved ether. The Maxwell equations require modification in other inertial frames of reference.
3. The Maxwell equations are correct and have the same form in all inertial reference frames. There is some transformation other than the Galilean transformation that makes both electromagnetic and mechanical equations transform in an invariant way.

As we know now, the third proposal is the correct one, but it was not accepted without resistance. The first attempt was abandoned rather quickly. When it was tried, the new terms that make the Maxwell equations Galilean invariant led to predictions of new phenomena that did not exist when tested experimentally. So the attempt was abandoned. The second attempt was ruled out only after extensive experiments. In the following section, we will review a critical experiment that eventually led us to give up the hypothetical ether, the Michelson–Morley experiment.

Before we proceed to describe the Michelson–Morley experiment, let us make a minor observation: The ether has to be assumed to have contradictory mechanical properties; it is the softest and also the hardest substance. It must be the softest because all material bodies can pass through it without any resistance from the ether. Otherwise, for example, the Earth would have slowed down and fallen into the sun during the billions of years of its traveling around the sun. On the other hand, ether must be harder than any material because light (ether vibration) travels with such a high speed that its elastic constant must be the highest of all known materials. Surprisingly, such contradictions did not prevent physicists of the 19th century from clinging to their belief in the hypothetical ether.

13.2 MICHELSON–MORLEY EXPERIMENT

The existence of ether and the law of velocity addition (according to the Galilean transformation) suggest that it should be possible to detect some variation of the speed of light as emitted by some terrestrial source. As the Earth travels through space at 30 km/s in an almost circular orbit around the sun, it is bound to have some relative velocity with respect to the ether. If this relative velocity is added to that of the light emitted from the source, then light emitted simultaneously in two perpendicular directions should be traveling at different speeds, corresponding to the two relative velocities of the light with respect to the ether.

In one of the most famous and important experiments in physics, Michelson set out, in 1887, to detect this variation in the velocity of the propagation of light. Michelson's ingenious way of doing this depends on the phenomenon of interference of light to determine whether the time taken for light to pass over two equal paths at right angles was different or not. He designed and constructed an interferometer, schematically shown in Figure 13.2. The interferometer is essentially composed of a light source S, a half-silvered glass plate A, and two mirrors B and C, all mounted on a rigid base. The two mirrors B and C are placed at equal distances L from the plate A. Light from S enters A and splits into two beams. One goes to mirror B, which reflects it back; the other beam goes to mirror C, also to be reflected back. On arriving back at A, the two reflected beams are recombined as two superposed beams, D and F, as indicated. If the time taken for light from A to B and back equals the time from A to C and back, the two beams D and F will be in phase and will reinforce each other. If the two times differ slightly, the two beams will be slightly out of phase, and they produce an interference pattern. A typical interference pattern is sketched in Figure 13.3.

We now calculate the two times to see whether they are the same or not. First, calculate the time required for the light to go from A to B and back. If the line AB is parallel to the Earth's motion in its orbit, and if the Earth is moving at a speed u and the speed of light in the ether is c, the time is

$$t_1 = \frac{L_{AB}}{c-u} + \frac{L_{AB}}{c+u} = \frac{2L_{AB}}{c[1-(u/c)^2]} \approx \frac{2L_{AB}}{c}\left(1+\frac{u^2}{c^2}\right) \tag{13.4}$$

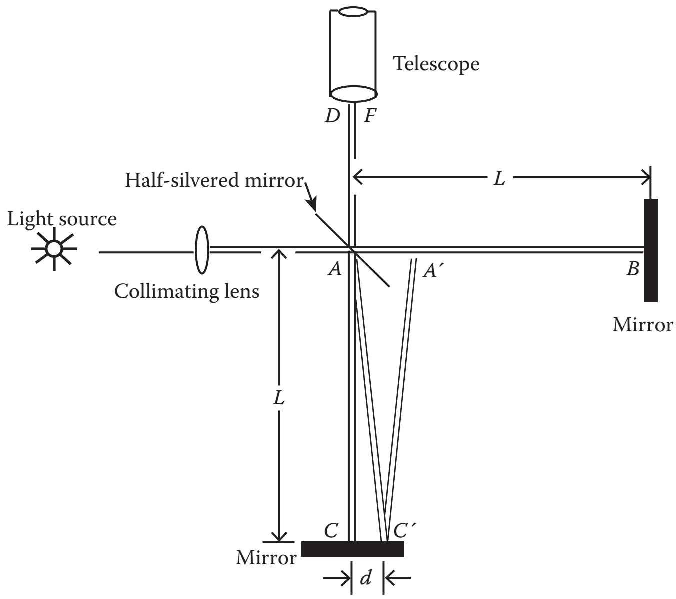


FIGURE 13.2 Schematic diagram of the Michelson–Morley experiment.

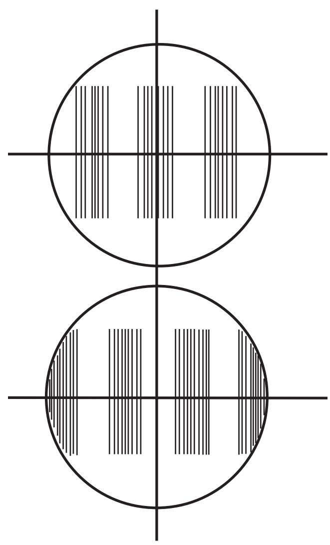


FIGURE 13.3 Sketch of a typical interference pattern.

where $(c - u)$ is the upstream speed of light with respect to the apparatus, and $(c + u)$ is the downstream speed.

Our next calculation is of the time t_2 for the light to go from A to C. We note that while light goes from A to C, the mirror C moves to the right relative to the ether through a distance $d = ut_2$ to the position C'; at the same time, the light travels a distance ct_2 along AC'. For this right triangle, we have

$$(ct_2)^2 = L_{AC}^2 + (ut_2)^2$$

from which we obtain

$$t_2 = \frac{L_{AC}}{\sqrt{c^2 - u^2}}.$$

Similarly, as the light is returning to the half-silvered plate, the plate moves to the right to the position B'. The total path length for the return trip is the same as can be seen from the symmetry of Figure 13.2. Therefore, if the return time is also the same, the total time for light to go from A to C and back is then $2t_2$, which we denote by t_3:

$$t_3 = \frac{2L_{AC}}{\sqrt{c^2 - u^2}} = \frac{2L_{AC}}{c\sqrt{1-(u/c)^2}} \approx \frac{2L_{AC}}{c}\left(1 + \frac{u^2}{2c^2}\right). \tag{13.5}$$

In Equations 13.3 and 13.4, the first factors are the same and represent the time that would be taken if the apparatus were at rest relative to the ether. The second factors represent the modifications in the times caused by the motion of the apparatus. Now, the time difference Δt is

$$\Delta t = t_3 - t_1 = \frac{2(L_{AC} - L_{AB})}{c} + \frac{L_{AC}}{c}\beta^2 - \frac{2L_{AB}}{c}\beta^2 \tag{13.6}$$

where $\beta = u/c$.

It is most likely that we cannot make $L_{AB} = L_{AC} = L$ exactly. In that case, we can rotate the apparatus 90°, so that AC is in the line of motion, and AB is perpendicular to the motion. Any small difference in length becomes unimportant. Now, we have

$$\Delta t' = t'_3 - t'_1 = \frac{2(L_{AB} - L_{AC})}{c} + \frac{2L_{AB}}{c}\beta^2 - \frac{L_{AC}}{c}\beta^2. \tag{13.7}$$

Thus,

$$\Delta t' - \Delta t = \frac{(L_{AB} + L_{AC})}{c}\beta^2. \tag{13.8}$$

This difference yields a shift in the interference pattern across the crosshairs of the viewing telescope. If the optical path difference between the beams changes by one λ (wavelength), for example, there will be a shift of one fringe. If δ represents the number of fringes moving past the crosshairs as the pattern shifts, then

$$\delta = \frac{c(\Delta t' - \Delta t)}{\lambda} = \frac{L_{AB} + L_{AC}}{\lambda}\beta^2 = \frac{\beta^2}{\lambda/(L_{AB} + L_{AC})}. \tag{13.9}$$

In the Michelson–Morley experiment of 1887, the effective length L was 11 m; sodium light of $\lambda = 5.9 \times 10^{-5}$ cm was used. The orbit speed of the Earth is 3×10^4 m/s, so $\beta = 10^{-4}$. From Equation 13.6, the expected shift would be about 4/10 of a fringe:

$$\delta = \frac{22m \times (10^{-4})^2}{5.9 \times 10^{-5}} = 0.37. \tag{13.10}$$

A shift of 0.005 fringe can be detected by the Michelson–Morley interferometer. However, no fringe shift in the interference pattern was observed. Thus, no effect at all resulting from the Earth's motion through the ether was found. This null result was very puzzling and most disturbing at the time. How could it be? It was suggested, including by Michelson, that the ether might be dragged along by the Earth, eliminating or reducing the ether wind in the laboratory. This is hard to square with the picture of the ether as an all-pervasive, frictionless medium. The ether's status as an absolute reference frame was also gone forever. Many attempts to save the ether failed (Resnick and Halliday 1985); we just mention one here—namely, the contraction hypothesis.

George Francis Fitzgerald pointed out in 1892 that a contraction of bodies along the direction of their motion through the ether by a factor $(1 - u^2/c^2)^{1/2}$ would give the null result. Because Equation 13.4 must be multiplied by the contraction factor $(1 - u^2/c^2)^{1/2}$, Δt, Equation 13.6, reduces to zero. The magnitude of this time difference is completely unaffected by rotation of the apparatus through 90°.

Lorentz obtained a contraction of this sort in his theory of electrons. He found the field equations of electron theory remain unchanged if a contraction by the factor $(1 - u^2/c^2)^{1/2}$ takes place, provided also that a new measure of time is used in a uniformly moving system. The outcome of the Lorentz theory is that an observer will observe the same phenomena, no matter whether he or she is at rest in the ether or moving with velocity. Thus, different observers are equally unable to tell whether they are at rest or moving in the ether. This means that for optical phenomena, just as for mechanics, ether is unobservable.

Poincare offered another line of approach to the problem. He suggested that the result of the Michelson–Morley experiment was a manifestation of a general principle that absolute motion cannot be detected by laboratory experiments of any kind, and the laws of nature must be the same in all inertial frames.

13.3 POSTULATES OF SPECIAL THEORY OF RELATIVITY

Einstein realized the full implications of the Michelson–Morley experiment, the Lorentz theory, and Poincare's principle of relativity. Instead of trying to patch up the accumulating difficulties and contradictions connected with the notion of ether, Einstein rejected the ether idea as unnecessary or unsuitable for the description of the physical world and returned to the pre-ether idea of a completely empty space. Along with the exit of ether from the stage of physics, the notion of absolute motion through space is also gone. The Michelson–Morley experiment proves unequivocally that no such special frame of reference exists. All frames of reference in uniform relative motion are equivalent for mechanical motions and also for electromagnetic phenomena. Einstein extended this as a fundamental postulate, now known as the principle of relativity. Furthermore, he argued that the velocity of light predicted by electromagnetic theory must be a universal constant, the same for all observers. He took an epoch-making step in 1905 and developed the special theory of relativity from these two basic postulates (assumptions), which are rephrased as follows:

1. **The principle of relativity**: the laws of physics are the same in all inertial frames. No preferred inertial frame exists.

   Two people standing in the aisle of an airplane going 700 km/h can play catch exactly as they would on the ground. On that airplane, if you drop a heavy and a light ball together, they fall at the same rate as they would if you dropped them on the ground, and they hit the cabin floor at the same time. When you are moving uniformly, you do not experience the physical sensation of motion. Experiments indicate that the principle of relativity also applies to electromagnetism; it is very general. That radios and tape recorders work the same on an airplane as they do in the house is a simple example.
2. **The principle of the constancy of the velocity of light**: the velocity of light in empty space is the same in all inertial frames and is independent of the motion of the emitting body.

> According to Einstein, sometime in 1896, after he entered the Zurich Polytechnic Institute to begin his education as a physicist, he asked himself the question of what would happen if he could catch up to a light ray, that is, move at the speed of light. Maxwell's theory says that light is a wave of electric and magnetic fields that moves like a water wave through space. But if you could catch up to one of Maxwell's light waves the way a surfboard rider catches an ocean wave for a ride, then the light wave would not be moving relative to you, but instead would be standing still. The light wave would then be a standing wave of electric and magnetic fields that is not allowed if Maxwell's theory is right. So, he reasoned, there must be something wrong with the assumption that you can catch a light wave as you can catch a water wave. This idea was a seed from which the fundamental postulate of the constancy of the speed of light and the special theory of relativity grew nine years later.

All the seemingly very strange results of special relativity came from the special nature of the speed of light. Once we understand this, everything else in relativity makes sense. So we take a brief look at this special nature of the speed of light. The speed of light is very great, 3×10^8 m/s. However, the bizarre fact of the speed of light is that it is independent of the motion of the observer or the source emitting the light. Michelson hoped to determine the absolute speed of the Earth through ether by measuring the difference in times required for light to travel across equal distances that are at right angles to each other. What did he observe? No difference in travel times for the two perpendicular light beams. It was as if the Earth were absolutely stationary. The conclusion is that the speed of light does not depend on the motion of the observer. We saw earlier that the speed of light is also independent of the speed of the source. The special nature of the speed of light is not something we would expect from common sense. The same common sense was once objecting to the idea that the Earth is round. Hence, common sense is not always right.

13.3.1 Time Is Not Absolute

The constancy of the velocity of light puts an end to the notion of absolute time. We know that Newtonian mechanics abolished the notion of absolute space but not of absolute time. Now, time is also not absolute anymore. Because all inertial observers must agree on how fast light travels but not how far light travels, space is not absolute. Now, time taken is the distance light has traveled divided by the speed of light; thus, they must disagree over the time that the journey took. And time lost its universal nature. In fact, we shall see later that moving clocks run slow. This is known as time dilation.

13.4 LORENTZ TRANSFORMATIONS

Because the Galilean transformations are inconsistent with Einstein's postulate of the constancy of the speed of light, we must modify it in such a way that the new transformation will incorporate Einstein's two postulates and make both mechanical and electromagnetic equations transform in an invariant way.

To this end, we consider two inertial frames S and S'. For simplicity, let the corresponding axes of the two frames be parallel with frame S' moving at a constant velocity u relative to S along the x_1-axis. The apparatuses used to measure distances and times in the two frames are assumed identical, and both clocks are adjusted to read zero at the moment the two origins coincide. Figure 13.4 represents the viewpoint of observers in S.

Suppose that an event occurred in frame S at the coordinates (x, y, z, t) that is observed at (x', y', z', t') in frame S'. Because of the homogeneity of space and time, we expect the transformation relationships between the coordinates (x, y, z, t) and (x', y', z', t') to be linear; otherwise, there would not be a simple one-to-one relationship between events in S and S' frames. For instance, a nonlinear

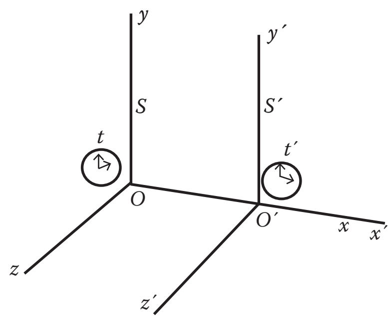


FIGURE 13.4 Relative motion of two coordinate systems.

transformation would predict acceleration in one system even if the velocity were constant in the other, obviously an unacceptable property for a transformation between inertial systems.

We consider the transverse dimensions first. Because the relative motion of the coordinate systems occurs only along the x-axis, we expect the linear relationships are of the forms $y' = k_1 y$ and $z' = k_2 z$. The symmetry requires that $y = k_1 y'$ and $z = k_2 z'$. These can both be true only if $k_1 = 1$ and $k_2 = 1$. Therefore, for the transverse direction, we have

$$y' = y \text{ and } z' = z. \tag{13.11}$$

These relationships are the same as in Galilean transformations.

For the longitudinal dimension, we expect that the relationship between x and x' must involve some change in the time coordinate, so we consider the most general linear relationship

$$x' = ax + bt. \tag{13.12}$$

Now the origin O', where $x' = 0$, corresponds to $x = ut$. Substituting these into Equation 13.12, we have

$$0 = aut + bt$$

from which we obtain

$$b = -au$$

and Equation 13.12 simplifies to

$$x' = a(x - ut). \tag{13.13}$$

By symmetry, we also have

$$x = a(x' + ut'). \tag{13.14}$$

Now we apply Einstein's second postulate of the constancy of the speed of light. If a pulse of light is sent out from the origin O of frame S at $t = 0$, its position along the x-axis later is given by $x = ct$, and its position along the x'-axis is $x' = ct'$. Putting these in Equations 13.13 and 13.14, we obtain

$$ct' = a(c - u)t \text{ and } ct = a(c + u)t'.$$

From these, we obtain

$$\frac{t}{t'} = \frac{c}{a(c-u)} \quad \text{and} \quad \frac{t}{t'} = \frac{a(c+u)}{c}.$$

Therefore,

$$\frac{c}{a(c-u)} = \frac{a(c+u)}{c}.$$

Solving for a,

$$a = \frac{1}{\sqrt{1-(u/c)^2}}.$$

Then,

$$b = -au = -\frac{u}{\sqrt{1-(u/c)^2}}.$$

Substituting these into Equations 13.13 and 13.14 gives

$$x' = \frac{x-ut}{\sqrt{1-\beta^2}} \tag{13.14a}$$

and

$$x = \frac{x'+ut'}{\sqrt{1-\beta^2}} \tag{13.14b}$$

where $\beta = u/c$. Eliminating either x or x' from Equations 13.14a and 13.14b, we obtain

$$t' = \frac{t-ux/c^2}{\sqrt{1-\beta^2}} \tag{13.14c}$$

and

$$t = \frac{t'+ux'/c^2}{\sqrt{1-\beta^2}}. \tag{13.14d}$$

Combining all of these results, we obtain the Lorentz transformations:

$$\left.\begin{array}{ll} x' = \gamma(x-ut) & x = \gamma(x'+ut) \\ y' = y & y = y' \\ z' = z & z = z' \\ t' = \gamma(t-ux/c^2) & t = \gamma(t'+ux/c^2) \end{array}\right\} \tag{13.15}$$

where

$$\gamma(=1/\sqrt{1-\beta^2}), \quad \beta = u/c \tag{13.16}$$

is the Lorentz factor.

If $\beta \ll 1$, $\gamma \cong 1$, then Equation 13.15 reduces to the Galilean transformations. Thus, the Galilean transformation is a first approximation to the Lorentz transformations for $\beta \ll 1$.

When the velocity, $\vec{u}$, of S' relative to S is in some arbitrary direction, Equation 13.15 can be given a more general form in terms of the components of $\vec{r}$ and $\vec{r}'$ perpendicular and parallel to $\vec{u}$:

$$\left.\begin{array}{ll} \vec{r}'_{\text{II}} = \gamma(\vec{r}_{\text{II}} - \vec{u}t) & \vec{r}_{\text{II}} = \gamma(\vec{r}'_{\text{II}} + \vec{u}t) \\ \vec{r}'_{\perp} = \vec{r}_{\perp} & \vec{r}'_{\perp} = \vec{r}_{\perp} \\ t' = \gamma(t - \vec{u}\cdot\vec{r}/c^2) & t = \gamma(t' + \vec{u}\cdot\vec{r}/c^2) \end{array}\right\}. \tag{13.17}$$

The Lorentz transformations are valid for all types of physical phenomena at all speeds. As a consequence of this, all physical laws must be invariant under a Lorentz transformation.

The Lorentz transformations that are based on Einstein's postulates contain a new philosophy of space and time measurements. We now examine the various properties of these new transformations. In the following discussion, we still use Figure 13.4.

13.4.1 Relativity of Simultaneity, Causality

Two events that happen at the same time but not necessarily at the same place are called "simultaneous." Now, consider two events in S' that occur at (x'_1, t'_1) and (x'_2, t'_2), and they would appear in frame S at (x_1, t_1) and (x_2, t_2). The Lorentz transformations give

$$t_2 - t_1 = \gamma\left[(t'_2 - t'_1) + \frac{u(x'_2 - x'_1)}{c^2}\right]. \tag{13.18}$$

It is easy to observe the following:

(1) If the two events take place simultaneously in S', then $t'_2 - t'_1 = 0$. But the events do not occur simultaneously in the S frame, for there is a finite time lapse:

$$\Delta t = t_2 - t_1 = \gamma \frac{u(x'_2 - x'_1)}{c^2}.$$

Thus, two spatially separated events that are simultaneous in S' would not be measured to be simultaneous in S. In other words, the simultaneity of spatially separated events is not an absolute property as it was assumed to be in Newtonian mechanics.

Moreover, depending on the sign of $(x'_2 - x'_1)$, the time interval Δt can be positive or negative, that is, in the frame S, the "first" event in S' can take place earlier or later than the "second" one. The sole exception is the case when two events occur coincidentally in S', then they also occur at the same place and at the same time in frame S.

(2) If the order of events in frame S is not reversed in time, then $\Delta t = t_2 - t_1 > 0$, which implies that

$$(t'_2 - t'_1) + \frac{u(x'_2 - x'_1)}{c^2} > 0$$

or

$$\frac{x_2' - x_1'}{t_2' - t_1'} < \frac{c^2}{u}$$

which will be true as long as

$$\frac{x_2' - x_1'}{t_2' - t_1'} < c.$$

Thus, the order of events will remain unchanged if no signal can be transmitted with a speed greater than c, the speed of light.

The preceding discussion also illustrates clearly that the theory of relativity is incompatible with the notion of action at a distance.

13.4.2 Time Dilation, Relativity of Co-Locality

Two events that happen at the same place but not necessarily at the same time are called co-local. Now consider two co-local events in S' taking place at times t_1' and t_2' but at the same place. For simplicity, consider this to be on the x'-axis so that $y' = z' = 0$. These two events would appear in frame S at (x_1, t_1) and (x_2, t_2). According to the Lorentz transformations, we have

$$\Delta x = \frac{u\Delta t'}{\sqrt{1-\beta^2}} = \gamma u \Delta t', \quad \Delta t = \frac{\Delta t'}{\sqrt{1-\beta^2}} = \gamma \Delta t' \tag{13.19}$$

where $\beta = u/c$, $\Delta t' = t_2' - t_1'$, and so forth. It is easy to observe the following:

(1) The two events that happened at the same place in S' do not occur at the same place in S, and so t_1 and t_2 must be measured by spatially separated synchronized clocks. Einstein's prescription for synchronizing two stationary separated clocks is to send a light signal from clock 1 at a time t_1 (measured by clock 1) and reflected back from clock 2 at a time t_2 (measured by clock 2). If the reflected light returns to clock 1 at a time t_3 (measured by clock 1), then clocks 1 and 2 are synchronous if $t_2 - t_1 = t_3 - t_2$, that is, if the time measured for light to go one way is equal to the time measured for light to go in the opposite direction.
(2) The time interval between two events taking place at the same point in an inertial reference is measured by a single clock at that point, and it is called the proper time interval between two events. In the second equation of Equation 13.19, $\Delta t' = t_2' - t_1'$ is the proper time interval between the events in S'. Because $\gamma \geq 1$, the time interval $\Delta t = t_2 - t_1$ in S is longer than $\Delta t'$; this is called time dilation, often described by the statement that "moving clocks run slow" (as seen by the stationary observer in S). This apparent asymmetry between S and S' in time is a result of the asymmetric nature of the time measurement.

From the above discussion, we see that proper time interval is the minimum time interval that any observer can measure between two events. Note that in Equation 13.19, $\Delta t'$ is the proper time interval.

In addition to the common types of clock with which we are all familiar, there are atomic and nuclear processes that can and are being used for measuring time intervals. Among them are the emission and absorption of radiation and the decay of subatomic particles. These particles are usually in motion, and a measurement of a time interval, particularly when $v \approx c$, will be greatly influenced by its velocity. One of these events is the decay of muons (μ) that has become a classical demonstration of time dilation. The muon is an unstable particle that spontaneously decays into an electron and two neutrinos:

$$\mu^{\pm} \rightarrow e^{\pm} + v + \overline{v}$$

where e^- stands for electron, e^+ for positron, v for neutrino, and $\overline{v}$ for antineutrino. The decay of the muon is typical of radioactive decay processes: if there are $N(0)$ muons at $t = 0$, the number at time t is

$$N(t) = N(0)e^{t/\tau}$$

where τ, the mean lifetime, is 2.15×10^{-6} s. Muons can be observed by stopping them in dense absorbers and detecting the decay electron, which comes off with an energy of approximately 40 MeV.

Muons are formed in abundance when high-energy cosmic ray protons enter the Earth's upper atmosphere. The protons lose energy rapidly, and at the altitude of a typical mountaintop, 2000 m, there are few left. But the muons penetrate far through the Earth's atmosphere and may reach the ground. The muons descend through the atmosphere with a velocity close to c. The minimum time to descend 2000 m is approximately

$$T = \frac{2 \times 10^3 \text{m}}{3 \times 10^5 \text{m/s}} = 7 \times 10^{-6} \text{s}.$$

This is more than 3 times the mean lifetime, τ, of the muon.

The experiment consists of comparing the flux of muons at the mountaintop with the flux at sea level:

$$\frac{\text{flux at sea level}}{\text{flux at mountaintop}} = e^{T/\tau'}.$$

Here τ' is different from τ, which is the mean lifetime to decay of a muon at rest. When the muon moves rapidly in the atmosphere, the lifetime τ' observed is increased by time dilation according to the relationship

$$\tau' = \frac{\tau}{\sqrt{1-\beta^2}} = \gamma\tau.$$

The measured flux ratio is 0.7. To account for this measured ratio, we require $\gamma = 10$. This was found to be the case: by measuring the energy of the muons, γ was determined to be 10 within the experimental error.

Time dilation between observers in uniform relative motion is not an artifact of the clock we choose to construct. It is a very real thing. All processes, including atomic and biological processes, slow down in moving systems.

13.4.3 Length Contraction

Consider a rod of length L_0 in the S' frame in which the rod lies at rest along the x' axis: $L_0 = x_2' - x_1'$. L_0 is the proper length of the rod measured in the rod's rest frame S'. Now, the rod is moving lengthwise with velocity u relative to the S frame. An observer in the S frame makes a *simultaneous* measurement of the two ends of the rod. The Lorentz transformations give

$$x_1' = \gamma[x_1 - ut_1], \quad x_2' = \gamma[x_2 - ut_2]$$

from which we get

$$x_2' - x_1' = \gamma[x_2 - x_1] - \gamma u(t_2 - t_1).$$

Let $L(u) = x_2 - x_1$, the length of the rod moving with speed u relative to the observer in S, and remember that $t_1 = t_2$, $\gamma = 1/(1 - \beta^2)^{1/2}$. The result then becomes

$$L(u) = \sqrt{1-\beta^2}\, L_0. \tag{13.20}$$

Thus, the length of a body moving with velocity u relative to an observer is measured to be shorter by a factor of $(1 - \beta^2)^{1/2}$ in the direction of motion relative to the observer.

Because all inertial frames of reference are equally valid, if $L' = \gamma L$, does not the expression $L = \gamma L'$ have to be equally true? The answer is no. The reason is that the measurement was not carried out in the same way in the two frames of reference. The two events of marking the positions of the two ends of the rod were simultaneous in the S frame but not simultaneous in the S' frame. This difference gives the asymmetry of the result. Length L' was equal to length L_0 only because the rod was at rest in the S' frame. As a general expression, $\Delta x' = \gamma\Delta x$ is not true. The full expression relating distances in two frames of reference is $\Delta x' = \gamma(\Delta x - u\Delta t)$. The symmetrical inverse relationship is $\Delta x = \gamma(\Delta x' + u\Delta t')$. In the case that was considered earlier, $\Delta t = 0$, so $\Delta x' = \gamma\Delta x$; but $\Delta t' \neq 0$, so $\Delta x \neq \Delta x'$.

A body of proper volume V_0 can be divided into thin rods parallel to u. Each one of these rods is reduced in length by a factor $(1 - \beta^2)^{1/2}$ so that the volume of the moving body measured by an observer in S is $V = (1 - \beta^2)^{1/2} V_0$.

Example 13.1

A ruler of length L_0 lies in the $x'y'$-plane of its rest system and makes an angle θ_0 with the x'-axis. What is the length and orientation of the ruler in the observer's system with respect to which the ruler moves to the right with velocity u?

Solution:

Let the ends of the ruler be designated by A and B. In the system S' in which the ruler lies, A and B have the following coordinates:

$$(x_A', y_A') = (0,0) \quad (x_B', y_B') = (L_0\cos\theta_0, L_0\sin\theta_0).$$

Using $x' = \gamma(x - ut)$ and $y' = y$, we obtain the coordinates of A and B in the observer's (S) system at time t:

$$x_A' = 0 = \gamma(x_A - ut), \quad y_A' = y_A$$

$$x_B' = L_0\cos\theta_0 = \gamma(x_B - ut), \quad y_B' = L_0\sin\theta_0 = y_B.$$

From these two equations, we find

$$x_B - x_A = \gamma^{-1} L_0 \cos\theta_0, \quad y_B - y_A = L_0 \sin\theta_0.$$

The length L of the ruler in the system S is

$$L = \sqrt{(x_B - x_A)^2 + (y_B - y_A)^2} = L_0\sqrt{1-\beta^2\cos^2\theta_0}$$

which indicates that the moving ruler is contracted. And the angle θ that the ruler makes with the x-axis is

$$\theta = \tan^{-1}\left(\frac{y_B - y_A}{x_B - x_A}\right) = \tan^{-1}\left(\gamma\tan\theta_0\right).$$

The moving ruler is both contracted and rotated.

Length contraction opens the possibility of space travel. The nearest star, besides the sun, is Alpha Centauri, which is about 4.3 light years away: light from Alpha Centauri takes 4.3 years to reach us. Even if a spaceship could travel at the speed of light, it would take 4.3 years to reach Alpha Centauri. This is certainly true from the point of view of an observer on Earth. But from the point of view of the crew of the spaceship, the distance between the Earth and Alpha Centauri is shortened by a factor $\gamma = (1 - \beta^2)^{1/2}$, where $\beta = v/c$ and v is the speed of the spaceship. If v is, say, $0.99c$, then $\gamma = 0.14$, and the distance appears to be only 14% of the value as seen from the Earth. The crew, therefore, deduces that light from Alpha Centauri takes only $0.14 \times 4.3 = 0.6$ year to reach Earth. But the crew sees Alpha Centauri coming toward it with a speed of $0.99c$ and expects to get there in $0.60/0.99 = 0.606$ year without having to suffer a long tedious journey. But, in practice, the power requirements to launch a spaceship near the speed of light are prohibitive.

13.4.4 Visual Apparent Shape of Rapidly Moving Object

An interesting consequence of the length contraction is the visual apparent shape of a rapidly moving object. This was shown first by Terrell (1959). The act of seeing involves the simultaneous reception of light from the different parts of the object. In order for light from different parts of an object to reach the eye or a camera at the same time, light from different parts of the object must be emitted at different times to compensate for the different distances the light must travel. Thus, taking a picture of a moving object or looking at it does not give a valid impression of its shape. Interestingly, the distortion that makes the Lorentz contraction seem to disappear instead makes an object seem to rotate by an angle $\theta = \sin^{-1}(u/c)$ as long as the angle subtended by the object at the camera is small. If the object moves in another direction or if the angle it subtends at the camera is not small, the apparent distortion becomes quite complex.

Figure 13.5 shows a cube of side l moving with a uniform velocity u with respect to an observer some distance away; the side AB is perpendicular to the line of sight of the observer. In order for light from corners A and D to reach the observer at the same instant, light from D, which must travel a distance l farther than from A, must have been emitted when D was at position E. The length DE is equal to $(l/c)u = l\beta$. The length of the side AB is foreshortened by the Lorentz contraction to $l\sqrt{1-\beta^2}$. The net result corresponds to the view that the observer would have if the cube were rotated through an angle $\sin^{-1}\beta$. The cube is not distorted; it undergoes an apparent rotation. Similarly, a moving sphere will not become an ellipsoid; it still appears as a sphere. An interesting discussion of apparent rotations at high velocity is given by Weisskopf (1960).

13.4.5 Relativistic Velocity Addition

Another very important kinematic consequence of the Lorenz transformation is that the Galilean transformation for velocity is no longer valid. The new and more complicated transformation for

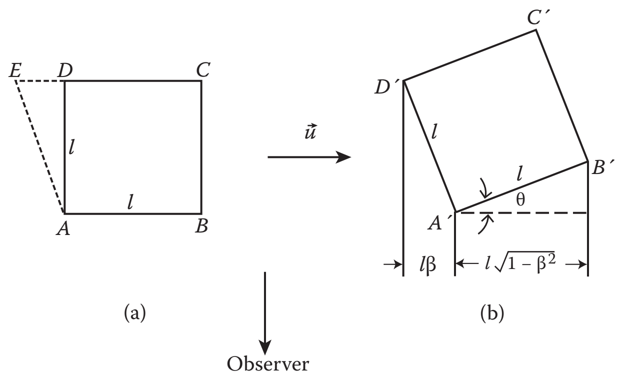


FIGURE 13.5 Visual apparent shape of a rapidly moving object. (a) A cub moving perpendicular to an observer's sight, (b) the observer view the cube rotated through an angle $\sin^{-1}\beta$.

velocities can be deduced easily. By definition, the components of velocity in S and S' are given by, respectively,

$$v_x = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{x_2 - x_1}{t_2 - t_1}$$

and

$$v_y = \frac{\mathrm{d}y}{\mathrm{d}t} = \frac{y_2 - y_1}{t_2 - t_1}$$

$$v'_x = \frac{\mathrm{d}x'}{\mathrm{d}t} = \frac{x'_2 - x'_1}{t'_2 - t'_1}$$

and

$$v'_y = \frac{\mathrm{d}y'}{\mathrm{d}t} = \frac{y'_2 - y'_1}{t'_2 - t'_1}$$

and so on.

Applying the Lorentz transformations to x_1 and x_2 and then taking the difference, we get

$$\mathrm{d}x = \frac{\mathrm{d}x' + u\mathrm{d}t'}{\sqrt{1-\beta^2}}, \quad \beta = \frac{u}{c}.$$

Similarly,

$$\mathrm{d}x' = \frac{\mathrm{d}x - u\mathrm{d}t}{\sqrt{1-\beta^2}}.$$

Do the same for the time intervals $\mathrm{d}t$ and $\mathrm{d}t'$:

$$\mathrm{d}t = \frac{\mathrm{d}t' + udx'/c^2}{\sqrt{1-\beta^2}}.$$

and

$$dt' = \frac{dt - u\,dx/c^2}{\sqrt{1-\beta^2}}.$$

From these, we obtain

$$\frac{dx}{dt} = \frac{dx' + u\,dt'}{dt' + u\,dx'/c^2}.$$

Dividing both the numerator and the denominator of the right-hand side by dt' yields the right transformation equation for the x component of the velocity:

$$v_x = \frac{v'_x + u}{1 + uv'_x/c^2}. \tag{13.21a}$$

Similarly, we can find the transverse components:

$$v_y = \frac{v'_y\sqrt{1-\beta^2}}{1 + uv'_x/c^2} = \frac{v'_y}{\gamma(1 + uv'_x/c^2)} \tag{13.21b}$$

$$v_z = \frac{v'_z\sqrt{1-\beta^2}}{1 + uv'_x/c^2} = \frac{v'_z}{\gamma(1 + uv'_x/c^2)}. \tag{13.21c}$$

In these formulas, $\gamma = (1 - \beta^2)^{-1/2}$ as before. We note that the transverse velocity components depend on the x-component. For $v \ll c$, we obtain the Galilean result $v_x = v'_x + u$.

Solving explicitly or merely switching the sign of u would yield (v'_x, v'_y, v'_z) in terms of (v_x, v_y, v_z).

It follows from the velocity transformation formulas that the value of an angle is relative and changes in transition from one reference frame to another. For an object in the S frame moving in the xy-plane with velocity v that makes an angle θ with the x-axis, we have

$$\tan\theta = v_y/v_x, \quad v_x = v\cos\theta, \quad v_y = v\sin\theta.$$

In the S' frame, we have

$$\tan\theta' = \frac{v'_y}{v'_x} = \frac{v\sin\theta}{\gamma(v\cos\theta - u)} \tag{13.22}$$

where

$$\gamma = 1/\sqrt{1-\beta^2} \text{ and } \beta = u/c.$$

As an application, consider the case of star light, that is, $v = c$; then,

$$\tan\theta' = \frac{\sin\theta}{\gamma(\cos\theta - u/c)}.$$

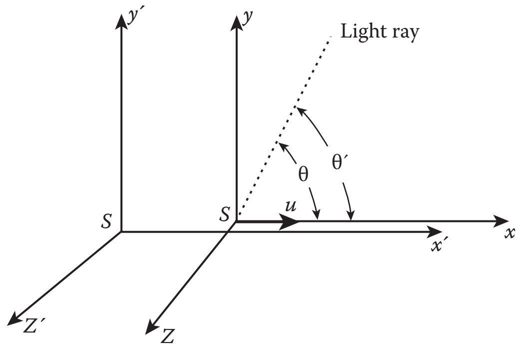


FIGURE 13.6 Aberration. The angles of a light ray with x-axis and x'-axis are different.

Let $\theta = \pi/2$, $\theta' = \pi/2 - \varphi$ (Figure 13.6); then,

$$\tan\phi = \frac{-u/c}{\sqrt{1-\beta^2}}, \quad \sin\phi = -\frac{u}{c}$$

which is the star aberration formula; to see a star overhead, tilt the telescope at angle φ.

Example 13.2: The Velocity of Light Is the Limiting Speed

A bullet is fired in the forward direction from a moving platform whose speed is u, as shown in Figure 13.7. The muzzle velocity of the bullet is v_1'. What is the velocity of the bullet relative to the ground?

Solution:

The S' system is with the moving platform, and the S system is attached to the ground. The muzzle velocity of the bullet v_1' is measured relative to the gun. The velocity of the bullet relative to the ground v_1 is

$$v_1 = \frac{v_1' + u}{1 + uv_1'/c^2}.$$

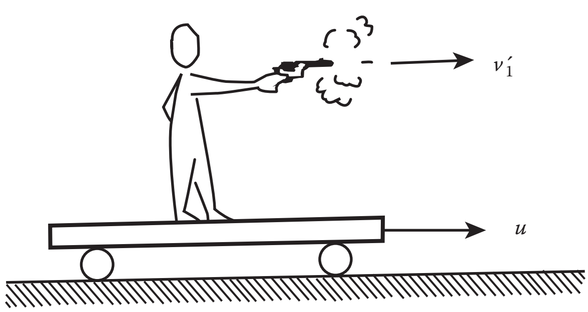


FIGURE 13.7 Velocity of light is the same for all observers.

Now, let us take a hypothetical case: $u = 0.9c$ and $v_1' = 0.9c$. Then,

$$v_1 = \frac{v_1' + u}{1 + uv_1'/c^2} = \frac{0.9c + 0.9c}{1 + (0.9c)(0.9c)/c^2} = 0.9945c < c$$

whereas, according to the Galilean transformation, $v_1 = v_1' + u = 1.8c$. In the limiting case, if we take both u and v_1' to be c, then

$$v_1 = \frac{v_1' + u}{1 + uv_1'/c^2} = \frac{c + c}{1 + c \cdot c/c^2} = c$$

which agrees with the postulate originally built into the Lorentz transformations: the velocity of light is the same for all observers. Thus, the relativistic transformation of velocities ensures that we cannot exceed the velocity of light by changing reference frames.

13.5 DOPPLER EFFECT

The Doppler effect occurs for light and sound. It is a shift in frequency resulting from the motion of the source or the observer. Knowledge of the motion of distant receding galaxies comes from studies of the Doppler shift of their spectral lines. The Doppler effect is also used for satellite tracking and radar speed traps. We examine the Doppler effect in light only.

Consider a source of light or radio waves moving with respect to an observer or a receiver, at a speed u and at an angle θ with respect to the line between the source and the observer (Figure 13.8). The light source flashes with a period τ_0 in its rest frame (the S' frame in which the source is at rest). The corresponding frequency is $v_0 = 1/\tau_0$, and the wavelength is $\lambda_0 = c/v_0 = c\tau_0$.

While the source is going through one oscillation, the time that elapses in the rest frame of the observer (the S frame) is $\tau = \gamma\tau_0$ because of time dilation, where $\gamma = (1 - \beta^2)^{1/2}$ and $\beta = u/c$. The emitted wave travels at speed c, and therefore, its front moves a distance $\gamma\tau_0 c$; the source moves toward the observer with a speed $u\cos\theta$, so a distance $\gamma\tau_0 u\cos\theta$. Then, the distance D that separates the fronts of the successive waves (the wavelength) is

$$D = \gamma\tau_0 c - \gamma\tau_0 u\cos\theta,$$

that is,

$$\lambda = \gamma\tau_0 c - \gamma\tau_0 u\cos\theta = \gamma\tau_0 c[1 - (u/c)\cos\theta],$$

but $c\tau_0 = \lambda_0$; we can rewrite the last expression as

$$\lambda = \lambda_0 \frac{1 - \beta\cos\theta}{\sqrt{1 - \beta^2}}. \tag{13.23}$$

In terms of frequency, this Doppler effect formula becomes

FIGURE 13.8 Doppler effect.

$$\nu = \nu_0 \frac{\sqrt{1-\beta^2}}{1-\beta\cos\theta}. \tag{13.24}$$

Here, ν is the frequency at the observer, and θ is the angle measured in the rest frame of the observer. If the source is moving directly toward the observer, then $\theta = 0$ and $\cos\theta = 1$. Equation 13.24 reduces to

$$\nu = \nu_0 \frac{\sqrt{1-\beta^2}}{1-\beta} = \nu_0\sqrt{\frac{1+\beta}{1-\beta}}. \tag{13.24a}$$

For a source moving directly away from the observer, $\cos\theta = -1$, Equation 13.24 reduces to

$$\nu = \nu_0 \frac{\sqrt{1-\beta^2}}{1+\beta} = \nu_0\sqrt{\frac{1-\beta}{1+\beta}}. \tag{13.24b}$$

At $\theta = \pi/2$, that is, the source moving at right angles to the direction toward the observer, Equation 13.24 reduces to

$$\nu = \nu_0\sqrt{1-\beta^2}. \tag{13.24c}$$

This transverse Doppler effect is a result of time dilation.

13.6 RELATIVISTIC SPACE–TIME (MINKOWSKI SPACE)

As we have seen previously, there is no absolute standard for the measurement of time or of space; the relative motion of observers affects both kinds of measurement. Lorentz transformations treat (x, y, z) and t as equivalent variables. In 1907, Hermann Minkowski proposed that the three dimensions of space and the dimension of time should be treated together as four dimensions of space–time. Minkowski said, "Henceforth space by itself and time by itself, are doomed to fade away into mere shadows, and only a kind of union of the two will preserve an independent reality." He called the four dimensions of space–time "the world space" (also known as the Minkowski space); an event is a point in a space–time diagram, and the path of an individual particle is a "world line" as shown in Figure 13.9. Clearly, we cannot draw in four dimensions, but we can draw easily a diagram in which we measure time in vertical direction and distance in horizontal direction. In Figure 13.9, dots are events. Line *A* represents a stationary particle, line *B* a particle moving with constant velocity, and line *C* an accelerating particle (starting from rest). The space–time diagram is a valuable aid in visualizing many relativity problems. We will describe one such graph in the last section as it is not our main concern.

By analogy, with the three-dimensional case, the coordinates of an event (ct, x, y, z) can be considered as the components of a four-dimensional radius vector, or a radius four-vector (for short) in a four-dimensional Minkowski space. The square of the length of the radius four-vector for any event E is a variable quantity:

$$s^2 = c^2t^2 - (x^2 + y^2 + z^2). \tag{13.25}$$

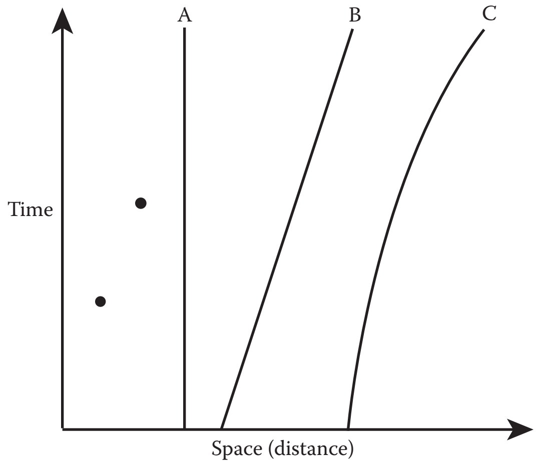


FIGURE 13.9 In the Minkowski space, the path of an individual particle is a world line.

This quadratic expression is commonly known as an interval, and we denote it by s^2; it is an invariant, independent of the frame used to measure these coordinates. To see it, let us calculate the value of $x^2 - c^2t^2$ (for simplicity) in S frame in terms of x' and t' in S' frame:

$$\begin{aligned} x^2 - c^2t^2 &= \gamma^2(x' + ut')^2 - c^2\gamma^2(t' + ux/c^2)^2 \\ &= -\gamma^2\left\{(c^2 - u^2)t'^2 + 2ux't' - 2ux't' + x'^2(u^2/c^2 - 1)\right\} \\ &= x'^2 - c^2t'^2. \end{aligned}$$

As an interval is an invariant quantity, we may use it to classify the possible events in space–time. In particular, we may divide events into three types: those for which the invariant label is positive, zero, or negative. Events for which the interval is positive are called time-like with respect to the origin because they are of the class containing those with $x = 0$ and $t \neq 0$, which are just changes in time of a clock at the origin of space. Those for which the invariant is negative are called space-like with respect to the origin because that includes events with $t = 0$ and $x \neq 0$, which are just spatially separated but simultaneous with an event at the origin of space–time. The final class of events with the interval being zero is called light-like with respect to the origin because a ray of light can pass to or from the origin of space–time to them. This division of space–time into three regions is shown in Figure 13.10, the light cone of a two-dimensional space and a one-dimensional time continuum:

(1) $c^2t^2 - x^2 > 0$: time-like with respect to the origin
(2) $c^2t^2 - x^2 < 0$: space-like with respect to the origin
(3) $c^2t^2 - x^2 = 0$: light-like with respect to the origin

Such a classification of events can also be considered from the point of view of causality.

It is now a common practice to treat t as a zeroth or a fourth coordinate:

$$x^0 = ct,\ x^1 = x,\ x^2 = y,\ x^3 = z$$

or

$$x^1 = x,\ x^2 = y,\ x^3 = z,\ x^4 = ix^0.$$

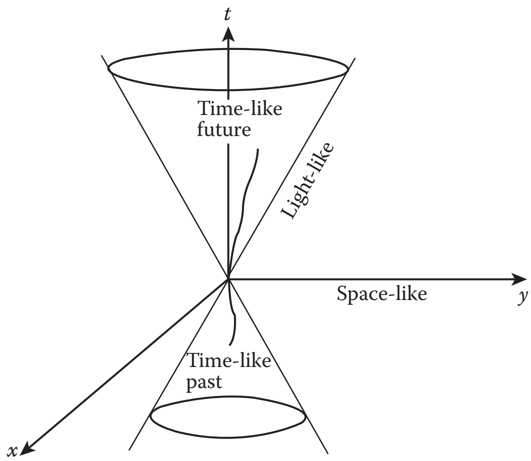


FIGURE 13.10 Space–time diagram of a three-dimensional world, showing the light cone.

Then the Lorentz transformations take on the form

$$\begin{aligned} x^{1\prime} &= \gamma(x^1 + i\beta x^4) \\ x^{2\prime} &= x^2 \\ x^{3\prime} &= x^3 \\ x^{4\prime} &= \gamma(-i\beta x^1 + x^4) \end{aligned}$$

or

$$\begin{aligned} x^{0\prime} &= \gamma(x^0 - \beta x^1) \\ x^{1\prime} &= \gamma(-\beta x^0 + x^1) \\ x^{2\prime} &= x^2 \\ x^{3\prime} &= x^3. \end{aligned} \tag{13.26}$$

In matrix form, we have

$$\begin{pmatrix} x^{1\prime} \\ x^{2\prime} \\ x^{3\prime} \\ x^{4\prime} \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x^1 \\ x^2 \\ x^3 \\ x^4 \end{pmatrix}$$

or

$$\begin{pmatrix} x^{0\prime} \\ x^{1\prime} \\ x^{2\prime} \\ x^{3\prime} \end{pmatrix} = \begin{pmatrix} \gamma & -\beta\gamma & 0 & 0 \\ -\beta\gamma & \gamma & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{13.27}$$

It is customary to use Greek indices (μ and ν, etc.) to label four-dimensional variables and Latin indices (i and j, etc.) to label three-dimensional variables.

The Lorentz transformations can be distilled into a single equation:

$$x^{\mu\prime} = \sum_{\nu=1}^{4} L^{\mu}_{\nu} x^{\nu} = L^{\mu}_{\nu} x^{\nu}\mu, \quad \nu = 1,2,3,4 \tag{13.28}$$

where L^{μ}_{ν} is the Lorentz transformation matrix in Equation 13.27. The summation sign is eliminated by Einstein's summation convention: the repeated indexes appearing once in the lower and once in the upper position are automatically summed over. However, the indexes repeated in the lower part or upper part alone are not summed over.

If Equation 13.28 reminds us of the orthogonal rotations, it is no accident. The general Lorentz transformations can indeed be interpreted as an orthogonal rotation of axes in Minkowski space. Note that the *xt*-submatrix of the Lorentz matrix in Equation 13.27 is

$$\begin{pmatrix} \gamma & i\beta\,\gamma \\ -i\beta\,\gamma & \gamma \end{pmatrix}$$

which is to be compared with the *xy*-submatrix of the two-dimensional rotation about the *z*-axis:

$$\begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Upon identification of matrix elements $\cos\theta = \gamma$ and $\sin\theta = i\beta\gamma$, we see that the rotation angle θ (for the rotation in the *xt*-plane) is purely imaginary.

Some books prefer to use a real angle of rotation φ, defining $\varphi = -i\theta$. Then note that

$$\cos\theta = \frac{e^{i\theta} + e^{-i\theta}}{2} = \frac{e^{-\phi} + e^{\phi}}{2} = \cosh\phi$$

$$\sin\theta = \frac{e^{i\theta} - e^{-i\theta}}{2i} = \frac{i[e^{\phi} - e^{-\phi}]}{2} = \sinh\phi$$

and the submatrix becomes

$$\begin{pmatrix} \cosh\phi & i\sinh\phi \\ -i\sinh\phi & \cosh\phi \end{pmatrix}.$$

We should note that the mathematical form of Minkowski space looks exactly like a Euclidean space; however, it is not physically so because of its complex nature as compared to the real nature of the Euclidean space.

13.6.1 Four-Velocity and Four-Acceleration

How do we define four vectors of velocity and acceleration? It is evident that the set of the four quantities $\mathrm{d}^{\mu}_{x}/\mathrm{d}t$ does not have the properties of a four-vector because d*t* is not an invariant. But we know that the proper time $\mathrm{d}\tau$ is an invariant. Although observers in different frames may disagree about the time interval between events because each is using his or her own time axis, all agree on the value of the time interval that would be observed in the frame moving with the particle. The components of the four-velocity are therefore defined as

$$u^{\mu} = \frac{dx^{\mu}}{d\tau}. \tag{13.29}$$

The second equation of Equation 13.19 relates the proper time $d\tau$ (was dt' there) to the time dt read by a clock in frame S relative to which the object (S' frame) moves at a constant u:

$$d\tau = dt\sqrt{1-\beta^2}.$$

We can rewrite u^{μ} completely in terms of quantities observed in frame S as

$$u^{\mu} = \gamma\frac{dx^{\mu}}{dt} = \frac{1}{\sqrt{1-\beta^2}}\frac{dx^{\mu}}{dt} \tag{13.30}$$

where

$$\gamma = 1/\sqrt{1-\beta^2}.$$

In terms of the ordinary velocity components v_1, v_2, v_3, we have

$$u^{\mu} = (\gamma c, \gamma v_i),\ i = 1, 2, 3. \tag{13.31}$$

The length of four-velocity must be invariant. We can verify this easily:

$$\sum_{\mu=1}^{4}(u^{\mu})^2 = -c^2. \tag{13.32}$$

Similarly, a four-acceleration is defined as

$$w^{\mu} = \frac{d^2x^{\mu}}{d\tau^2} = \frac{du^{\mu}}{d\tau}. \tag{13.33}$$

Now differentiating Equation 13.32 with respect to τ, we obtain

$$w^{\mu}u^{\mu} = 0. \tag{13.34}$$

That is, the four vectors of velocity and acceleration are mutually perpendicular.

13.6.2 Four-Energy and Four-Momentum Vectors

Now, it is obvious that Newtonian dynamics cannot hold totally. How do we know what to retain and what to discard? The answer is found in the generalizations that grew from the laws of motion but transcend it in their universality. These are the conservation of momentum and energy.

Thus, we now generalize the definitions of momentum and energy so that, in the absence of external forces, the momentum and energy of a system of particles are conserved. In Newtonian mechanics, the momentum $\vec{p}$ of a particle is defined as $m\vec{v}$, the product of the particle's inertial mass and its velocity. A plausible generalization of this definition is to use the four-velocity u^μ and an invariant scalar m_0 that truly characterize the inertial mass of the particle and define the momentum four-vector (four-momentum, for short) P^μ as

$$P^\mu = m_0 u^\mu. \tag{13.35}$$

To ensure that the "mass" of the particle is truly a characteristic of the particle, this mass must be that measured in the frame of reference in which the particle is at rest. Thus, the mass of the particle must be its proper mass. We customarily call this mass the rest mass of the particle and denote it by m_0. Using Equation 13.31, we write P^μ in component form:

$$\begin{aligned} P^j &= \frac{m_o v_j}{\sqrt{1-\beta^2}} = \gamma m_0 v_j, \quad j = 1,2,3 \\ P^0 &= \frac{m_o c}{\sqrt{1-\beta^2}} = \gamma m_0 c. \end{aligned} \tag{13.36}$$

We see that as $\beta = v/c \to 0$, the first three components of the four-momentum P^μ reduce to $m_o v_j$, the components of the ordinary momentum. This indicates that Equation 13.35 appears to be a reasonable generalization.

Let us write the time component of P^0 as

$$P^0 = \frac{m_0 c}{\sqrt{1-\beta^2}} = \frac{E}{c}. \tag{13.37}$$

Now, what is the meaning of the quantity E? For low velocities, the quantity E reduces to

$$E = \frac{m_0 c^2}{\sqrt{1-\beta^2}} \cong m_0 c^2 + \frac{1}{2} m_0 v^2.$$

The second term on the right-hand side is the ordinary kinetic energy of the particle; the first term can be interpreted as the rest energy of the particle (it is an energy the particle has even when it is at rest), which must contain all forms of internal energy of the object, including heat energy, internal potential energy of various kinds, or rotational energy if any. Hence, we can call the quantity E the total energy of the particle (moving at speed v).

We now write the four-momentum as

$$P^\mu = \left(\frac{E}{c}, P^j\right). \tag{13.38}$$

The length of the four-momentum must be invariant, just as the length of the velocity four-vector is invariant under the Lorentz transformations. We can show this easily:

$$\sum_\mu P^\mu P^\mu = \sum_\mu (m_0 u^\mu)(m_0 u^\mu) = -\, m_0^2 c^2. \tag{13.39}$$

But Equation 13.38 gives

$$\sum_{\mu} P^{\mu} P^{\mu} = P^2 - \frac{E^2}{c^2}.$$

Combining this with Equation 13.39, we arrive at the relationship

$$P^2 - \frac{E^2}{c^2} = -m_0^2 c^2$$

or

$$E^2 - P^2 c^2 = m_0^2 c^4. \tag{13.40}$$

The total energy E and the momentum P^{μ} of a moving body are different when measured with respect to different reference frames. But the combination $P^2 - E^2/c^2$ has the same value for all frames of reference, namely, $m_0^2 c^2$. This relationship is very useful. Another very useful relationship is $\vec{P} = \vec{v}(E/c^2)$. From Equation 13.35, we see that $\gamma m_0 = E/c^2$; combining this with the first equation of Equation 13.36 gives the very useful relationship $\vec{P} = \vec{v}(E/c^2)$.

The relativistic momentum, however, is not quite the familiar form found in general physics because its spatial components contain the Lorentz factor γ. We can bring it into the old sense, and the traditional practice was to introduce a "relativistic mass" m:

$$m = m_0 \gamma = \frac{m_0}{\sqrt{1-\beta^2}}.$$

With this introduction of m, P^j takes the old form: $P^j = mv_j$. However, some feel that the introduction of the relativistic mass is not a purely methodological issue; it often causes misunderstanding and vague interpretations of relativistic mechanics. So they prefer to include the factor γ with v_j forming the proper four-velocity component u_j and treating the mass as simply the invariant parameter m_0. For details, see Okun (1989).

13.6.3 Particles of Zero Rest Mass

A surprising consequence of the relativistic energy–momentum generalization is the possibility of "massless" particles, which possess momentum and energy but no rest mass. From the expression for the energy and momentum of a particle,

$$E = \frac{m_0 c^2}{\sqrt{1 - v^2/c^2}}, \quad \vec{P} = \frac{m_0 \vec{v}}{\sqrt{1 - v^2/c^2}}. \tag{13.41}$$

We can define a particle of zero rest mass possessing finite energy and momentum. To this purpose, we allow $v \to c$ in some inertial system S and $m_0 \to 0$ in such a way that

$$\frac{m_0}{\sqrt{1 - v^2/c^2}} = \chi \tag{13.42}$$

remains constant. Then, Equation 13.41 takes the simple form

$$E = \chi c^2 \quad \vec{P} = \chi c \hat{e}$$

where $\hat{e}$ is a unit vector in the direction of motion of the particle. Eliminating χ from the last two equations, we obtain

$$E = Pc \tag{13.43}$$

which is consistent with Equation 13.40: $E^2 - P^2 \cdot c^2 = m_0^2 c^4$.

Now as $(E/c, \vec{P})$ is a four-vector, $(\chi c, \chi c\hat{e})$ is also a four-vector, the energy and momentum four-vector of a zero rest-mass particle in frame S and in any other inertial frame such as S'. It can be shown that the transformation of the energy and momentum four-vector $(\chi c, \chi c\hat{e})$ of a zero rest-mass particle is identical with that of a light wave, provided χ is made proportional to the frequency ν. Thus, if we associate a zero rest-mass particle with a light wave in one inertial frame, this association holds in all other inertial frames. The ratio of the energy of the particle to the frequency has the dimensions of action (or angular momentum). This suggests that we can write this association by the following equations:

$$E = h\nu \text{ and } P = \chi c = h\nu/c$$

where h is Planck's constant. This massless particle of light is called a photon. Einstein introduced it in his pioneering paper on the photoelectric effect published a few months before his work on special relativity out of concern with the photoelectric effect and consideration of Planck's quantum hypothesis.

13.7 EQUIVALENCE OF MASS AND ENERGY

We have learned that Einstein's theory of special relativity drastically revised our concepts about space and time. So we must follow Einstein and rethink the notions of mass, energy, and other important quantities. The equivalence of mass and energy is the best-known relationship Einstein gave in his theory of special relativity in 1905:

$$E = mc^2 \tag{13.44}$$

where E is the energy, m is the mass, and c is the speed of light.

We can get this general idea of the equivalence of mass and energy from the consideration of electromagnetic theory. An electromagnetic field possesses energy E and momentum p, and there is a simple relationship between E and p:

$$P = E/c.$$

Thus, if an object emits light in one direction with momentum p, in order to conserve momentum, the object itself must recoil with a momentum $-p$. If we stick to the definition of momentum as $p = mv$, we may associate a "mass" with a flash of light:

$$m = \frac{p}{v} = \frac{p}{c} = \frac{E}{c^2}$$

which leads to the famous formula

$$E = mc^2.$$

This mass is not merely a mathematical fiction. Let us consider a simple thought experiment. Refer to Figure 13.11; there is an observer S in S frame and an observer S' and an atom at rest in the

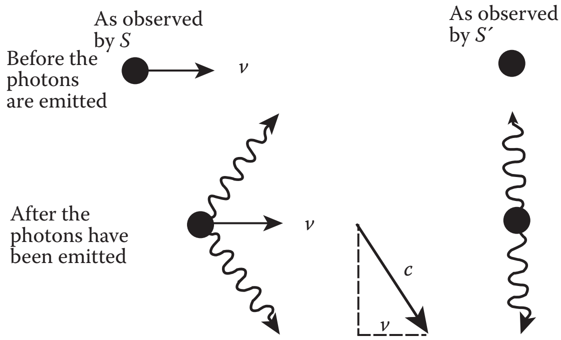


FIGURE 13.11 Thought experiment to show the "equivalence" mass and energy.

S' frame. The atom emits two flashes of light (photons) of equal energy, traveling back-to-back and normal to the direction of the frames' relative motion. Figure 13.8 depicts what S and S' observe. Before the lights are emitted, the atom is at rest in S' frame and moves rightward with speed v as observed by observer S in the S system.

We now analyze the emission process, first from the point of view of S' and then according to the point of view of S. In S' frame, the sum of the momentum of the two lights is zero, and the atom remains at rest after the emission of lights. As observed by observer S, the situation is quite different. The two flashes of light move along diagonal directions. Each light's component of velocity parallel to the atom's velocity is v. Thus, the momentum of the two flashes of light parallel to the atom's velocity is

$$2\frac{E}{c}\frac{v}{c}.$$

We next study energy and momentum changes for the atom as observed by S. The atom's energy is decreased by an amount equal to $-2E$:

$$\Delta E_{\text{atom}} = 2E$$

and its momentum change is

$$\Delta p_{\text{atom}} = -2\frac{E}{c}\frac{v}{c} = \Delta E_{\text{atom}}\frac{v}{c^2}.$$

Now, the definition of momentum for the atom ($p_{\text{atom}} = m_{\text{atm}}\, v$) implies that

$$\Delta p_{\text{atom}} = \Delta m_{\text{atom}} \cdot v.$$

Comparing the two expressions for the change in the atom's momentum, we obtain

$$\Delta E_{\text{atom}}\frac{v}{c^2} = \Delta m_{\text{atom}} \cdot v$$

which gives

$$\Delta E_{\text{atom}} = \Delta m_{\text{atom}} \cdot c^2.$$

In short,

$$\Delta E = \Delta m \cdot c^2.$$

Einstein also provided a thought experiment some time ago. An emitter and an absorber of light are firmly attached to the ends of a box of mass M and length L (Figure 13.12). The box is initially stationary but is free to move. If the emitter sends a short light pulse of energy ΔE and momentum $\Delta E/c$ toward the right, the box will recoil toward the left by a small distance Δx with momentum $p_x = -\Delta E/c$ and velocity v_x, where v_x is given by

$$v_x = p_x/M = -\ \Delta E/cM.$$

The light pulse reaches the right end of the box approximately in time $\Delta t = L/c$ and is absorbed. The small recoil distance is then given by

$$\Delta x = v_x \Delta t = -\ \Delta EL/Mc^2.$$

Now, the center of mass of the system cannot move by purely internal changes, and there are no external forces. It must be that the transport of energy ΔE from the left end of the box to the right end is accompanied by the transport of mass Δm, so the change in the position of the center of mass of the box (denoted by δx) vanishes. The condition for this is

$$\delta x = 0 = \Delta mL + M\Delta x.$$

From this, we find

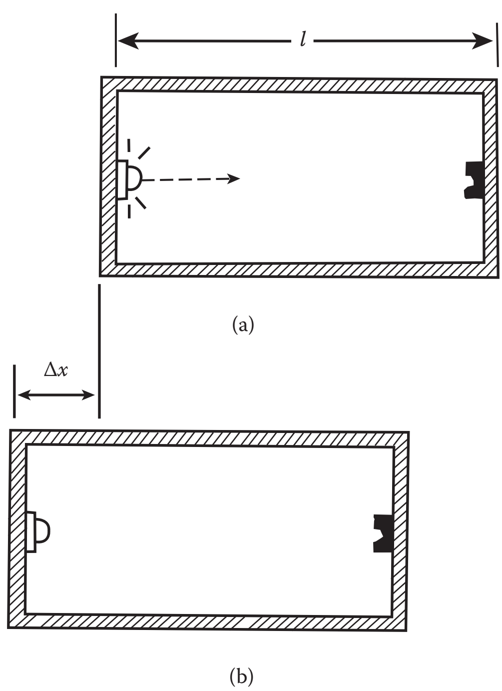


FIGURE 13.12 Einstein's thought experiment. (a) A box with an emitter, (b) after the emitter sends out a light pulse, the box recoiled to the left by a distance x.

$$\Delta m = -\frac{M}{L}\Delta x = \frac{M}{L}\frac{\Delta EL}{Mc^2} = \Delta E/c^2$$

or

$$\Delta E = \Delta m \cdot c^2.$$

We should not confuse the notions of equivalence and identity. The energy and mass are different physical characteristics of particles; here, the word "equivalence" established only their proportionality to each other. This is similar to the relationship between the gravitational mass and inertial mass of a body: the two masses are indissolubly connected with each other and proportional to each other but, at the same time, have different characteristics. The equivalence of mass and energy has been beautifully verified by experiments in which matter is annihilated and converted totally into energy. For example, when an electron and a positron, each with a rest mass m_0, come together, they disintegrate, and two gamma rays emerge, each with the measured energy of m_0c^2.

Based on Einstein's mass–energy relationship $E = mc^2$, we can show that the mass of a particle depends on its velocity. Let a force F act on a particle of momentum mv. Then,

$$F\mathrm{d}t = \mathrm{d}(mv). \tag{13.45}$$

If there is no loss of energy by radiation resulting from acceleration, then the amount of energy transferred in dt is

$$\mathrm{d}E = c^2\mathrm{d}m.$$

This is put equal to the work done by the force F to give

$$Fv\mathrm{d}t = c^2\mathrm{d}m.$$

Combining this with Equation 13.45, we have

$$v\mathrm{d}(mv) = c^2\mathrm{d}m.$$

Multiply this by m.

$$vm\mathrm{d}(mv) = c^2m\mathrm{d}m.$$

Integrating, we have

$$(mv)^2 = c^2m^2 + K$$

where K is the integration constant. Now, $m = m_0$ when $v = 0$; we find $K = -c^2m_0^2$, and

$$m^2v^2 = c^2\left(m^2 - m_0^2\right).$$

m_0 is known as the rest mass of the particle. Solving for m, we obtain the so-called relativistic mass of the particle:

$$m = \frac{m_0}{\sqrt{1-(v/c)^2}}. \tag{13.46}$$

It is now easy to see that a material body cannot have a velocity greater than the velocity of light. If we try to accelerate the body, as its velocity approaches the velocity of light, its mass becomes larger and larger as it becomes increasingly more difficult to accelerate it further. In fact, because the mass m becomes infinite when $v = c$, we can never accelerate the body up to the speed of light.

Example 13.3: Variation of Mass with Velocity

For the sake of simplicity, we shall consider the case of a central inelastic collision between two particles of equal mass whose velocities, as viewed in the S' system, are w and $-w$ and are parallel to the x-axis. The center of mass of this two-particle system will remain fixed; that is, its velocity will be zero at all times, including the instant of collision. In the S system, the masses of the two particles will not be the same; let us call the masses m_1 and m_2, and their velocities u_2 and $-u_2$, respectively. At the instant of collision, the two particles are at rest in the S' system but have a common velocity v in the S system, the velocity of S' relative to S (Figure 13.13). Applying the conservation of momentum in the S system, we obtain

$$m_1u_1 + (-m_2u_2) = (m_1 + m_2)v.$$

Solving for m_1/m_2,

$$m_1/m_2 = (1 + u_2/v)/(1 - u_1/v).$$

Now,

$$u_1 = (v + w)/(1 + vw/c^2), \quad u_2 = (-v + w)/(1 - vw/c^2).$$

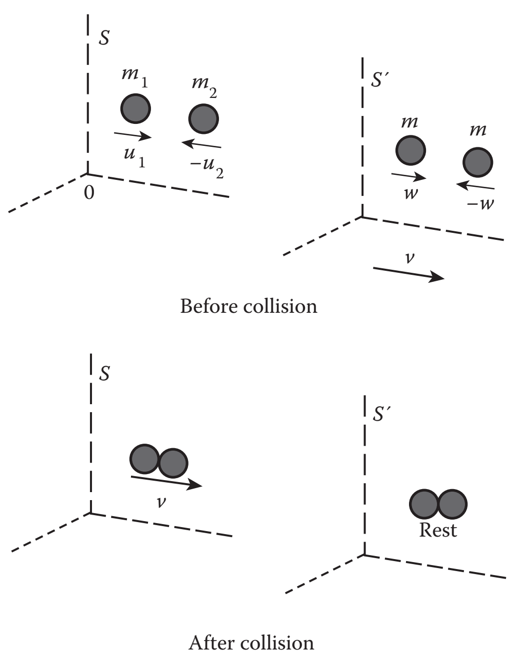


FIGURE 13.13 Collision of two particles in frame S and S'.

Substituting u_1 and u_2 into the expression for m_1/m_2, we obtain

$$\frac{m_1}{m_2} = \frac{1 + vw/c^2}{1 - vw/c^2}.$$

What is desired is a relationship between the mass of a particle and its velocity in the same S system. This can be done by expressing vw/c^2 in terms of u_1 and u_2. Let us first compute $(1 - u_1^2/c^2)$ and $(1 - u_2^2/c^2)$:

$$1 - \frac{u_1^2}{c^2} = 1 - \frac{1}{c^2}\left(\frac{v + w}{1 + vw/c^2}\right)^2 = \frac{(1 - v^2/c^2)(1 - w^2/c^2)}{(1 + vw/c^2)^2}.$$

Similarly, we have

$$1 - \frac{u_2^2}{c^2} = \frac{(1 - v^2/c^2)(1 - w^2/c^2)}{(1 - vw/c^2)^2}.$$

From these two equations, we find

$$\frac{1 - u_2^2/c^2}{1 - u_1^2/c^2} = \frac{(1 + vw/c^2)^2}{(1 - vw/c^2)^2}$$

or

$$\frac{1 + vw/c^2}{1 - vw/c^2} = \frac{\left(1 - u_2^2/c^2\right)^{1/2}}{\left(1 - u_1^2/c^2\right)^{1/2}}.$$

Accordingly, the ratio m_1/m_2 becomes

$$\frac{m_1}{m_2} = \frac{1 + vw/c^2}{1 - vw/c^2} = \frac{\left(1 - u_2^2/c^2\right)^{1/2}}{\left(1 - u_1^2/c^2\right)^{1/2}}.$$

Suppose we now consider the special case of a collision between these two particles when one of them, say, particle 2, has a velocity $u_2 = 0$. Then the last equation reduces to

$$m_1 = \frac{m_0}{\sqrt{1 - (u_1/c)^2}}$$

where m_0 is the mass of particle 2 when it is at rest. It is also the rest mass of particle 1 when it is at rest. Therefore, in general, if m_0 is the rest mass of a particle, its mass m when its velocity is v will be given by

$$m = \frac{m_0}{\sqrt{1 - (v/c)^2}}. \tag{13.47}$$

High-energy physicists do not use relativistic mass, and they prefer the quantity momentum as it is measurable.

13.8 CONSERVATION LAWS OF ENERGY AND MOMENTUM

It is now clear that the linear momentum and energy of a particle should not be regarded as different entities but simply as two aspects of the same attributes of the particle because they appear as separate components of the same four-vector P^μ, which transforms according to Equation 13.28:

$$P'^\mu = L^\mu_\nu P^\nu$$

in matrix form

$$\begin{pmatrix} P^{1\prime} \\ P^{2\prime} \\ P^{3\prime} \\ P^{4\prime} \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & i\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -i\beta\gamma & \gamma & 0 & 0 \end{pmatrix} \begin{pmatrix} P^1 \\ P^2 \\ P^3 \\ P^4 \end{pmatrix} = \begin{pmatrix} \gamma(P^1 + i\beta P^4 \\ P^2 \\ P^3 \\ \gamma(-i\beta P^1 + P^4) \end{pmatrix}.$$

Thus,

$$P^{1\prime} = \gamma(P^1 + i\beta P^4) = \gamma(P^1 - \beta P^0),$$

$$P^{2\prime} = P^2,$$

$$P^{3\prime} = P^3 \tag{13.48a}$$

$$P^{4\prime} = \gamma(-i\beta P^1 + P^4) \quad \text{or} \quad P^{0\prime}\gamma(P^0 - \beta P^1) \tag{13.48b}$$

which show clearly that what appears as energy in one frame appears as momentum in another frame, and vice versa.

So far, we have not discussed explicitly the conservation laws. Because linear momentum and energy are not regarded as different entities but as two aspects of the same attributes of an object, it is no longer adequate to consider linear momentum and energy separately. A natural relativistic generalization of the conservation laws of momentum and energy would be the conservation of the four-momentum. Consequently, the conservation of energy becomes one part of the law of conservation of four-momentum. This is exactly what has been found to be correct experimentally, and in addition, this generalized conservation law of four-momentum holds for a system of particles even when the number of particles and their rest energies are different in the initial and final states. It should be emphasized that what we mean by energy E is the total energy of an object. It consists of rest energy, which contains all forms of internal energy of the body, and kinetic energy. The rest energies and kinetic energies need not be individually conserved, but their sum must be. For example, in an inelastic collision, kinetic energy may be converted into some form of internal energy or vice versa; accordingly, the rest energy of the object may change.

Energy and momentum conservation go together in special relativity; we cannot have one without the other. This seems a bit puzzling for some readers for in classical mechanics, the conservation laws of energy and momentum are on different footing. That is because energy and momentum are regarded as different entities. Moreover, classical mechanics does not talk about rest energy at all.

13.9 GENERALIZATION OF NEWTON'S EQUATION OF MOTION

A natural relativistic generalization of the Newton's equations of motion follows:

$$K^\mu = \frac{dP^\mu}{d\tau} = \frac{d}{d\tau}(m_0 u^\mu) = m_0 \frac{d^2 x^\mu}{d\tau^2} \tag{13.49}$$

where K^μ is the four-force vector; it is also called the Minkowski force.

Using Equations 13.19 and 13.36, we obtain the three components for μ = 1, 2, 3:

$$K^j\sqrt{1-\beta^2} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{m_0 v_j}{\sqrt{1-\beta^2}} = F_j \tag{13.50}$$

and $K^4 = iK^0$, and K^0 is given by

$$K^0 = \frac{\mathrm{d}P^0}{\mathrm{d}\tau} = \frac{\gamma}{c}\frac{\mathrm{d}E}{\mathrm{d}\tau} = \frac{1}{c\sqrt{1-\beta^2}}\frac{\mathrm{d}E}{\mathrm{d}t}. \tag{13.51}$$

Thus, K^0 is proportional to the time rate of change of the energy.

In practical calculations, we usually do not need to deal with the four-force but prefer to use just the three components given by Equation 13.50. In vector form, we have

$$\vec{F} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{m_0\vec{v}}{\sqrt{1-\beta^2}}. \tag{13.52}$$

This relativistic equation of motion reduces to the Newtonian form $\vec{F} = \mathrm{d}\vec{P}/\mathrm{d}t$, provided we use the relativistic momentum $\vec{P}$ given by the first equation of Equation 13.36.

We can show that the dot product of the four-force K^μ with the four-velocity u^μ vanishes:

$$\sum_\mu K^\mu u^\mu = \sum_\mu \frac{\mathrm{d}(m_0 u^\mu)}{\mathrm{d}\tau}u^\mu = \sum_\mu m_0\frac{\mathrm{d}(u^\mu u^\mu/2)}{\mathrm{d}\tau} = \sum_\mu \frac{m_0}{2}\frac{\mathrm{d}(-c^2)}{\mathrm{d}\tau} = 0. \tag{13.53}$$

But the dot product can also be written as

$$\begin{aligned}\sum_{\mu=1}^{4} K^\mu u^\mu &= \frac{\vec{F}\cdot\vec{v}}{1-\beta^2} + K^4u^4 = \frac{\vec{F}\cdot\vec{v}}{1-\beta^2} - K^0u^0 \quad (K^4 = iK^0,\ u^4 = iu^0)\\ &= \frac{\vec{F}\cdot\vec{v}}{1-\beta^2} - \frac{cK^0}{\sqrt{1-\beta^2}}.\end{aligned}$$

Combining this with Equation 13.53, we find

$$\frac{\vec{F}\cdot\vec{v}}{1-\beta^2} - \frac{cK^0}{\sqrt{1-\beta^2}} = 0$$

from which we obtain

$$K^0 = \frac{\vec{F}\cdot\vec{v}}{c\sqrt{1-\beta^2}}. \tag{13.54}$$

Now,

$$K^0 = \frac{\mathrm{d}P^0}{\mathrm{d}\tau} = \frac{1}{\sqrt{1-\beta^2}}\frac{\mathrm{d}}{\mathrm{d}t}\frac{m_0c^2}{\sqrt{1-\beta^2}}.$$

Combining this with Equation 13.54 gives a very useful relationship:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{m_0c^2}{\sqrt{1-\beta^2}} = \vec{F}\cdot\vec{v}. \tag{13.55}$$

13.9.1 Force Transformation

Consider a particle of rest mass m_0 having velocity $\vec{v}'$ and momentum $\vec{P}'$ relative to frame S' and having velocity $\vec{v}$ and momentum $\vec{P}$ relative to frame S. Then, the force acting on the particle measured in S' is, from Equation 13.50,

$$F'_x = \mathrm{d}P'_x/\mathrm{d}t' \tag{13.56}$$

where

$$F'_x = F'_1, \quad P'_x = P'_1 = m_0v_x/\sqrt{1-\beta^2} \quad (v_x = v_1).$$

Now, Equation 13.48 gives, with some modifications in notations,

$$P'_x = \gamma\left(P_x - \frac{u}{c}\frac{E}{c}\right), \quad P'_y = P_y, \quad P'_z = P_z, \quad E' = \gamma(E - uP_x).$$

Hence, Equation 13.56 becomes, after it is made use of the first transformation relationship,

$$F'_x = \gamma\frac{\mathrm{d}}{\mathrm{d}t'}(P_x - uE/c^2). \tag{13.57}$$

But

$$\frac{\mathrm{d}}{\mathrm{d}t'} = \frac{\mathrm{d}t}{\mathrm{d}t'}\frac{\mathrm{d}}{\mathrm{d}t}$$

and

$$\frac{\mathrm{d}t'}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}t}\gamma\left(t - \frac{ux}{c^2}\right) = \gamma(1 - uv_x/c^2).$$

Substituting these into Equation 13.57,

$$\begin{aligned} F'_x &= \frac{\gamma}{\gamma(1-uv_x/c^2)}\frac{\mathrm{d}}{\mathrm{d}t}(P_x - uE/c^2) \\ &= \frac{1}{(1-uv_x/c^2)}\left(F_x - \frac{u}{c^2}\frac{\mathrm{d}E}{\mathrm{d}t}\right). \end{aligned}$$

Using Equations 13.55 and 13.36, we have

$$\frac{\mathrm{d}E}{\mathrm{d}t} = \vec{F}\cdot\vec{v}.$$

Hence,

$$F'_x = \frac{1}{(1 - uv_x/c^2)}\left(F_x - \frac{u}{c^2}\vec{F}\cdot\vec{v}\right).$$

Now,

$$\vec{F}\cdot\vec{v} = F_x v_x + F_y v_y + F_z v_z.$$

Hence,

$$F'_x = F_x - \frac{uv_y}{c^2(1 - uv_x/c^2)}F_y - \frac{uv_z}{c^2(1 - uv_x/c^2)}F_z. \tag{13.58}$$

Next, we consider

$$F'_y = \frac{\mathrm{d}P'_y}{\mathrm{d}t'} = \frac{\mathrm{d}P_y}{\mathrm{d}t'} = \frac{\mathrm{d}P_y}{\mathrm{d}t}\frac{\mathrm{d}t}{\mathrm{d}t'}, \quad (P_y = Py).$$

But

$$\frac{\mathrm{d}t}{\mathrm{d}t'} = 1/\gamma(1 - uv_x/c^2).$$

Hence,

$$F'_y = \frac{F_y}{\gamma(1 - uv_x/c^2)}. \tag{13.59}$$

Similarly, because

$$P'_z = P_z$$

we have

$$F'_z = \frac{F_z}{\gamma(1 - uv_x/c^2)}. \tag{13.60}$$

The inverse transformations are

$$F_x = F'_x + \frac{uv'_y}{c^2(1 + uv'_x/c^2)}F'_y + \frac{uv'_z}{c^2(1 + uv'_x/c^2)}F'_z \tag{13.61}$$

$$F_y = \frac{F'_y}{\gamma(1 + uv'_x/c^2)} \tag{13.62}$$

$$F_z = \frac{F'_z}{\gamma(1 + uv'_x/c^2)} \tag{13.63}$$

$$E = \chi c^2 \quad \vec{P} = \chi c\hat{e}.$$

13.10 RELATIVISTIC LAGRANGIAN AND HAMILTONIAN FUNCTIONS

As in nonrelativistic mechanics, equations of motion can be written in generalized coordinates in the form of Lagrange's or Hamilton's equations. To do this, we must first find the Lagrangian or Hamiltonian. This is a relatively easy task for a free particle. Because the action integral for a free particle must be invariant under Lorentz transformations, it follows that the action integral must be taken over a scalar, and the latter must have the form of a differential of the first order. The only scalar of this kind that can be associated with a free particle is a quantity proportional to the interval ds. So for a free particle, the action must have the form

$$S = \alpha \int_a^b \mathrm{d}s \tag{13.64}$$

where α is some constant characterizing the particle, and the integral is along the world line of the particle between two world points (i.e., between two particular events of the arrival of the particle at the initial position and at the final position at definite times t_1 and t_2). Now $\mathrm{d}\tau = \mathrm{d}s/c = \mathrm{d}t\sqrt{1-\beta^2}$, and the action integral (Equation 13.63) can be rewritten as an integral with respect to the time:

$$S = \alpha \int_{t_1}^{t_2} c\sqrt{1 - v^2/c^2}\, \mathrm{d}t = \int_{t_1}^{t_2} L\,\mathrm{d}t$$

where v is the velocity of the particle. We arrive at the conclusion that the Lagrangian for a free relativistic particle is

$$L = \alpha c\sqrt{1 - v^2/c^2}. \tag{13.65}$$

At the limit $\ll c$, our expression for L must be reduced to the Newtonian expression $L = 2\, m_0 v^2$. To carry out this transition, we expand Equation 13.64 in powers of v/c. Ignoring the terms of higher orders, we obtain

$$L = \alpha c\sqrt{1 - v^2/c^2} \cong \alpha c - \alpha v^2/2c.$$

We may discard the constant term that does not affect the equation of motion. Consequently, in the Newtonian approximation $L = -av^2/2c$, a comparison with the Newtonian expression shows that $\alpha = -m_0 c$. We have thus established the form of the Lagrangian for a free particle:

$$L = -m_0 c^2\sqrt{1 - v^2/c^2}. \tag{13.66}$$

The next problem is to extend the free particle Lagrangian so that it includes the effects of external forces that act on the particle. If the forces acting on the particle were conservative forces independent of velocity, a suitable Lagrangian for such a particle would be

$$L = -m_0 c^2\sqrt{1 - v^2/c^2} - V \tag{13.67}$$

where V is the potential energy of the particle, depending on position only. Note that the Lagrangian is no longer $L = T - V$. That this is the correct Lagrangian can be shown by demonstrating that the

Lagrange's equation resulting from it agrees with Equation 13.50. We shall leave this as homework for the reader.

Having established the Lagrangian, we can now find the Hamiltonian H. If the Lagrangian L does not depend explicitly on time, then, by definition,

$$\begin{aligned} H &= \sum_j \dot{q}_j p_j - L \\ &= m_0 v^2/\sqrt{1-v^2/c^2} + m_0 c^2\sqrt{1-v^2/c^2} + V \end{aligned}$$

which, on collecting terms, reduces to

$$H = m_0 c^2/\sqrt{1-v^2/c^2} + V = T + V. \tag{13.68}$$

The Hamiltonian H is seen again to be the total energy.

The relativistic Hamiltonian can be expressed in terms of the momentum of the particle. By means of Equation 13.36, we have

$$p_1^2 + p_2^2 + p_3^2 = p^2 = m_0^2 v^2 (1 - v^2/c^2)^{-1}.$$

Then a simple calculation gives

$$H = \sqrt{p^2 c^2 + m_0^2 c^4} + V. \tag{13.69}$$

When $p \gg m_0 c$, the Hamiltonian of a free particle attains the simple form

$$H \cong pc.$$

A motion with such large momentum, for which the above approximation is valid, is called an ultra-relativistic motion. It is clear that for particles with zero rest mass, the expression is valid.

Example 13.4

As an application of the Lagrangian Equation 13.66, we consider a particle of rest mass m_0 moving under a central force - $\mathrm{d}V/\mathrm{d}r = -V'$. As in the nonrelativistic case, the orbit is in a plane, and so we employ plane polar coordinates (r, θ). Then L is

$$L = -m_0 c^2 \left[1 - \frac{\dot{r}^2 + r^2\dot{\theta}^2}{c^2}\right]^{1/2} - V(r) \tag{13.70}$$

from which we find

$$\frac{\partial L}{\partial \dot{r}} = \gamma m_0 \dot{r} \qquad \frac{\partial L}{\partial r} = \gamma m_0 r\dot{\theta}^2 - \frac{\partial V}{\partial r}$$
$$\frac{\partial L}{\partial \dot{\theta}} = \gamma m_0 r^2 \dot{\theta} \qquad \frac{\partial L}{\partial \theta} = 0.$$

The Lagrange's equation for coordinate r gives

$$\frac{\mathrm{d}(\gamma\dot{r})}{\mathrm{d}t} - \gamma r\dot{\theta}^2 + \frac{V'}{m_0} = 0 \tag{13.71}$$

where

$$\gamma = \left[1 - \frac{\dot{r}^2 + r^2\dot{\theta}^2}{c^2}\right]^{1/2}. \tag{13.72}$$

Because θ is cyclic, the Lagrange's equation for coordinate θ yields an integral of the motion:

$$\frac{\partial L}{\partial\dot{\theta}} = m_0\gamma r^2\dot{\theta} = A \text{ (constant)}. \tag{13.73}$$

This is the relativistic law of area, from which we have

$$\dot{\theta} = \frac{A}{m_0\gamma r^2}. \tag{13.74}$$

Then, we can write

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\mathrm{d}}{\mathrm{d}\theta}\frac{\mathrm{d}\theta}{\mathrm{d}t} = \dot{\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}$$

and so forth.

In terms of the new variable $y = 1/r$, Equation 13.71 becomes

$$\frac{\mathrm{d}^2 y}{\mathrm{d}\theta^2} + y - \frac{\gamma m_0 V'}{A^2 y^2} = 0. \tag{13.75}$$

It is desirable to eliminate γ from Equation 13.74. To this end, we seek help from the energy integral Equation 13.67, which can be written as

$$H = \gamma m_0 c^2 + V$$

from which we have

$$\gamma = \frac{H - V}{m_0 c^2}.$$

Substituting this into Equation 13.75, we obtain the differential equations for the general central orbit:

$$\frac{\mathrm{d}^2 y}{\mathrm{d}\theta^2} + y - \frac{(H - V)\,V'}{c^2 A^2 y^2} = 0. \tag{13.76}$$

If $V(r) = -k/r = -ky$ (the inverse square law of force; $k > 0$ for an attractive force), Equation 13.75 reduces to the form

$$\frac{\mathrm{d}^2 y}{\mathrm{d}\theta^2} + y(1 - a) - C = 0 \tag{13.77}$$

where

$$a = \frac{k^2}{c^2 A^2}, \quad C = \frac{kH}{c^2 A^2}.$$

With an appropriate choice of initial conditions, its solution is

$$r = \frac{1}{y} = \frac{l}{1 + \varepsilon \cos \eta\theta} \tag{13.78}$$

with

$$\eta^2 = 1 - a, \; l = (1 - a)/C. \tag{13.79}$$

The apses of the orbit are

$$r_{min} = \frac{l}{1+\varepsilon}, \quad \text{at } \eta\theta = 0, 2\pi, 4\pi, \ldots \tag{13.80}$$

and

$$r_{max} = \frac{l}{1-\varepsilon}, \quad \text{at } \eta\theta = \pi, 3\pi, \ldots. \tag{13.81}$$

The angle between successive apse lines is given by π/η:

$$\frac{\pi}{\eta} = \frac{\pi}{\sqrt{1-a}} \tag{13.82}$$

which reduces to $\cong (\pi + k^2/2c^2A^2)$ for the case where a is small. Thus, the orbit advances the perihelion as shown in Figure 13.14, where the circles have radii r_{max} and r_{min}.

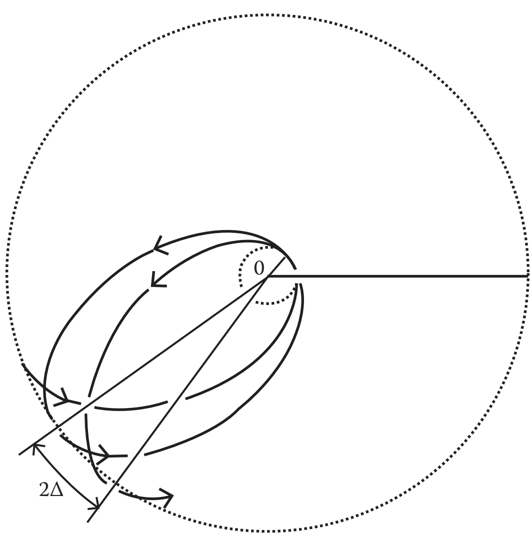


FIGURE 13.14 Perihelion advance due to the change of mass with velocity $\Delta = k^2/m^2c^2A^2$.

13.11 RELATIVISTIC KINEMATICS OF COLLISIONS

The subject of collisions is of considerable interest in experimental high-energy physics. Although the interactions (forces) between elementary particles are non-classical, as long as the particles involved in a reaction are outside the region of mutual interaction, their mean motion can be described by classical mechanics. The fundamental principles involved in the analysis of collisions are the conservation laws of momentum and energy. Because energy and momentum form a four-momentum vector, the conservation equations can be written as one four-vector equation. We shall see that it is easy to work with this four-vector equation.

We shall restrict our discussion to two-particle collisions that can be illustrated symbolically by the reaction

$$A + B \to C + D + E + \dots.$$

This is a generalization of the two-body collision considered in Chapter 9; it allows for the possibility of two particles A and B colliding and producing a group of particles C, D, E, and so on. Conservation of the energy–momentum four-vector gives

$$\vec{p}_A + \vec{p}_B = \vec{q}_C + \vec{q}_D + \vec{q}_E + \dots$$

where $\vec{p}$'s denote the momenta before the collision, and $\vec{q}$'s denote the momenta after the collision. The problem is that we know the four-vector for A and B; we are given some information about the four-vectors of C, D, and so forth, and we have to find the unknown momenta and energies. The solution to such a problem can be found by the use of a particular technique. A simple example will illustrate the flavor of this technique. For simplicity and clarity, we consider the case where only two particles, which may or may not be identical, are produced after a collision. Suppose we are told nothing about the four-momentum vector of particle D, but we are asked about the dynamic state of particle C. The technique is to rearrange the equation for the conservation of the four-momentum vectors so that the four-momentum vector for the particle, which we are not interested in (i.e., particle D), stands alone on one side of the equation:

$$\vec{q}_D = \vec{p}_A + \vec{p}_B - \vec{q}_C.$$

Taking the scalar product of each side with itself and using the result that the length of the four-momentum vector squared is an invariant and equal to (rest mass × c)^2, we obtain

$$(m_D c)^2 = (m_A c)^2 + (m_B c)^2 + (m_C c)^2 + 2\vec{p}_A \cdot \vec{p}_B - 2\vec{p}_A \cdot \vec{q}_C - 2\vec{p}_B \cdot \vec{q}_C$$

in which the only four-vectors remaining are those we know or the one we wish to find, that is, $\vec{p}_A$, $\vec{p}_B$, and $\vec{q}_C$. Let us illustrate this technique by looking at a few specific examples.

Because our examples involve photons, let us first take a brief review of the nature of photons. As shown a little earlier, a consequence of the relativistic energy–momentum relationship is the possibility of "massless" particles that possess energy and momentum but no rest mass: $E = pc$. Furthermore, massless particles must travel at the speed of light.

Photons interact electromagnetically with electrons and other charged particles. Neutrinos and gravitons (not yet detected) are two other possible massless particles.

Example 13.5: Compton Scattering

As a simple example, we consider the Compton scattering, where an electron at rest scatters an incident photon (Figure 13.15). Given the incident photon energy, what is the energy of the scattered photon?

FIGURE 13.15 Compton scattering.

Solution:

We denote the incident photon by γ_1 and the scattered photon by γ_2. Because we are not interested in the electron after the scattering, the energy–momentum four-vector conservation equation is written as

$$\vec{q}_e = \vec{p}_e + \vec{p}_{\gamma_1} - \vec{q}_{\gamma_2}. \tag{13.83}$$

Taking the scalar product of each side with itself and using the invariance of the square of the length and the special properties of the photon four-vector, we obtain

$$\begin{aligned} m^2c^2 = m^2c^2 &+ 2c^2\frac{E_{\gamma_1}}{c^2}\frac{mc^2}{c^2} - 2c^2\frac{E_{\gamma_2}}{c^2}\frac{E_{\gamma_2}}{c^2} \\ &+ 2\vec{p}_{\gamma_1}\cdot\vec{q}_{\gamma_2} - 2c^2\frac{mc^2}{c^2}\frac{E_{\gamma_2}}{c^2} \end{aligned} \tag{13.84}$$

where $E_{\gamma 1}$ and $E_{\gamma 2}$ are the energies of the incident photon and the scattered photon, respectively. If φ is the angle between the scattered photon and the incident photon, then

$$\vec{p}_{\gamma_1}\cdot\vec{q}_{\gamma_1} = p_{\gamma_1}q_{\gamma_1}\cos\phi = \frac{E_{\gamma_1}}{c^2}\frac{E_{\gamma_2}}{c^2}\,. \tag{13.85}$$

Solving Equations 13.84 and 13.85 for $E_{\gamma 2}$, we find

$$E_{\gamma_2} = \frac{E_{\gamma_1}}{1+(E_{\gamma_1}/mc^2)(1-\cos\phi)}. \tag{13.86}$$

For a photon $E = hc/\lambda$, Equation 13.86 can be rewritten as

$$\lambda_2 = \lambda_1\left[1+\frac{h}{mc\lambda_1}(1-\cos\phi)\right]$$

from which we obtain

$$\lambda_2 - \lambda_1 = \lambda_c\,(1 - \cos\phi) \tag{13.87}$$

where $\lambda_c = h/mc$ is known as the Compton wavelength of the electron. Here, we have treated photons as quantum particles. In classical physics, a photon would be treated as a wave, and consequently, there would be no change in wavelength after scattering.

The electron was assumed free and at rest. For sufficiently high photon energies, this is a good approximation for electrons in the outer shells of light atoms. If the binding energy of the electron is comparable to the photon energy, momentum and energy can be transferred to the atom as a whole, and the photon can be completely absorbed; there would be no scattering.

The special theory of relativity was not widely accepted in the 1920s partly because of the radical nature of its space–time concepts but also because of a lack of experimental evidence. Now, the result of Compton scattering left little doubt the relativistic dynamics are valid.

Example 13.6: Electron–Positron Pair Annihilation

Positrons, which are antiparticles of electrons, can be found in nature. They are detected in cosmic radiation and as products of radioactivity from a few radioactive elements. However, they cannot live free very long because of their interaction with electrons. When positrons and electrons are in near proximity, they annihilate each other to produce two photons:

$$e^+ + e^+ \rightarrow \gamma + \gamma.$$

The reaction is illustrated in Figure 13.16, where the positron is at rest.

Conservation of the energy–momentum four-vector gives

$$\vec{p}_+ + \vec{p}_- = \vec{q}_{\gamma_1} + \vec{q}_{\gamma_2}. \tag{13.88}$$

Of course, there is no difference between the two photons, so let us label them 1 and 2. If we want to know what the energy of photon 1 is as a function of the angle between the two photons, we rearrange Equation 13.88 in the following form:

$$\vec{q}_{\gamma_2} = \vec{p}_+ + \vec{p}_- - \vec{q}_{\gamma_1}$$

and then take the square of both sides with the following result:

$$\begin{aligned}\vec{q}_{\gamma_2}^{\,2} = \vec{q}_{\gamma_2} \cdot \vec{q}_{\gamma_2} &= \vec{p}_+ \cdot \vec{p}_+ + \vec{p}_- \cdot \vec{p}_- + \vec{q}_{\gamma_1} \cdot \vec{q}_{\gamma_1} \\ &+ 2\vec{p}_+ \cdot \vec{p}_- - 2\vec{p}_+ \cdot \vec{q}_{\gamma_1} - 2\vec{p}_- \cdot \vec{q}_{\gamma_1}.\end{aligned} \tag{13.89}$$

Equation 13.88 can be simplified with the following facts:

1. The energy-momentum four-vector is an invariant; $E^2 = p^2c^2 + m^2c^4$.
2. The photon has no rest mass.
3. The energy and momentum of a photon satisfy the simple relation $E = pc$.
4. $p_+ = 0$ (the positron is initially at rest), and $E_+ = mc^2$.

With these aids, Equation 13.89 reduces to

$$\begin{aligned}0 = m^2c^2 + m^2c^2 + 2c^2 \frac{E_-}{c^2}\frac{mc^2}{c^2} \\ - 2c^2 \frac{E_+}{c^2}\frac{E_{\gamma_1}}{c^2} - 2c^2 \frac{E_-}{c^2}\frac{E_{\gamma_1}}{c^2} + 2\vec{p}_- \cdot \vec{q}_{\gamma_1}.\end{aligned}$$

FIGURE 13.16 Electron–positron pair annihilation.

If φ_1 is the angle between the direction of the propagation of photon 1 and the direction of the incoming electron, then

$$2\vec{p}_{-}\cdot\vec{q}_{\gamma_1} = p_{-}q_{\gamma_1}\cos\phi_1 = p_{-}E_{\gamma_1}\cos\phi_1/c.$$

We finally obtain

$$E_{\gamma_1} = \frac{mc^2(E_{-}+mc^2)}{E_{-}+mc^2-cp_{-}\cos\phi_1}.$$

For fixed values of E_{-} and p_{-}, the photon energy will be at a maximum when it is emitted in the forward direction (i.e., $\varphi_1 = 0$, $\cos\varphi_1 = 1$) and is at a minimum when it is emitted in the backward direction.

The results just given should apply equally well to photon 2. If photon 1 is emitted in the forward direction with the maximum energy, then, because of conservation of energy and momentum and because the initial momentum is in the forward direction, photon 2 must be emitted in the backward direction with minimum energy.

13.12 COLLISION THRESHOLD ENERGIES

The majority of collision experiments in high-energy physics produce one, two, or more particles of the same or different types as the initial ones; the total mass-produced often is greater than the mass of the particles producing the interaction. How can this reaction occur? It could happen because some of the incoming kinetic energy is converted to mass. This leads to the concept of the threshold energy for a reaction (i.e., the minimum kinetic energy of the incoming particle so that the reaction will just occur).

Consider a simple reaction

$$1 + 2 \rightarrow 3 + 4$$

where particle 2 is initially at rest in the laboratory frame, and incident particle 1 has a total energy E_1 and momentum $\vec{p}_1$. We use unprimed quantities for the laboratory frame and primed quantities for the center-of-mass frame that is moving with a velocity v relative to the laboratory frame along the x-axis.

The energy–momentum four-vector $(E/c, \vec{p})$ is an invariant:

$$E^2/c^2 - p^2 = M^2c^2.$$

For any system of particles, the total energy and total momentum also form a four-vector and is an invariant. Now the total energy is

$$E_0 = M_2c^2 + E_1$$

and in the center-of-mass frame, the total momentum is zero. The invariance of the energy–momentum four-vector gives us

$$E_0^2/c^2 - p_1^2 = E_0'^2/c^2.$$

Thus, we obtain

$$E_0'^2 = E_0^2 - p_1^2c^2 = (M_2c^2+E_1)^2 - (E_1^2 - M_1^2c^4) = M_1^2c^4 + M_2^2c^4 + 2M_2E_1c^2.$$

Solving for E_1,

$$E_1 = \frac{E_0'^2 - M_1^2c^4 - M_2^2c^4}{2M_2c^2}. \tag{13.90}$$

The center-of-mass frame is moving with a velocity v relative to the laboratory frame along the x-axis. The Lorentz transformation then gives

$$E_0 = \gamma E_0'$$

where

$$\gamma = \frac{1}{\sqrt{1-\beta^2}} \quad \text{and} \quad \beta = \frac{v}{c} = \frac{cp_1}{E_1 + M_2c^2}. \tag{13.91}$$

Before we proceed further, let us digress for a moment to derive the last expression. For a particle with velocity v, we have

$$p = \gamma M v = \gamma M c \beta$$

and

$$E = \gamma Mc^2 = c(p^2 + M^2c^2)^{1/2}$$

where the last step is from the invariance of the energy–momentum four-vector. From the last expressions, we obtain

$$\beta = \frac{cp}{E}$$

which can be applied to the two-particle system and yields the second expression in Equation 13.91.

We now return to our main problem. The energy released, denoted by Q, of the reaction is

$$Q = (M_1 + M_2 - M_3 - M_4)c^2.$$

If Q is positive, the reaction can proceed for all values of E_1. But if Q is negative, the reaction has a threshold; that is, there is a minimum value of E_1, denoted by E_1th, for which the reaction can occur. At threshold,

$$E_0' = (M_3 + M_4)c^2$$

and then, Equation 13.90 gives

$$E_1^{th} = \frac{\left[(M_3 + M_4)^2 - M_1^2 - M_2^2\right]c^2}{2M_2}$$

or

$$T_1^{th} = \frac{[(M_3 + M_4)^2 - (M_1 + M_2)^2]c^2}{2M_2}$$

where T_1 is the kinetic energy of the incident particle M_1.

PROBLEMS

1. Two inertial frames S and S' have their respective x-axes parallel with S' moving with constant velocity v along the positive x-direction of the system S. A rod makes an angle of 30° with respect to the x'-axis. What is the value of v if the rod makes an angle of 45° with respect to the x-axis?
2. The wave equation

$$\nabla^2\psi - \frac{1}{c^2}\frac{\partial^2\psi}{\partial t^2} = 0$$

   is satisfied in a vacuum for any component of $\vec{E}$ or $\vec{B}$, and it is a consequence of Maxwell's equations. Show that this wave equation is invariant under a Lorentz transformation. For simplicity, you can restrict yourself to the case in which the wave propagates along the x-axis.
3. A light signal is emitted by observer 0 at exactly the moment when observer 0′ passes him with velocity $\vec{v}$ (relative to 0). Show that the velocity of light in any direction is the same as measured by 0 and 0′.
4. A track star on Earth runs the 100-m dash in 10.0 s, as measured on Earth. What is his or her time as measured by a clock on a spaceship receding from the Earth with speed $0.99c$?
5. A spaceship moves away from the Earth with constant speed $v = c/\sqrt{2}$. The astronaut observes that a rod-like external probe is 6 m long and makes an angle of 45° with the ship's line of motion. To an observer on Earth, how long is the probe and what angle does it make with the line of motion?
6. A spaceship moving away from the Earth at a velocity $v_1 = 0.75c$ with respect to the Earth launches a rocket with a velocity $v_2 = 0.75c$ in the direction away from the Earth. What is the velocity of the rocket with respect to the Earth?
7. An observer on Earth sees two spaceships A and B approaching her along the same straight line: A approaches from the left with speed $c/2$, and B from the right with speed $3c/4$. With what speed does each spaceship approach the other?
8. A man on a station platform sees two trains approaching each other at the rate $7c/5$, but an observer on one of the trains sees the other train approaching her with a velocity $35c/37$. What are the velocities of the trains with respect to the station?

$$v = \frac{v' + u}{1 + uv'/c^2}$$

9. Show that if two Lorentz transformations with relative velocities given by v_1 and v_2, respectively, are carried out consecutively, the result is the same as that of a single Lorentz transformation with relative velocity v given by

$$\beta = \frac{\beta_1 + \beta_2}{1 + \beta_1\,\beta_2}$$

   where

$$\beta = V/c,\ \beta_1 = v_1/c,\ \text{and}\ \beta_2 = v_2/c.$$

10. Show that even if the relativistic mass is used, the expression $mv^2/2$ is not the correct relativistic energy.
11. To an observer O, a particle moves with the velocity $\vec{v}$ and orientation specified by the usual spherical coordinates θ and ϕ. Find its apparent velocity $\vec{v}'$ for a second observer O′ moving with a uniform velocity u in the positive x-direction relative to O.
12. A pion of mass π comes to rest and then decays into a muon of mass μ and a neutrino of mass zero. Show that the kinetic energy of the muon is $c^2(\pi - \mu)^2/2\pi$.
13. Two identical bodies, each with rest mass m_0, approach each other with equal velocities u, collide, and stick together in a perfectly inelastic collision.
    (a) What is the rest mass of the composite body?
    (b) What is the rest mass of the composite body as determined by an observer who is at rest with respect to one of the initial bodies?
14. A particle of rest mass m_0 and velocity $v = 3c/5$ collides with a stationary particle of rest mass m_0. Assuming that, after collision, the two particles stick together, find the velocity and the rest mass of the composite particle.
15. Consider a system of non-interacting particles. Show that its rest mass M exceeds the sum of rest masses of constituent particles by the total kinetic energy of the particles (divided by c^2):

$$M = \sum_j m_j + \frac{1}{c^2}\sum_j T_j.$$

16. A particle of rest mass m_0 moves under the action of a constant force. Find the time dependence of the particle's velocity.
17. A particle of rest mass M_0 moves with an instantaneous velocity $\vec{v}$ under the action of a force $\vec{F}$. Show that
    (a) If $\vec{F}$ is parallel to $\vec{v}$, then $\vec{F} = M_0\gamma^3\vec{a}$, where $\vec{a} = d\vec{v}/dt$, and $\gamma = \dfrac{1}{\sqrt{1 - v^2/c^2}}$.
    (b) If $\vec{F}$ is perpendicular to $\vec{v}$, then $\vec{F} = M_0\gamma a$.
18. As viewed by an observer in the laboratory, a proton collides with another proton initially at rest. After the collision, a proton and an antiproton come off in addition to the original protons. Find the minimum kinetic energy that the incoming proton must have to make this reaction energetically possible.
19. The relativistic Doppler effect. A light source flashes with period $\tau_0 = 1/v_0$ in its rest frame, and the source moves toward an observer with velocity v. Show that the frequency of the pulses v received by the observer is

$$v = v_0\left[\frac{1 + v/c}{1 - v/c}\right]^{1/2}$$

and that if the observer is at angle θ from the line of motion, v_D is given by

$$v = v_0\frac{(1 - v^2/c^2)^{1/2}}{(1 - v\cos\theta/c)}.$$

20. Find the relativistic path of an electron moving in the field of a fixed-point charge.
21. Derive an expression for the acceleration four-vector that gives the rate of change of the velocity four-vector of a particle with respect to its proper time. What are its components in the case when either the magnitude or the direction of velocity is invariable?

22. The result of any two Galilean transformations is another transformation of the same type.
   (a) Show that the same is true of two Lorentz transformations if the directions of $\vec{v}_1$ and $\vec{v}_2$ are the same.
   (b) If $\vec{v}_1$ and $\vec{v}_2$ have different directions, then there is a small resultant space rotation, known as the Thomas precession. Show that the small angle of the space rotation is given by

$$d\vec{\theta} = \vec{v}_1 \times \vec{v}_2 / 2c^2.$$

   This result was applied by Thomas to the case of the spinning electron in the hydrogen atom.

REFERENCES

Okun, L.B. The concept of mass. *Phys. Today*, 31–36, 1989.
Resnick, R., and Halliday, D. *Basic Concepts in Relativity,* Wiley, 1985.
Terrell, J.L. The invisibility of the Lorentz contraction, *Phys. Rev.,* 116, 1041, 1959.
Weisskopf, V.F. The visual appearance of rapidly moving objects. *Phys. Today*, 14, 9, 24, 1960.

14 Newtonian Gravity and Newtonian Cosmology

14.1 NEWTON'S LAW OF GRAVITY

We were introduced to Newton's law of gravity briefly earlier in the discussion of the concept of force in Chapter 2. Because it is another fundamental principle, let us revisit it and take a close look. This law states that "between any two particles there is a force of attraction that is directly proportional to the product of the masses of the particles and inversely proportional to the square of their distance apart." The justification for regarding this law as one of our fundamental principles lies in its universality. Newton was led to formulate this law through a study of planetary motion. Before the time of Newton, Kepler, based on Tycho's observational record of Mars' motion, had found that the motions of the planets followed certain empirical laws: the celebrated three laws of planetary motion. These laws express certain regularities found in the observed data, but they are descriptive only of the motion of planets about the sun and do not explain why planets revolve about the sun. Newton showed that if he assumed the law of gravitation, then, with the aid of his laws of motion, the motion of the planets could be described and Kepler's laws derived (Chapter 7). Newton next showed that his laws could be applied to the motion of the moon around the Earth and to bodies falling near the Earth's surface; he compared the accelerations toward the Earth's center in the two cases and found that they satisfied the inverse square law. In his work, Newton first assumed and later proved that if the mass of a large sphere is symmetrically distributed, then the gravitational attraction on an external particle is the same as if the mass of the sphere were concentrated at its center. We shall consider this theorem later.

The law of gravitation has been tested in delicate laboratory experiments, such as the Cavendish experiment, in which one directly measures the gravitational attraction between suspended and fixed spheres that are a few centimeters apart. The law is valid over a wide range in the distance of separation of the masses involved. It is also found to be independent of physical properties (other than the mass) of the bodies and of the medium separating them. Hence, it is truly a universal law.

In equation form, the force of attraction experienced by a mass m in the field of, and at a distance r from, a mass M is, as we saw earlier,

$$\vec{F} = -\frac{GmM}{r^2}\hat{r} = -\frac{GmM}{r^3}\vec{r} \tag{14.1}$$

in which the origin is placed at M, and $\hat{r} = (\vec{r}/r)$ is a unit vector pointing along the radius vector from the origin. The minus sign indicates that the force is directed opposite to $\hat{r}$; that is, it is directed toward the origin. The constant G is the so-called gravitational constant, which is determined experimentally and in cgs units has the value 6.67×10^{-8}. G is the least accurately known of the fundamental constants.

Now, consider gravitational force arising from N masses $m_1, m_2, \ldots, m_N$ at fixed positions $\vec{r}_1, \vec{r}_2, \ldots, \vec{r}_N$ (Figure 14.1), on mass m at $\vec{r}(t)$, which is changing. Newton's law of gravity gives

$$\vec{F} = -\sum_{i=1}^{N} \frac{Gmm_i(\vec{r} - \vec{r}_i)}{\left|\vec{r} - \vec{r}_i\right|^3}. \tag{14.2}$$

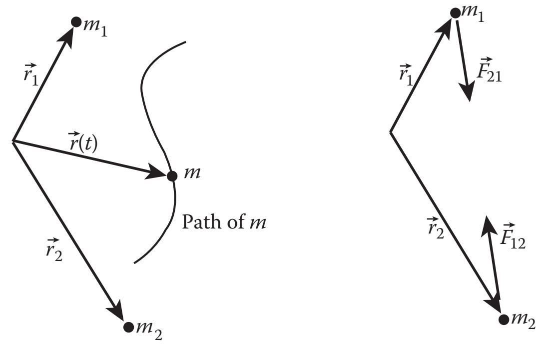


FIGURE 14.1 Path of a particle: action and reaction.

From Equation 14.2, we see that the mass m that governs a body's resistance to external applied force—the inertial mass—is also the mass that determines the magnitude of the gravitational force acting on it, that is, the passive gravitational mass. Gravitational mass m^ occurs in Newton's law of gravity in two ways: passively $\left(m_g^p\right)$ in determining the force on the body from others and actively $\left(m_g^a\right)$ in determining the force with which the body attracts others. Active and passive gravitational masses are proportional, and without loss of generality, we can define them as equal. This follows from the third law of motion. Consider two particles 1 and 2, at $\vec{r}_1$ and $\vec{r}_2$. The force $\vec{F}_{21}$ that 2 exerts on 1 is

$$\vec{F}_{21} = -\frac{Gm_{g1}^p m_{g2}^2(\vec{r}_1 - \vec{r}_2)}{\left|\vec{r}_1 - \vec{r}_2\right|^3}$$

while

$$\vec{F}_{12} = -\frac{Gm_{g2}^p m_{g1}^2(\vec{r}_2 - \vec{r}_1)}{\left|\vec{r}_2 - \vec{r}_1\right|^3}$$

is the force that 1 exerts on 2. The third law of motion then gives

$$m_{g1}^p m_{g2}^a = m_{g2}^p m_{g1}^a \quad \text{or} \quad \frac{m_{g1}^p}{m_{g2}^p} = \frac{m_{g1}^a}{m_{g2}^a}.$$

That is, active and passive gravitational masses are proportional.

This dual role played by mass has an astonishing consequence. Inserting Equation 14.2 into Newton's second law of motion, we see that m cancels, and we get

$$\vec{a} = -G\sum_{i=1}^{N} \frac{m_i(\vec{r} - \vec{r}_i)}{\left|\vec{r} - \vec{r}_i\right|^3}.$$

That is, the acceleration of a particle in any gravitational field is independent of its mass. Thus, if only gravitational forces act, all bodies similarly projected pursue identical trajectories. This was known to Galileo and Newton. Both Galileo and Newton conducted experiments to test the equivalence of inertial and gravitational masses, and their experiments suggest that

$$m_{\mathrm{I}} = m_g.$$

It has been established by Dicke in 1964 and by Panov and Braginski in 1971 that they are equal within one part in 10^{12}.

The assertion of the equivalence of the inertial mass and gravitational mass is known as the principle of equivalence. Pursuit of the consequences of the principle of equivalence led Einstein to formulate his theory of general relativity between 1911 and 1916.

14.2 GRAVITATIONAL FIELD AND GRAVITATIONAL POTENTIAL

We now attempt to reformulate Newton's theory of gravitation so that action at a distance is eliminated. This can be done through the field concept. It is often convenient to use field strength or field intensity when we are dealing with a force field. The gravitational field strength at a given point is defined as the gravitational force per unit mass that would act on a particle located at that point and is a vector quantity. If the origin of force is considered to be the mass *M*, the gravitational field strength at the given point where the mass *m* is located is

$$\vec{f} = \frac{\vec{F}}{m} = -\frac{GM}{r^2}\hat{r}_1. \tag{14.3}$$

If more than one particle is present, the gravitational field is the vector sum of the individual vector fields produced by each particle. Clearly, the gravitational field strength has the dimensions of acceleration. For the uniform field at the Earth's surface, the field intensity $\vec{f}$ is simply $\vec{g}$, the acceleration of gravity.

If more than one body is present, the gravitational field is the vector sum of the individual fields produced by each body. For a body that consists of a continuous distribution of matter, Equation 14.3 becomes

$$\vec{f} = -G\int_{vol} \frac{\rho(\vec{r}')\hat{r}}{r^2}\,\mathrm{d}V'. \tag{14.4}$$

Upon using the identity $\nabla(1/r) = -(1/r^2)\hat{r}$, Equation 14.4 becomes

$$\vec{f} = G\int_{Vol} \nabla(1/r)\rho(\vec{r}')\,\mathrm{d}V'.$$

Because ∇ does not operate on r', it can be factored out of the integral, and we have

$$\vec{f} = \nabla\int_{Vol} \frac{G\rho(\vec{r}')}{r}\,\mathrm{d}V'. \tag{14.5}$$

Equation 14.5 indicates that $\vec{f}$ can be expressed as the gradient of a scalar function denoted by Φ:

$$\vec{f} = -\nabla\Phi \tag{14.6}$$

with

$$\Phi = -\int_{Vol} \frac{G\rho(\vec{r}')}{r}\,\mathrm{d}V' = -\frac{GM}{r}. \tag{14.7}$$

Thus, the static gravitational field is a conservative vector field. The scalar function Φ is called the gravitational potential and has the dimension of energy per unit mass. Φ is the gravitational potential energy per unit mass.

The reader may question why we place a negative sign in front of $\nabla\Phi$ in Equation 14.6. Because the gravitational force vanishes at $r = \infty$, it is customary to consider $r = \infty$ as the zero of potential energy. Thus, the potential energy is negative for all attractive forces of the form $1/r^n$ with $n > 1$ and positive for all repulsive force varying in the same fashion.

Example 14.1

(a) Find the intensity of the gravitational field outside and at a distance r from the center of a uniform sphere of mass M.
(b) A particle of mass m is brought from ∞ to a point P at a distance r from the center of this uniform sphere. Find the potential energy of m at point P.

Solution:

(a) Consider a surface S of a sphere of radius r, concentric with the sphere of mass m. At any element of area dS, let θ be the angle between the direction of $\vec{f}$ and that of $d\vec{S}$ (Figure 14.2). We then have

$$\iint_S \vec{f}\cdot d\vec{S} = -\iint_S \frac{Gm}{r^2}\cos\theta\, dS = Gm\iint_S d\Omega = -Gm(4\pi) \tag{14.8}$$

where $d\Omega = \cos\theta\, dS/r^2$ is the solid angle subtended by dS at the center. From symmetry considerations, the vector field $\vec{f}$ must be constant in magnitude over S and directed radially, so that $\vec{f}\cdot d\vec{S} = f dS$. Then, we have

$$\iint_S \vec{f}\cdot d\vec{S} = \iint_S f\, dS = f\iint_S dS = f(4\pi r^2).$$

Combining this result with Equation 14.8, we have

$$4\pi f r^2 = -4\pi Gm$$

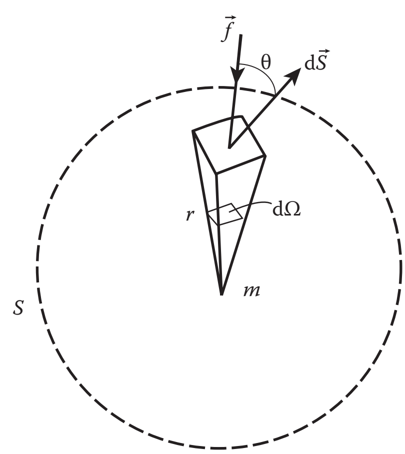


FIGURE 14.2 Gravitational field outside a uniform sphere.

or

$$f = -\frac{Gm}{r^2}\ .$$

The negative sign means that the field is inward. The field is seen to be the same as it would be if the sphere were replaced by a particle of mass m at its center. This was first proved by Newton: "If the mass of a large sphere is symmetrically distributed, the gravitational attraction on an external particle is the same as if the mass of the sphere were concentrated at its center."

(b) The potential energy of m at point P may be computing the negative work done by the force on m while it is brought from ∞ to r. A straight-line path is the most convenient one. We then have

$$V = -\int_{\infty}^{r} \vec{F}\cdot \mathrm{d}\vec{r} = -\int_{\infty}^{r} F\,\mathrm{d}r = GMm\int_{\infty}^{r}\frac{\mathrm{d}r}{r^2} = -\frac{GMm}{r}$$

which is a negative quantity in accordance with Equation 14.7.

14.3 GRAVITATIONAL FIELD EQUATIONS: POISSON'S AND LAPLACE'S EQUATIONS

Suppose we have a continuous distribution of mass, and the mass per unit volume or mass density at any point is ρ. Then, from Equation 14.8

$$\iint_S \vec{f}\cdot \mathrm{d}\vec{S} = -4\pi G \iiint_V \rho\,\mathrm{d}V.$$

We can transform the surface integral on the left side into a volume integral by using the divergence theorem:

$$\iint_S \vec{f}\cdot \mathrm{d}\vec{S} = \iiint_V \nabla\cdot\vec{f}\,\mathrm{d}V.$$

Then, we have

$$\iiint_\tau \nabla\cdot\vec{f}\,\mathrm{d}V = -4\pi G \iiint_V \rho\,\mathrm{d}V.$$

Because V is an arbitrary volume, we get

$$\nabla\cdot\vec{f} = -4\pi G\rho. \qquad (14.9)$$

Substituting Equation 14.8 into Equation 14.9, we get

$$\nabla^2\Phi = 4\pi G\rho. \qquad (14.10)$$

This is known as Poisson's equation. At points where no matter is present, $\rho = 0$, and

$$\nabla^2\Phi = 0. \tag{14.11}$$

This is known as Laplace's equation.

14.4 GRAVITATIONAL FIELD AND POTENTIAL OF EXTENDED BODY

As it stands, Equation 14.1 applies to particles or a spherical symmetric body. Thus, to compute the force between two extended bodies, it is necessary to consider the contributions of all the elements of mass of each body. The procedure for finding the forces or the field strength by a direct integration is difficult except for symmetrically shaped bodies. For example, we want to find the field at point P produced by the body in Figure 14.3. Two mass elements dm_i and dm_j are shown; each gives a tiny contribution $d\vec{F}_i$ and $d\vec{F}_j$, respectively, to the field at P. These must be added *vectorially*. Then, the contributions of the other mass elements of the body must be vectorially compounded with this result. Obviously, this procedure is often a laborious one. A better method, for many instances, is first to compute the potential of the body. This is generally a simpler procedure because the contribution of each mass element of the body is a scalar and can thus be added together by simple algebraic addition. The field strength may then be obtained by computing the gradient of the potential.

Example 14.2: Field and Potential of a Homogeneous Spherical Shell

The origin is at the center of the shell that has a radius a and a mass σ per unit area. A ring element is selected, the axis of which is OP (Figure 14.4). The potential at P resulting from this element is

$$d\Phi = -\frac{G\,dm}{R} = -\frac{G\sigma(2\pi a\sin\theta)a\,d\theta}{R} = -\frac{2\pi Ga^2\sigma\sin\theta\,d\theta}{R}.$$

Now,

$$R^2 = a^2 + r^2 - 2ar\cos\theta$$

and

$$2R\,dR = 2ar\sin\theta\,d\theta.$$

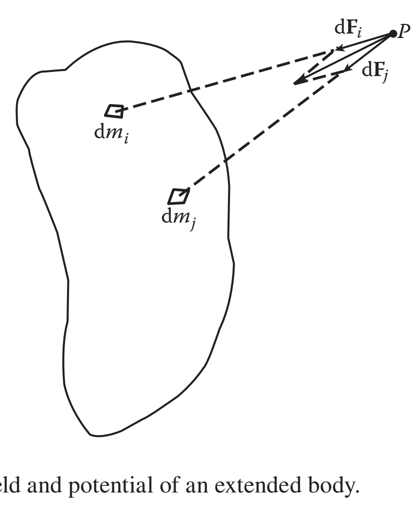


FIGURE 14.3 Gravitational field and potential of an extended body.

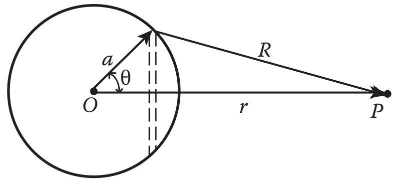


FIGURE 14.4 Gravitational field and potential of a homogeneous spherical shell.

Substituting for sinθ dθ in dΦ, we get

$$d\Phi = -\frac{2\pi G a\sigma}{r} dR.$$

Integrating between the limits $r - a$ and $r + a$, we obtain

$$\Phi = -\frac{2\pi G a\sigma}{r}\int_{r-a}^{r+a} dR = -\frac{GM}{r}$$

where M (= $4\pi a^2\sigma$) is the total mass of the shell. This result states that the potential of a homogeneous spherical shell at an external point is the same as if all the mass were concentrated at the center. This result is also the potential of a homogeneous solid sphere at an external point in which M is the total mass of the sphere. Why? A homogeneous solid sphere may be regarded as a large number of concentric spherical shells.

The field can be computed at once by using Equation 14.6, and we have

$$\vec{f} = -\nabla\Phi = -\frac{d\Phi}{dr}\hat{r} = -\frac{GM}{r^2}\hat{r}$$

where the unit vector $\hat{r}$ points in the direction of increasing r.

14.5 TIDES

The moon travels around the Earth in almost a circular orbit, and the Earth is so large that the moon's gravitational attraction on Earth is not uniform, and appreciable non-local effects such as the tides can be observed. To a lesser extent, the sun is also responsible for Earth tides. Although the sun is 400 times more massive than the moon, the sun is also 400 times further away from the Earth than the moon with the sun's attraction on Earth's different parts almost with equal force. As a result, the moon is the major tide generator. The atmosphere, the ocean, and the solid Earth all experience tidal forces; only the ocean tides are commonly observed.

The tide-generating force is the difference between the gravitational force on the surface of the Earth and at the center of the Earth. The gravitational field of the moon at the center of the Earth is

$$\vec{f}_0 = Gm\frac{\hat{n}}{a^2}$$

where m is the moon's mass, a is the distance between the center of the moon and the center of the Earth, and $\hat{n}$ is the unit vector from the Earth toward the moon (Figure 14.5).

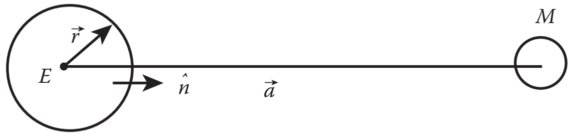


FIGURE 14.5 Moon's gravitational attraction on earth is not uniform.

If $\vec{f}(\vec{r})$ is the gravitational force of the moon at some point $\vec{r}$ on the Earth, where the origin of $\vec{r}$ is the center of the Earth, then the force on mass m at $\vec{r}$ is $m\vec{f}(\vec{r})$. The apparent force to an earth-bound observer is

$$\vec{F} = m\vec{f}(\vec{r}) - m\vec{f}_0 = m[\vec{f}(\vec{r}) - \vec{f}_0]$$

and the apparent field is

$$\vec{f}'(\vec{r}) = \vec{f}(\vec{r}) - \vec{f}_0.$$

Figure 14.6 shows the true field $\vec{f}(\vec{r})$ at different points on the Earth's surface. f_a is larger than f_0 because point a is closer to the moon than the center. Similarly, f_c is less than f_0. The magnitudes of $\vec{f}_b$ and $\vec{f}_d$ are approximately equal to the magnitude of $\vec{f}_0$, but their directions are slightly different. The apparent field $\vec{f}'$ is shown in Figure 14.7. We now evaluate the apparent fields at different points on the Earth's surface. For simplicity, we choose a static model, that is, rotations are not included.

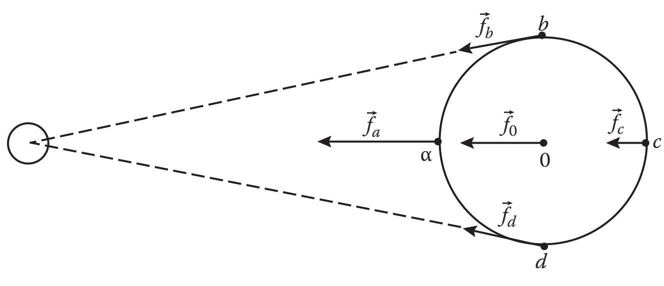


FIGURE 14.6 $\vec{f}(\vec{r})$ at different points on the earth's surface.

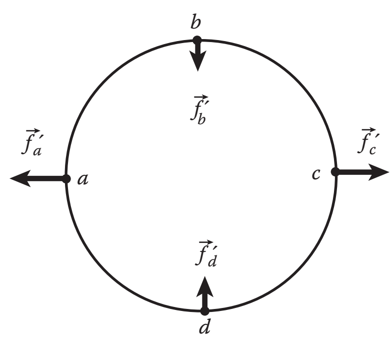


FIGURE 14.7 Apparent field $\vec{f}'$ at different points on the earth's surface.

Consider a point $\vec{r}$ on the Earth's surface; $\vec{a}$ is the distance of the moon relative to the Earth's center. The gravitational potential at the point is

$$\Phi(\vec{r}) = -\frac{Gm}{|\vec{r}-\vec{a}|} \tag{14.12}$$

where m is the mass of the moon (note that our analysis also applies to the sun) (Figure 14.8).

Now, $|\vec{r}-\vec{a}|^2 = r^2 - 2ar\cos\theta + a^2$. Hence, because $r \ll a$

$$\begin{aligned}\frac{1}{|\vec{r}-\vec{a}|} &= \frac{1}{a}\left(1 - 2\frac{r}{a}\cos\theta + \frac{r^2}{a^2}\right) \\ &= \frac{1}{a}\left[1 - \frac{1}{2}\left(-2\frac{r}{a}\cos\theta + \frac{r^2}{a^2}\right) + \frac{3}{8}\left(-2\frac{r}{a}\cos\theta + \frac{r^2}{a^2}\right)^2 - \ldots\right]. \\ &= \frac{1}{a} + \frac{r}{a^2}\cos\theta + \frac{r^2}{a^3}\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right) + \ldots\end{aligned}$$

Taking the direction of the moon to be the z direction, and using Equation 14.12, we obtain

$$\Phi(\vec{r}) = -Gm\left(\frac{1}{a} + \frac{r}{a^2}\cos\theta + \frac{r^2}{a^3}\left[\frac{3}{2}\cos^2\theta - \frac{1}{2}\right] + \ldots\right). \tag{14.13}$$

The first term is a constant and does not yield force. The major effect of the moon's gravitational force is from the second term, which may be rewritten as, with $z = r\cos\theta$,

$$\Phi_2 = -\frac{Gm}{a^2}z.$$

Thus, it is to accelerate the Earth as a whole and is irrelevant for tide generation. The third quadratic term

$$\Phi_3 = -Gm\frac{r^2}{a^3}\left[\frac{3}{2}\cos^2\theta - \frac{1}{2}\right]$$

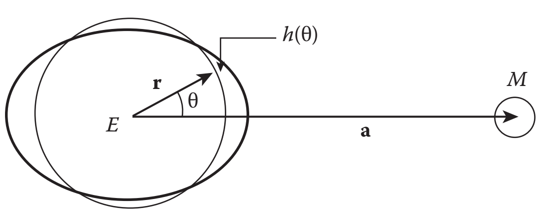


FIGURE 14.8 Gravitational potential at the point $\vec{r}$ on the Earth's surface.

is responsible for the tide generation. Let us see why. Φ_3 leads to a gravitational force

$$f_r' = -\frac{\partial \Phi_3}{\partial r} = \frac{Gmr}{a^3}(3\cos^2\theta - 1) \tag{14.14}$$

$$f_\theta' = -\frac{1}{r}\frac{\partial \Phi_3}{\partial \theta} = -\frac{3Gmr}{a^3}\cos\theta\sin\theta.$$

The field f_r' is directed outward along the z-axis (i.e., toward the moon) and f_θ' inward in the xy-plane.
We now evaluate f_r' at each of the points indicated in Figure 14.7.

1. f_a' and f_c'
   At point a, $\theta = 0°$ and $\cos\theta = 1$; thus,

$$f_a' = 2Gmr/a^3 = 2f_0(r/a)$$

where

$$\vec{f}_0 = Gm\frac{\hat{n}}{a^2}.$$

The analysis at c is similar except that the distance to the moon is $r + a$ instead $r - a$. We obtain

$$f_c' = -2Gmr/a^3 = -2f_0(r/a).$$

Note that both f_a' and f_c' point radially out from the Earth (Figure 14.7).

2. f_b' and f_d'
   Point b and d are, to excellent approximation, the same distance from the moon as the center of the Earth. However, f_b' is not parallel to f_0. The angle between them is $\alpha \approx r/a = 4.3 \times 10^{-5}$. To this approximation,

$$f_b' = f_0\alpha = 2f_0(r/a).$$

By symmetry, f_d' is equal and opposite to f_b'. Both f_b' and f_d' point toward the center of the Earth (Figure 14.9).

The sketch in Figure 14.10 shows $f'(r)$ at various points on the Earth's surface. This diagram is the starting point for analyzing the tides. The forces at a and c tend to lift the oceans, and the forces at b and d tend to depress them. If the Earth were uniformly covered

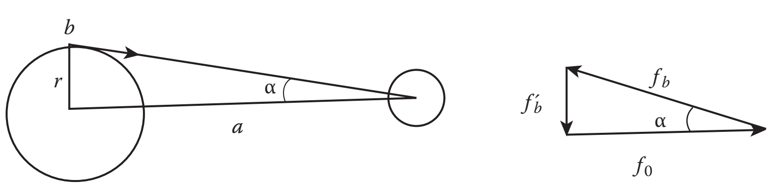


FIGURE 14.9 Evaluation of the apparent field f_b' at point b on the earth's surface.

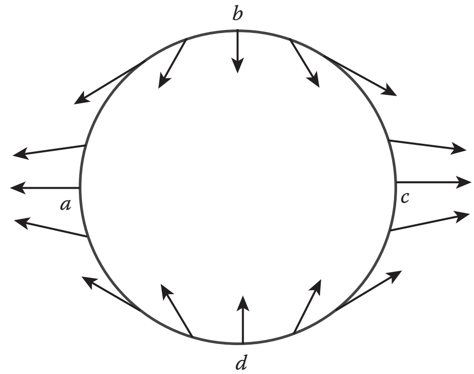


FIGURE 14.10 Sketch of $f'(r)$ at various points on the Earth's surface.

with water, the tangential force components would cause the two tidal bulges to sweep around the globe with the moon. This picture explains the twice daily ebb and flood of the tides, but the actual motions depend in a complicated way on the response of the oceans as the Earth rotates and on features of local topography.

We can estimate the equilibrium height of the tide easily. In equilibrium, the surface of the water must be an equipotential surface. We can thus calculate the height $h(\theta)$ as a function of the angle θ to the moon's direction (see Figure 14.8). Because $h(\theta)$ is small, the change in the Earth's gravitational potential is approximately $g_0 h(\theta)$, where $g_0 = GM/r^2$, and M is the mass of the Earth. This change must be balanced by the quadratic term in the potential Φ because of the moon, so that

$$g_0 h(\theta) = \frac{Gmr^2}{a^3}\left(\frac{3}{2}\cos^2\theta - 1\right).$$

Thus, using $g_0 = GM/r^2$, we find

$$h(\theta) = h_0\left(\frac{3}{2}\cos^2\theta - \frac{1}{2}\right)$$

where

$$h_0 = \frac{mr^4}{Ma^3}. \tag{14.15}$$

From the numerical values,

$$M = 5.98 \times 10^{27}\ \text{g} \quad r = 6.37 \times 10^{8}\ \text{cm}$$

$$m = 7.34 \times 10^{25}\ \text{g} \quad a = 3.84 \times 10^{10}\ \text{cm}$$

$$M_{\text{sun}} = 1.99 \times 10^{33}\ \text{g} \quad a_s = 1.49 \times 10^{13}\ \text{cm}$$

We find for the moon, h_0 = 36 cm = 0.36 m, and for the sun, h_0 = 16 cm = 0.16 m. We see that the moon's effect is about twice as large as the sun's.

These figures must not be taken as more than order-of-magnitude estimates of the height of the tides. Even on a completely ocean-covered Earth, there would be important effects modifying them. First, we have neglected the gravitational attraction of the ocean itself. The tidal bulges exert an attraction that tends to increase h_0 slightly, in fact, by approximately 12%. More important is the fact that the Earth is not perfectly rigid and is itself distorted by the tidal forces. Because the observed tidal range refers not to absolute height but to the relative heights of sea surface and sea floor, this reduces the effective height by a substantial factor. Indeed, in the extreme case of a fluid Earth enclosed by a completely flexible crust, there would be essentially no observable tide at all.

The values obtained from Equation 14.15 may seem rather small, especially when further reduced by the effect just mentioned. However, it is important to remember that they refer to the unrealistic case of an Earth without continents. The observed tidal range in the mid-ocean is in fact quite small, normally less than a meter. Large tides are a feature of continental shelf areas and strongly dependent on the local topography.

The strongest tides, called the spring tides, occur at the new and full moon when the moon and sun act together. Midway between, at the quarters of the moon, occur the weak neap tides. The ratio of the driving forces in these two cases is (Figure 14.11)

$$\frac{h_{0,\text{spring}}}{h_{0,\text{neap}}} = \frac{h_{0,\text{moon}} + h_{0,\text{sun}}}{h_{0,\text{moon}} - h_{0,\text{sun}}} \approx 3.$$

Tidal forces can have surprising effects on the Earth–moon system. Friction with the ocean beds drags the tidal bulges eastward out of a direct Earth–moon line as shown in Figure 14.12. (Note that the Earth rotates about its axis much faster than the moon revolves around the Earth.) These tidal bulges contain a large amount of mass, and their gravitational hold pulls the moon forward in its orbit. As a result, the moon's orbit is growing larger, and it is receding from Earth at about 4 cm per year, an effect that astronomers can measure by bouncing laser beams off reflectors left on the lunar surface by the astronauts. As the moon moves away from the Earth, its sidereal period is becoming longer. Meanwhile, the friction of the ocean waters with the seabed slows the rotation of the Earth and makes the day grow by 0.0023 s per century. Fossils of marine animals and tidal sediments confirm that only 900 million years ago, Earth's day was 18 hours long. At some point in the distant future, the Earth will be rotating so slowly that a solar day will equal a lunar month. At that time, the Earth's tidal bulge will be aimed directly at the moon, and

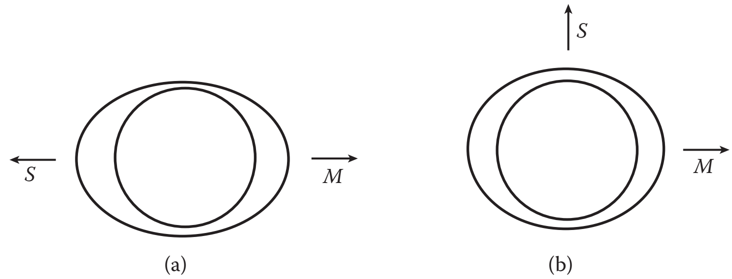


FIGURE 14.11 (a) Spring tides; (b) neap tides.

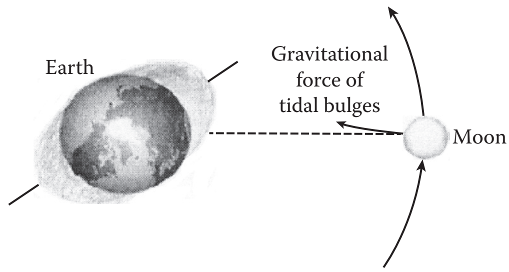


FIGURE 14.12 Frictional effect of the tides on the Earth.

the moon will stop spiraling away from the Earth. The Earth will then keep the same side facing the moon. Pluto and its moon have already achieved this stable configuration.

The tidal forces can also trigger moonquakes. Just as the gravitational forces of the sun and moon deform the Earth, the sun and the Earth deform the moon. The moon is alternately squeezed and stretched as it orbits our planet. Because the Earth is more massive than the moon, the Earth produces relatively large tidal deformations of the lunar surface. The strongest gravitational stresses occur when the moon is nearest the Earth in its orbit—that is, at perigee—which is just the time when the Apollo seismometers report the highest frequency of moonquakes.

14.6 GENERAL THEORY OF RELATIVITY: RELATIVISTIC THEORY OF GRAVITATION

In the Newtonian theory of gravitation, the gravitational effects travel at infinite speed instead of at or below the speed of light, the limiting speed set by the special theory of relativity. Furthermore, the concept of inertial frames of reference is incompatible with gravitational phenomena. An inertial frame is defined as one in which a free particle (no force acting on it) moves with a constant velocity. But gravity is long-range and cannot be screened. Consequently, the only way to visualize an inertial frame is to imagine it far away from any matter. A concept like this is clearly of little use to someone doing experiments on Earth or to astronomers whose observations relate to distant massive galaxies. Attempts to modify Newtonian gravity to make it compatible with special relativity have not been successful.

Because Newtonian gravity is proportional to the mass of an object, Einstein quite naturally turned to the consideration of the concept of mass, and this leads him to establish the principle of equivalence, the equivalence of the inertial mass of a body and its gravitational mass. How does one go about putting this important fact to use in unearthing the secrets of nature? Here is Einstein's genius. He made use of the principle of equivalence to show, with a thought experiment, that locally any accelerated motion is equivalent to a gravitational field (for details, see Section 12.7). This is another version of the principle of equivalence. If we accept this version of the principle of equivalence, then there is no meaning to be attached to the notion of gravitational force. Gravity in this context behaves like an inertial force, a fictitious force that arises as a result of the acceleration of the frame of reference from which observations are being made. The most familiar examples of inertial forces are centrifugal and Coriolis forces in a rotating system fixed on the Earth's surface.

According to Einstein, if gravitation is long-range and unscreened, it has something of a permanent character, and it must be intrinsic to the region in which it is located. Einstein identified this intrinsic property of a space–time with its geometry. Space–time is not flat, as had been previously assumed, but it is curved or wrapped by the distribution of mass and energy in it. So the space–time

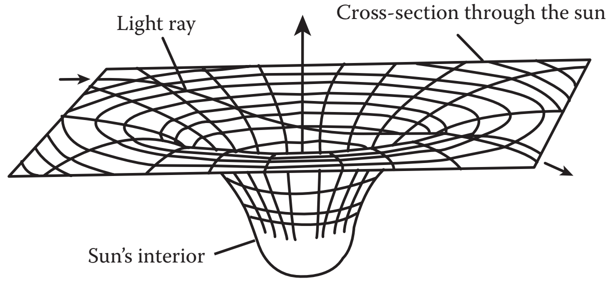


FIGURE 14.13 Bending of star light by a space–time warp caused by the sun.

geometry is non-Euclidean. In 1915, he made the revolutionary suggestion that gravity is not a force like other forces; it is a consequence of the curved space–time. Bodies such as the Earth or the moon are not made to move on curved orbits by a force called gravity; instead, they follow the nearest thing to a straight path in a curved space–time called a geodesic.

We can see the effect of space–time curvature in one of the famous tests of the general relativity: the bending of starlight by a space–time warp or "sink" caused by the sun as shown in Figure 14.13. This has analogy to the deviation of a rolling ball caused by a depression on a horizontal surface. Such bending can be observed during a total solar eclipse because the starlight is not concealed by sunlight scattered in the atmosphere (that is why we cannot see stars in the daytime). The space program offers new evidence. When Mars is on the far side of the sun, radio signals from a lander must pass through the space–time warp caused by the sun to reach us. It has been observed that the signals were delayed by about 100 μs, and that is very close to what is predicted by general relativity.

Soon after Einstein announced his general theory of relativity, Austrian physicists Josef Lense and Hans Thirring suggested in 1918 that a massive rotating body could literally drag space–time around with it as the body rotates and so modify the space–time curvature. This is known as the Lense–Thirring effect or frame dragging. The more massive the rotating body, the larger the effect.

To test this effect on the space–time around the Earth, two experiments have been proposed. One is gravity probe B (or GP-B, for short) that involves launching a gyroscope in orbit around the Earth. Normally, the axis of a gyroscope is fixed in space. However, if the above-mentioned general relativistic effect is present, the axis should process about a fixed direction. The predicted precession rate is about 7 in. per year at a height of 800 km, and this is not too small to be measured. The experiment is developed by Stanford University and NASA and was launched on April 20, 2004, from Vandenberg Air Force Base in California. For the past 3 years, GPB has circled the Earth, collecting data to determine the frame-dragging effect and the other effect (the geodetic effect, the amount by which the mass of the Earth warps the local space–time in which it resides). The first results confirm the two predictions of Einstein's general relativity theory; the results of May 2011 from the GT-B give the two critical predictions a resounding confirmation. It is critically important to thoroughly analyze the data to ensure their accuracy and integrity prior to releasing the results.

We now return to the principle of equivalence, which leads to two testable conclusions about the propagation of light. If the effects of gravitation and acceleration are locally indistinguishable, then rays of light should bend in a gravitational field. As well, light moving up through a gravitational field should be red-shifted. Let us look at gravitational red shift first.

14.6.1 Gravitational Shift of Spectral Lines (Gravitational Red Shift)

Consider an observer inside an enclosed space cabin that is accelerating with acceleration g through a gravitation-free region of space. T and R are two fixed points on a straight line parallel to the direction of g. A light wave of frequency v is emitted at T. It will not have the same frequency when observed at R. To analyze the situation, we first note that the light ray will take a time $\Delta t = l/c$ to reach

R (l is the distance between T and R), and during this time, the point R gains an additional velocity $v = g\Delta t = gl/c$. This apparent relative velocity with respect to T will cause a Doppler frequency shift:

$$\nu' = \nu \frac{1-\beta}{\sqrt{1-\beta^2}} = \nu\left(1-\beta+\frac{1}{2}\beta^2+\ldots\right),\ \beta = v/c.$$

Neglecting the higher order term, we have

$$\nu' = \nu\,(1-\beta) = \nu\,(1-gl/c^2) \tag{14.16}$$

and the period τ'

$$\tau' = \frac{1}{\nu'} = \tau\left(1+\frac{gl}{c^2}\right). \tag{14.17}$$

Now, using the equivalence principle, we can conclude that the same phenomenon must be observed in a gravitational field. We note that gl is the difference in gravitational potential between the source and the receiver. Let us denote this difference by $\Delta\phi$. We see that in passing up through a gravitational potential difference of $\Delta\phi$, the light has become redder and its wavelength is increased by the amount $\Delta\phi/c^2$. If light falls through a gravitational potential difference $\Delta\phi$, it gains energy and becomes bluer by the amount $\Delta\phi/c^2$, that is, its wavelength is Doppler-shifted toward the blue by the amount $\Delta\phi/c^2$. Pound and Rebka have verified this remarkable prediction in a terrestrial laboratory in 1960. They allowed a 14.4-keV γ-ray, emitted by the radioactive decay of ^{57}Fe, to fall 22.6 m down an evacuated tower and measured the change in its frequency. The predicted blue shift is $z = -2.46 \times 10'15$, and they measured $z = (-2.57 \pm 0.26) \times 10'15$, thus directly verifying the equivalence principle. Such high precision was possible because of the Mossbauer effect; this is the emission of radiation from an atomic nucleus in a crystal, which gives a spectral line with a very precisely defined frequency.

Equation 14.17 indicates that the gravitational shift of spectral lines implies that in a gravitational field, a clock (periodic phenomenon) runs slower than does the same clock in a gravitation-free region of space.

14.6.2 Bending of Light Beam

Consider the space cabin again, accelerating through a gravitation-free region of space. A light ray enters the cabin from a window at P, and it is parallel to the floor of the cabin as it enters the cabin. Where will the light hit on the opposite wall? The cabin will move upward while the light is traveling to the opposite wall so that the light will hit the wall at Q', a little below Q. Thus, to an observer inside the cabin, the light ray curves downward as it travels through the cabin (Figure 14.14a). Now,

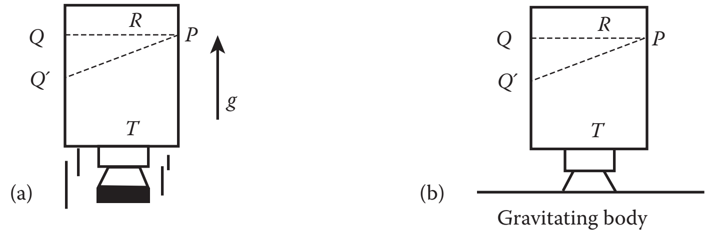


FIGURE 14.14 Bending of a light beam in a gravitational field. (a) A space cabin accelerating through a gravitation-free region, (b) the space cabin is a gravitational field.

using the equivalence principle, we conclude that in a gravitational field, light will not travel along a straight line, but its path will be curved (Figure 14.14b).

Although in the context of Newtonian gravity, we do not expect that gravity will influence the propagation of light, Newton somehow made a non-relativistic quantitative calculation of the bending of light in 1801. He got half the value given by the general relativity theory. We now make a semiclassical calculation, which suffers from a foundation built upon a Euclidean space–time. Our calculation is based on the following assumptions:

(1) Light is a beam of photons.
(2) A photon has energy $E = h\nu$, where h is Planck's constant and ν is photon frequency.
(3) A photon has effective $m = h\nu/c^2$, taken as equivalent to gravitational mass.

Now, consider a photon of effective mass m passing by a star of mass M and radius R as shown in Figure 14.15. For a star similar to the sun, the deflection angle φ proves to be very small. The photon feels the radial force of gravity, and as is clear from the symmetry in the picture, the star imparts a net transverse momentum to the photon, which bends its trajectory as shown. The calculation is successful only if m, the mass of the photon, cancels out of the calculation. The effect must be universal and independent of m. Following the original Newtonian calculation done by Soldner (a German mathematician) in 1801, the gravitational force F produces a change in the photon's transverse momentum P_T:

$$\mathrm{d}P_T = F\cos\theta\,\mathrm{d}t.$$

We see from Figure 14.15 that $x = R\tan\theta$, so we can trade the linear position of the photon for the angle θ. The angle θ varies from $-\pi/2$ to $\pi/2$ during the process. Because the deflection is small, we have the good approximation $\mathrm{d}x = c\mathrm{d}t$. Now, $\mathrm{d}x = R\sec^2\theta\,\mathrm{d}\theta$, and $R = r\cos\theta$, and so $\mathrm{d}P_T$ becomes

$$P_T = \int_{-\pi/2}^{\pi/2} \frac{GmM}{cR}\cos\theta\,\mathrm{d}\theta = \frac{2GmM}{cR}.$$

The deflection angle is approximately

$$\varphi = \frac{P_T}{P} = \frac{2GmM/cR}{mc} = \frac{2GmM}{c^2R}\ \text{radians}.$$

Substituting in the parameters for the sun, φ = 0.87 seconds of arc. The general relativity calculation predicts a value twice this value or 1.75 seconds of arc, a small but measurable effect. The experiment was first performed in 1919; the result reported was 1.7 seconds of arc. The experiment is difficult. The procedure of the experiment is to photograph a star field around the sun during a total eclipse (so that the stars are visible). Six months later, when the same stars are visible at night, the

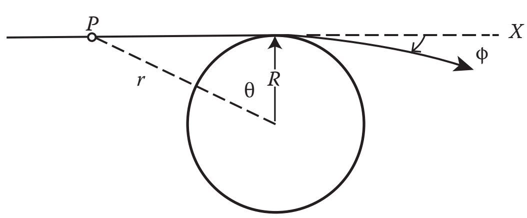


FIGURE 14.15 Semiclassical calculation of the bending of a light beam (photons) by a gravitating body.

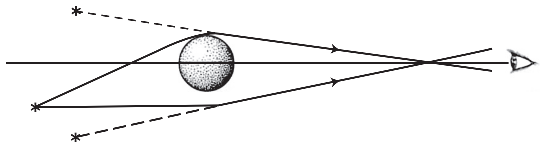


FIGURE 14.16 Gravitational lenses effect.

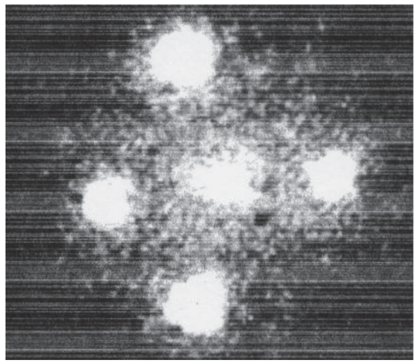


FIGURE 14.17 Optical image of Q2237+031, called the "Einstein Cross."

star field is photographed again. The displacement of the apparent positions of the stars can be measured by comparing the two photographs. Accuracy of this experiment is perhaps 20%. But general relativity prediction has been confirmed by better experimental measurements of φ.

When rays of light from a distant source are deflected by the gravitational field of a massive object, the rays that pass on opposite sides of the mass intersect at some large distance beyond (see Figure 14.16). An observer placed at such a large distance will simultaneously see the source at two locations in the sky; that is, we will see two images of the source. The massive objects that produce multiple images are called gravitational lenses.

Gravitational lensing is now regularly observed by astrophysicists, including the lensing of distant quasars by galaxies and lensing of stars in the galactic nucleus and in the Larger Magellanic Cloud by more nearby stars. About half a dozen quasars seem to have nearby companion quasars with essentially identical spectra but slightly different brightness and shapes. The existence of two so nearly identical objects so close together is unlikely; the companions are actually images of a single quasar created by a gravitational lens. Figure 14.17 is the optical image of Q2237+031, called the Einstein Cross, obtained with the Hubble Space Telescope. Four images of the quasar surround the central, fainter image of the galaxy that acts as gravitational lens (F.D. Macchetto, NASA/ESA).

14.7 INTRODUCTION TO COSMOLOGY

Cosmology is the study of the dynamic structure of the universe and seeks to answer questions regarding the origin, the evolution, and the future behavior of the universe as a whole. And cosmologists piece together the observed information about the universe into a self-consistent theory or model that describes the nature, origin, and evolution of the universe. All cosmological model constructions are based on the following basic assumptions:

1. The physical laws we know on Earth apply everywhere in the universe.
2. On a large scale (roughly 100 Mpc or more), the universe is homogeneous.
3. On a large scale, the universe looks the same in every direction (isotropy).

The assumptions of homogeneity and isotropy lead to the so-called cosmological principle that is often stated as follows:

> Any observer in any place (any galaxy) sees the same general features of the universe.

This means that our local sample of the universe is no different from more remote and inaccessible regions.

The expansion of the universe (established observationally in 1929 by Edwin Hubble) and the cosmic microwave background radiation are the basic observed information that allows us to probe the large-scale structure of the universe and construct a cosmological model.

Einstein's theory of general relativity is essential for a proper treatment of cosmology. But we do not have the required working knowledge of general relativity, so we will study cosmology from Newtonian dynamics and gravitation and the cosmological principle, that is, we study Newtonian cosmology. It reproduces most of the relativistic results. We will comment on the shortcomings of Newtonian cosmology at the end of this section.

14.8 BRIEF HISTORY OF COSMOLOGICAL IDEAS

Every culture has had its cosmology, its story of how the universe came into being, what it is made of, and where it is going. The mythological stories can be traced to the earliest writings of Babylonian, Egyptian, Greek, and Chinese civilizations. Transition from mythology to the birth of scientific inquiry occurred abruptly in the middle of the sixth century BC on the shore of Asia Minor. The earliest surviving attempt at a rational cosmology is probably that of Pythagoras. He taught that (1) the Earth is round and rotates on its axis; (2) the sun, moon, stars, and planets revolve in concentric spheres around a central fire. The "fixed" stars form the outermost sphere; and (3) The motion of the celestial bodies produces the harmony of the musical scale. Although Pythagorean philosophy prepared the way for a heliocentric cosmology and persisted for several centuries, its musical emphasis on celestial harmony based on the musical scale made it eventually obsolete.

Then the ideas of Plato and Aristotle appeared around fourth century BC. They advocated the idea of daily rotation of the heavens around the spherical immovable Earth. Some five centuries later, Ptolemy introduced a geocentric cosmology, which was adopted later by the Roman Catholic Church as an article of faith, so the Ptolemaic theory was not seriously challenged for 1400 years, until Nicolas Copernicus (1473–1543) reexamined it in the early 1500s.

Copernicus introduced a heliocentric system; he showed that the motion of the planets around the sun, with the moon orbiting around a rotating Earth, provided a far simpler and more elegant explanation of the planetary motion. Copernicus was primarily concerned with planets and did not take the logical step of scattering the stars through space. An Englishman, Thomas Digges, took that step in 1576.

The next great advance came as a result of serious observations of planets by Tycho Brahe (1546–1601) and his assistant Johannes Kepler (1571–1630). Kepler formulated, from Tycho's observational data, the three laws of planetary motion and made a lasting contribution of great significance.

Then came Galileo Galilei (1564–1642); he used the newly developed primitive telescope to discover the phases of Venus, much like our moon. This showed that Venus revolves around the sun. The discovery of four large satellites of Jupiter showed that Earth is not the center of all motion in the cosmos. Galileo himself did not contribute significantly to cosmological theory, but his discoveries made a path for others to follow. After Galileo, scientists have relied more and more on evidence, observation, and measurement.

Kepler and Galileo were unable to explain why the planets move around the sun in elliptical orbits and what keeps the solar system together. Isaac Newton (1643–1727) provided the underlying theory: the law of gravity. He used it to explain Kepler's laws of planetary motion. William Herschel found that binary stars in orbit around one another obey Newton's law of gravity. This discovery demonstrates the universality of Newton's law of gravity. Cosmology received an enormous boost when Herschel observed nebulae through his 72-in. reflecting telescope; and he considered these nebulae to be "island universes" of stars. Thomas Wright and I. Kant had previously speculated about such nebulae; Herschel's observations not only verified their speculation but also established extra-galactic astronomy as a new frontier.

14.8.1 Newton and Infinite Universe

The ancients never contemplated the possibility of an infinite universe. Both geocentric and heliocentric systems regarded the universe as having a finite space with the visible star fixed to an outermost sphere around the Earth or the sun. Thomas Digges introduced the concept of infinity to the modern picture of the universe; he dispersed stars in the star sphere in the geocentric and heliocentric systems into an endless infinity of space.

Newton believed Digges' infinite universe ideas. He argued that if the universe was finite or if stars were grouped in one part of the universe, the gravitational forces would soon cause all stars to collapse into a huge clump at the center, but an infinite universe has no center, so it cannot collapse. We know today that Newton's argument for an infinite, static universe does not hold water.

14.8.2 Newton's Law of Gravity Predicts Nonstationary Universe

Actually Newton's own law of gravity predicts a nonstationary universe. To this purpose, let us first introduce an important property of Newton's theory of gravity: a hollow spherically symmetric shell of matter does not create any gravitational field in its interior.

Consider a thin spherical shell of matter as shown in Figure 14.18. We are going to compare gravitational forces that pull a particle of mass *m* (located at an arbitrary point inside the shell) in two opposite directions *a* and *b*. The direction of the line *ab*, passing through *m*, is supposed to be arbitrary, too. The forces of gravitational attraction are created by the matter within the two surface elements cut out from the shell by two narrow cones with equal vertex angles. The areas of the surface elements, cut by these cones, are proportional to the squares of the cone heights. Namely, the ratio of the area S_a of an element *a* to the area S_b of an element *b* is equal to the ratio of the squares of the distance r_a and r_b from *m* to the shell surface along the line *ab*:

$$S_a/S_b = r_a^2/r_b^2. \tag{14.18}$$

If the mass is to be evenly distributed over the shell surface, we arrive at the same ratio for the masses of the surface elements:

$$M_a/M_b = r_a^2/r_b^2. \tag{14.19}$$

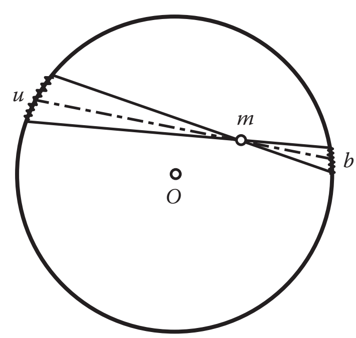


FIGURE 14.18 No gravitational forces inside any spherical shell.

Now, we can calculate the ratio of the forces with which surface elements attract the particle. According to Newton's law, the expressions for these forces are

$$F_a = GM_a m/r_a^2, \quad F_b = GM_b m/r_b^2.$$

Their ratio is given by

$$F_a/F_b = M_a r_b^2/M_b r_a^2. \tag{14.20}$$

Substituting for M_a/M_b in Equation 14.20 its value from Equation 14.19, we finally get

$$F_a/F_b = 1, F_a = F_b.$$

Hence, the two forces are equal in magnitude and act in opposite directions, thus canceling each other out. The argument can be repeated for any other direction. As a result, all forces pulling m in opposite directions cancel one another out, and the resultant force is exactly zero. The location of particle m was arbitrary. Hence, there truly are no gravitational forces inside any spherical shell.

Now, let us turn to the effect of gravitational forces in the universe. The distribution of matter in the universe is homogeneous on large scales. Because we discuss large scales only, we assume the matter to be uniformly distributed in space.

Let us single out of this uniform background an imaginary sphere of an arbitrary radius with the center at an arbitrary point as depicted in Figure 14.19. Consider first the gravitational forces that the matter inside the sphere exerts on the bodies at its surface, ignoring for a moment the effect of matter outside the selected sphere. Let the radius of the sphere be not too large, so the gravitational field generated by its interior is relatively weak and the Newtonian theory of gravity applies. Then, the galaxies at the surface of the sphere are attracted to its center by forces that are directly proportional to the mass M of the sphere and inversely proportional to the square of its radius R.

Next, consider the gravitational effect of all the remaining material in the universe, which lies outside the sphere under consideration. All this matter can be thought of as a sequence of concentric spherical shells with increasing radii, surrounding our selected sphere. But, as we have already shown, spherically symmetric layers of material create no gravitational forces in their interiors. As a result, all the spherical shells (i.e., all the remaining material of the universe) add nothing to the net force attracting some galaxy A at the surface of the sphere toward its center O.

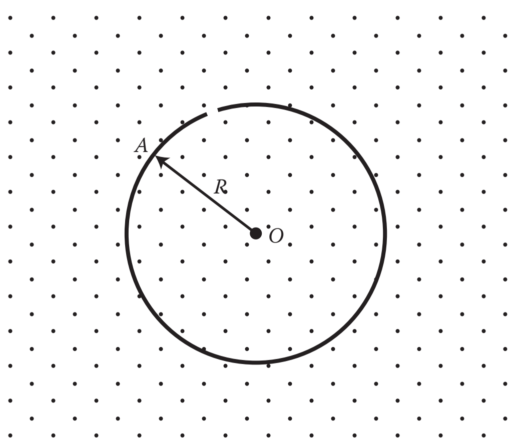


FIGURE 14.19 Nonstationary behavior of the universe, according to Newtonian theory.

So, we can calculate the acceleration of one galaxy A with respect to another galaxy O. We have associated O with the center of the sphere, and A is at a distance R from it. The acceleration to be found is a result of the gravitational attraction of matter inside the sphere of radius R only. According to Newton's law, it is given by

$$a = -GM/R^2. \tag{14.21}$$

The negative sign in Equation 14.21 reflects the fact that the acceleration corresponds to attraction rather than to repulsion.

Thus, any two galaxies in a homogeneous universe separated by a distance R experience a relative acceleration a as given by Equation 14.21. This means that the universe cannot be stationary. Indeed, even if one assumes that at some instant all the galaxies are at rest and the matter density in the universe does not change, the very next moment, the galaxies would acquire some speed as a result of mutual gravitational attraction resulting in the relative acceleration represented by Equation 14.21. In other words, the galaxies could remain motionless with respect to each other only for a brief instant. In the general case, they must move—either approach each other or recede from each other. The radius of the sphere R (see Figure 14.8) must change with time, and so must the density of matter in the universe.

The universe must be nonstationary because gravity acts in it; that is the basic conclusion from the theory. A.A. Friedmann first reached this conclusion in the framework of Einstein's relativistic gravitational theory in 1922 to 1924. Some years later, in the mid-1930s, E.A. Milne and W.H. McCrea pointed out that the nonstationary behavior of the universe can be also derived from Newtonian theory as outlined above.

14.8.3 An Infinite Steady Universe Is an Empty Universe

Furthermore, Newton's law of gravity also predicts that an infinite steady universe is an empty universe; that is, it contains no matter at all. To see this, let us first rewrite Newton's law of gravity

$$F = GMm/r^2$$

in terms of field strength. The field strength, Γ, produced by M at any point in space is defined as the gravitational force that a unit mass would experience if placed at that point. Thus, the force on m, when it is placed in the gravitational field of M, is the product of the field strength Γ and the mass m:

$$F = m\Gamma.$$

Thus,

$$\Gamma = F/m = GM/r^2$$

or

$$\Gamma = \frac{GM}{r^2} = \frac{4\pi GM}{4\pi r^2} = \frac{4\pi GM}{A}$$

where $A = 4\pi r^2$ is the surface area of a sphere of radius r centered at M. From the last equation, we obtain

$$A\Gamma = 4\pi GM.$$

In an infinite universe, this small volume V will be equally attracted in all directions, and so, on the average, Γ vanishes. Then, $M = 0$ as $4\pi G$ cannot be zero; that is, the small spherical volume contains no matter. Because the small volume V is arbitrary, it could be anywhere in the universe, and so $M = 0$ everywhere in the universe. That is, an infinite steady universe contains no matter, an empty universe—obviously not the case with the real universe.

14.8.4 Olbers' Paradox

Digges recognized in 1517 that in a static, infinite universe, the night sky should not be dark. In 1826, Heinrich Olbers detailed his discussion on the dark night puzzle in a paper. Hence, the dark night puzzle is known today as Olbers' paradox. Olbers' paradox tells us that there is something very wrong with the idea of an infinite, static universe.

To see this paradox, we consider a thin spherical shell of thickness t with a center at the observer O and the Earth (Figure 14.20) with an inner radius r; the number of stars (or galaxies if you prefer), N, in the shell is given by

$$N = 4\pi r^2 tn$$

where $4\pi r^2 t$ is the volume of the spherical shell, and n is the number of stars per unit volume. If l is the amount of light emitted by an individual star, then the amount of light emitted by the stars in the shell is given by

$$L = 4\pi r^2 tnl.$$

How much light from the shell will reach the observer O at the center of the spherical shell? Now, the amount of light reaching O from one star in the shell is given by $l/(4\pi r^2)$, so the amount of light the observer O receives from all the stars in the shell is

$$(4\pi r^2 tn) \times \frac{l}{4\pi r^2} = tnl.$$

We see that the radius r cancels out, and the amount of light reaching the observer O at the center from the shell is proportional simply to its thickness t. The same result will apply to any shell

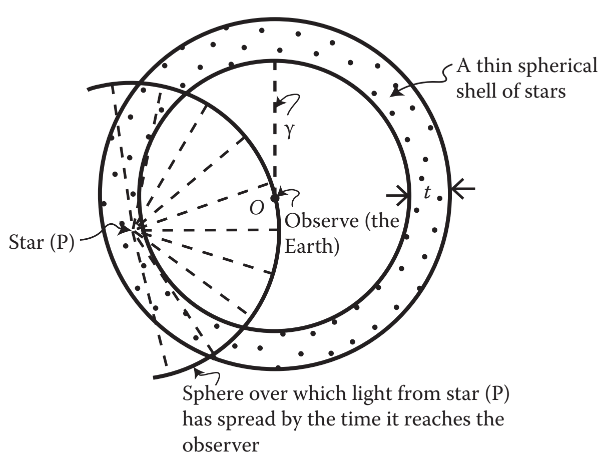


FIGURE 14.20 Olbers's paradox.

centered on O, whatever its radius is. Because there are an infinite number of such shells in an infinite universe, the total light to reach O is infinite.

A possible objection is that there is something like dust between the stars and the Earth, which absorbs the greater part of the light. Today, we know that this explanation is wrong because conservation of energy dictates that the dust must heat up. Eventually, it would begin to glow and finally would reach a state of thermal equilibrium, radiating all the energy it received and glowing as bright as the stars with, thus, no shielding effect.

The resolution to Olbers' paradox is the expansion of the universe. There are two aspects of the expansion of the universe that help: the finite age of the universe and red shift.

The expansion of the universe implies a finite age for the universe. And so the universe has a finite size. This means there is a finite cutoff to the number of shells that can contribute to the sky's brightness. The finite luminous lifetime of the stars should also contribute significantly to the resolution of Olbers' paradox.

The universe is expanding, and so light is red-shifted (the cosmic red shift—we will learn it later).

The red shift has a doubly weakening effect on light. First, because the wavelength of the incoming waves is increased, their frequency is reduced ($f = c/\lambda$); this diminishes their energy according to Planck's formula:

$$E = hf,\ h = \text{Planck's constant.}$$

Second, the lowering of the frequency means that not so many photons (light) arrive in 1 s, so energy received is still further reduced.

It is amazing that great scientists, such as Newton and Einstein, ignored Olbers' paradox and missed the opportunity to realize or conclude that the observable universe has a finite size and age.

14.9 DISCOVERY OF EXPANSION OF THE UNIVERSE, HUBBLE'S LAW

The development of spectroscopy led to many surprising discoveries in astronomy. Over the 15 years from 1912 to 1927, Vesto Slipher managed to obtain the red-shifted spectra of the light from some 39 relatively nearby galaxies. This means that all these 39 galaxies are receding away from us. Later, in 1929, Edwin Hubble determined the distances of the distances of those galaxies and found that the red shift was directly proportional to the distance of the galaxies; the greater the distance to a galaxy, the greater its apparent recession velocity. This proportionality relationship between velocity and galactic distance is known as Hubble's law (or the law of red shift):

$$V = H_0 r \tag{14.22}$$

where V is the speed of recession of a galaxy, r is its distance, and H_0 is a proportionality constant, called Hubble's constant today in honor of Hubble. The implication of Hubble's result was revolutionary: the universe is expanding! (Figure 14.21).

Figure 14.22 is a plot of the recession velocity versus apparent distances for a group of galaxies. The slope of the line is H_0.

The Hubble constant H is one of the most important numbers in all astronomy. It expresses the rate at which the universe is expanding and gives the age of the universe. As we shall see later, H is constant in the sense that it is the same at every place but can change with time. The measurement of the Hubble constant is difficult, and its value is constantly being updated. Currently accepted values of the Hubble constant H_0 fall in the range of 50 to 100 km/s/Mpc, where Mpc is million parsecs, and 1 parsec (pc) = 3.26 light-years. The number used often by many astronomers is 70 km/s/Mpc. That is to say that for each million parsecs distance to a galaxy, the recession speed of the galaxy increases by 70 km/s. The subscript zero on the H indicates that this is the current value.

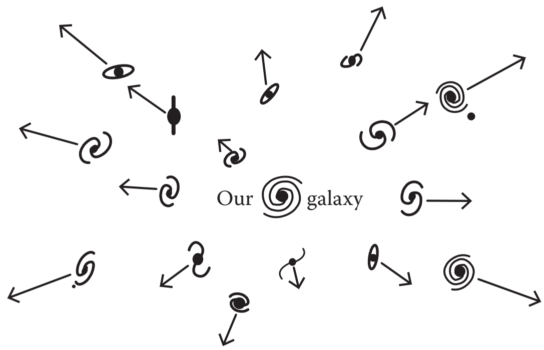


FIGURE 14.21 The universe is expanding.

Hubble's law asserts that the universe is in uniform expansion. By definition, in uniform expansion, $\Delta r/r$ = constant in any given time interval Δt, that is, $\Delta r/\Delta t \propto r$ or $v \propto r$. Because the universe is in uniform expansion, there are no privileged positions, and so an observer moving together with any galaxy sees the surrounding galaxies receding from him or her. As shown in Figure 14.23, consider the motion of galaxy G as seen by O and G'. As seen by O,

$$\vec{v} = H\vec{r}, \ \vec{v}' = H\vec{r}'.$$

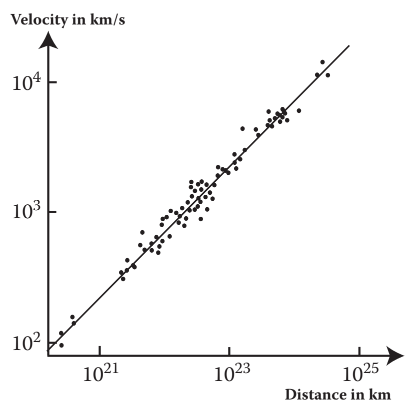


FIGURE 14.22 Plot of the recession velocity versus apparent distances for a group of galaxies.

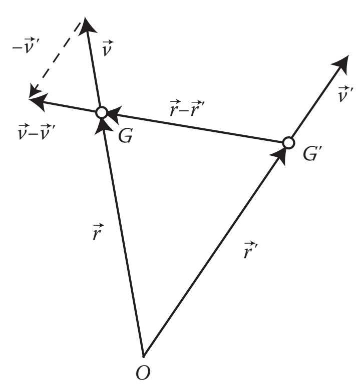


FIGURE 14.23 Expansion as seen by O and G'.

So $\vec{v} - \vec{v}'$, and the velocity of G relative to G', can be expressed as

$$\vec{v} - \vec{v}' = H\vec{r} - H\vec{r}' = H(\vec{r} - \vec{r}').$$

That is, G' also sees G and, therefore, all the other galaxies receding from itself. Although we have used Euclidean geometry and ignored possible changes in H with time, the general result nonetheless holds: each galaxy sees all the others receding from itself, and we are not at the center of this expanding universe.

14.10 BIG BANG

Because the universe is expanding, there must have been a time in the very distant past when everything in the universe came together. Presumably, it exploded and initiated the expansion of the universe. This scenario is called the Big Bang, and a model of the evolution of the universe from such a beginning is known as the Big Bang cosmology. So the phrases "before the Big Bang" or "at the moment of the Big Bang" are meaningless because time did not really exist until after that moment. We can view a Big Bang as an explosion of space at the beginning of time. As time elapses, space itself expands. So the expansion of the universe is the expansion of space. The universe has no center and no edge. If you take a balloon, blow it up, and mark a number of small dots on the surface and then blow it up some more, you will see all the dots move away from each other. This is rather like the expansion of the universe: the expansion of the universe is not by the galaxies moving through space, but that the space between the galaxies is expanding. The three "spatial" dimensions of our universe can be thought of as the two dimensions of the surface of the balloon. A creature can only crawl around the surface, never finding an edge and the center.

14.10.1 Age of the Universe

How long ago the Big Bang took place depends on the value of the Hubble constant (and the models of the Big Bang). We can make a crude estimate by asking how long a distant galaxy has been traveling, assuming that it has had a constant speed since the moment of the Big Bang (it is expected that the rate of expansion reduces as time goes on because the expansion is resisted by the gravitational attraction of the matter in the universe). Using Hubble's law, we see that this time is given by

$$t = \frac{r}{V} = \frac{1}{H_0} = \frac{3.09 \times 10^{22}\ \text{m}}{75 \times 10^3\ \text{m/s}} = 4.1 \times 10^{17}\,\text{s} = 1.3 \times 10^{10}\ \text{years.}$$

This is a rough estimate, but it is in the right order of magnitude.

14.11 FORMULATING DYNAMICAL MODELS OF THE UNIVERSE

There are some pertinent and fundamental questions about the expansion of the universe that need to be answered, such as the following:

(1) Will the expansion of the universe continue forever, or will it eventually reverse itself?
(2) How is the mass density of the universe changing with time?
(3) Is the Hubble constant changing with time?

To answer these and similar questions, we need to construct a model constrained by information revealed by observation. We can learn a lot about the evolution of an expanding universe by applying Newtonian gravitation. A fluid with zero pressure moving according to Newton's laws of motion and gravitation is known as a Newtonian dust. We now consider a Newtonian dust of uniform density ρ(*t*), which is in a state of uniform expansion; the only force on it is gravity. We will work with a co-moving coordinate system, which is carried along with the expansion. Because the expansion is uniform, the position vector $\vec{r}(t)$ of a fluid particle (a galaxy) at any time can be written in terms of the co-moving distance, which we call $\vec{c}$:

$$\vec{r}(t) = s(t)\vec{c} \tag{14.23}$$

where *s*(*t*) is known as the scale factor of the universe, and it is a function of time alone because of the homogeneity of the universe. What we should think of this equation is a coordinate grid that expands with time (Figure 14.24). The galaxies remain at fixed locations in the co-moving coordinate system. The $\vec{r}$ coordinate system does not expand. The scale factor *s*(*t*) measures the universal expansion rate, and it tells us how physical separations are growing with time. Thus, the expansion of the universe is the expansion of the space; galaxies do not expand, and the increase in distances between them is a result of the expansion of space between them. Stars within a galaxy do not expand, and our body does not expand. As velocity is proportional to distance, can faraway galaxies recede from us faster than the speed of light? Distant galaxies can appear to move away faster than the speed of light. Remember, it is space itself that is expanding; thus, there is no violation of the theory of special relativity.

Now, we return to Equation 14.23; differentiation gives

$$\dot{\vec{r}}(t) = \dot{s}\vec{c} = \frac{\dot{s}}{s}\vec{r}(t) = H(t)\vec{r}(t) \tag{14.24}$$

where

$$H(t) = \frac{1}{s(t)}\frac{\mathrm{d}s(t)}{\mathrm{d}t} \tag{14.25}$$

which is the Hubble constant associated with the scale factor, and it is a function of time.

The Newtonian continuity equation, $\partial\rho/\partial t + \nabla\cdot(\rho\dot{\vec{r}}) = 0$, gives

$$\dot{\rho} + 3\rho H = 0$$

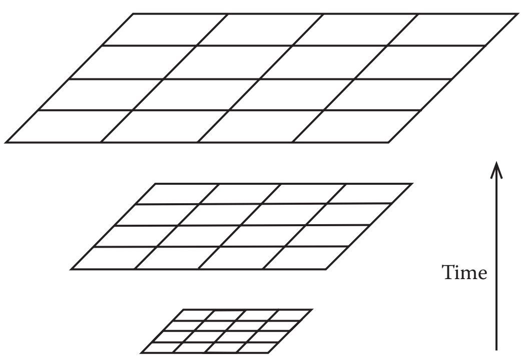


FIGURE 14.24 Newtonian theory: the expansion of the universe is the expansion of the space; galaxies do not expand.

or

$$s\dot{\rho} + 3\dot{s}\rho = 0.$$

Integration gives

$$\rho s^3 = \text{constant}$$

or

$$\rho s^3 = \rho_0 s_0^3 \tag{14.26}$$

where subscript zero indicates current values.

The cosmological principle assumes that the universe is isotropic; that is, the universe appears to be spherically symmetric from any point. Thus, any spherical volume evolves only under its own influence. The gravitational force exerted on the volume by material outside the volume sums to zero. If the volume in question has a radius r, and mass $M(r)$, the equation of motion of a particle of mass m on the surface of the sphere, as position r, is

$$m\ddot{\vec{r}} = -\frac{GmM(r)}{r^2}\hat{r} \tag{14.27}$$

where $\hat{r}$ is a unit vector in the r-direction. Now, in terms of mass density ρ, which is the same everywhere and can change with time, we can write $M(r)$ as

$$M(r) = (4\pi/3)r^3\rho.$$

Substituting this and Equation 14.23 into Equation 14.27, we obtain

$$\ddot{s} = -(4\pi/3)G\rho s. \tag{14.28}$$

This equation becomes, with the aid of Equation 14.26,

$$\ddot{s} = -\frac{4\pi G\rho_0}{3}\frac{1}{s^2} \tag{14.28a}$$

where we take $s_0 = 1$. Note that $\ddot{s}$ cannot be zero unless ρ_0 is zero. Thus, a universe with matter cannot be static; it must be expanding or contracting!

Multiplying both sides of Equation 14.28 by $\dot{s}$, we obtain

$$\dot{s}\ddot{s} + \frac{4\pi G\rho_0}{3}\frac{\dot{s}}{s^2} = 0.$$

This can be rewritten as

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{s}^2 - \frac{8\pi G\rho_0}{3}\frac{1}{s}\right) = 0.$$

Integration gives

$$\dot{s}^2 - \frac{8\pi G\rho_0}{3}\frac{1}{s} = -k \tag{14.29}$$

where k is a constant of integration. This is the standard form of the Friedmann equation that was obtained by Russian mathematician A. Friedmann from the theory of general relativity except that in the original Friedmann equation, k is either ±1 or 0. An expanding universe has a unique value of k, which it retains throughout its evolution. k tells about the geometry of the universe, and it is often called the curvature.

In fact, if $k \neq 0$, there is no loss of generality in taking it to be ±1. Then, we are led to exactly the same three models for the evolution of the universe. We shall not reproduce the three models here. The reader can consult Chow (2007).

One can further integrate Equation 14.29 for the three cases: $k = 0$, $k > 0$, and $k < 0$. But we shall make a qualitative discussion. We may regard the $\dot{s}^2$ term in Equation 14.29 as a measure of its kinetic energy, the $-1/s$ term as a measure of its gravitational potential energy, and $-k$ as a measure of its total energy. If $k < 0$, then $-k$ is positive, the total energy is positive, and the universe has an excess of kinetic energy that allows it to keep expanding. As s gets very large, the second term on the left-hand side approaches zero, and $\dot{s}$ approaches $\sqrt{-k}$. ($-k$ is a positive number.) The expansion continues forever, and we say that the universe is open. If $k = 0$, the total energy is zero, and the kinetic energy is just sufficient to allow the universe to keep expanding, but at a decreasing rate ($\dot{s} \to 0$ as $t \to \infty$ and $s \to \infty$). If $k = 1$, then $-k < 0$, and the universe does not have sufficient kinetic energy for continued expansion. Because the second term on the left-hand side gets smaller when s gets larger, a point will be reached eventually at some finite s, where $\dot{s} = 0$. After that, the universe starts to collapse. We say that the universe is closed. From Equation 14.29, we can find the maximum scale factor s_{max} where $\dot{s} = 0$:

$$-\frac{8\pi G\rho_0}{3}\frac{1}{s_{max}} = -k,$$

or

$$s_{max} = \frac{8\pi}{3}\frac{G\rho_0}{k}.$$

We plot the three possible fates for the universe in Figure 14.25.

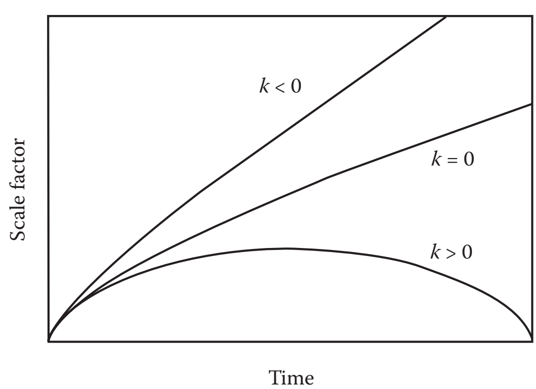


FIGURE 14.25 Three possible fates for the universe (according to Friedmann equation).

14.12 COSMOLOGICAL RED SHIFT AND HUBBLE CONSTANT *H*

In our discussion of Olbers' paradox, we said that the universe is expanding and light is red-shifted on its way to us. And this resolved Olbers' paradox. We can see now why photons from remote galaxies are red-shifted. Imagine a photon coming toward us from a distant galaxy. As a photon travels through the expanding space, its wavelength becomes stretched in proportion to the scale increase. This red shift caused by the expansion of the universe is called a cosmological red shift to distinguish it from a Doppler red shift or gravitational red shift.

Now, consider radiation emitted at wavelength λ_1 at time t_1 from a galaxy G and detected at wavelength λ_0 at time t_0 by a co-moving observer O. The scale of the universe grows from s_1 to s_0 during the time interval that radiation from G travels to observer O. The wavelength of radiation increases during transit by a factor s_0/s_1:

$$\frac{\lambda_0}{\lambda_1} = \frac{s(t_0)}{s(t_1)}.$$

Now,

$$z = \frac{\Delta\lambda}{\lambda_1} = \frac{\lambda_0 - \lambda_1}{\lambda_1}$$

or

$$\frac{\lambda_0}{\lambda_1} = 1 + z$$

and so we have

$$1 + z = \frac{s(t_0)}{s(t_1)}.$$

If G is not very far away from O, we can use the rate of change of s at t_0 to estimate the value of s at t_1:

$$s(t_1) \approx s(t_0) - \dot{s}(t_0)(t_0 - t_1)$$

from which we obtain

$$\frac{s(t_0)}{s(t_1)} = 1 + \frac{\dot{s}(t_0)}{s(t_1)}(t_0 - t_1).$$

On the other hand, we may compute the distance of galaxy G, d, from the time the radiation has taken to traverse it:

$$d = c(t_0 - t_1).$$

Elimination of $(t_0 - t_1)$ from the last two equations, we obtain

$$z = \left[\dot{s}(t_0)/s(t_1)\right](d/c)$$

or

$$cz = v = Hd$$

which is Hubble's law, and we identify

$$H = \dot{s}(t_0)/s(t_1). \tag{14.30}$$

14.13 CRITICAL MASS DENSITY AND FUTURE OF THE UNIVERSE

In Section 14.11, we saw that the future of the universe is determined by the value of k that is not an observable quantity. We now introduce a quantity that determines the future and the geometry of the universe, and it is also closer to observations: the critical mass density ρ_c. Let us start with Equation 14.29, the Friedmann equation, and replace ρ_0 in it by ρ. To this purpose, we can get help from Equation 14.26: $\rho_0 = \rho s^3$ (remember we take $s_0 = 1$), and so the Friedmann equation becomes

$$\left(\frac{\dot{s}}{s}\right)^2 - \frac{8\pi G}{3}\rho = -\frac{k}{s^2}.$$

Combining this equation with Equation 14.30, we obtain

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{s^2}. \tag{14.31}$$

For a given value of H, there is a special value of the density that would be required in order to make the universe marginally bound (the galaxy escape to infinity with no kinetic energy), $k = 0$. This is known as the critical density ρ_c, which we see is given by

$$\rho_c(t) = \frac{3H}{8\pi G}. \tag{14.32}$$

Note that the critical mass density changes with time because H does. For the current value, we have

$$\rho_c = 1.1 \times 10^{-29}\ \text{g/cm}^3 \text{ if } H_0 = 75\ \text{km/s/Mpc}$$

$$\rho_c = 1.1 \times 10^{-30}\ \text{g/cm}^3 \text{ if } H_0 = 50\ \text{km/s/Mpc}.$$

If the average density of matter throughout space is less than the critical density, then gravity is weak, and the expansion of the universe will continue forever; moreover, the universe is open or unbounded. Conversely, if the average density of matter throughout space is larger than the critical density, the resulting gravity will be a strong enough density to eventually halt the expansion of the universe, and the universe is closed or bounded. Separating the two scenarios is the situation in which we have the universe as marginally bounded; in this case, the average density of matter across space exactly equals the critical density, so the galaxies just barely manage to keep moving away from each other. A summary of possible type and destiny (fate) of the universe is given in Table 14.1.

TABLE 14.1
Possible Type and Fate of the Universe

Value of k	Density ρ	Type of Universe	Fate of Universe
> 0	$> \rho_c$	Closed	Big crunch
$= 0$	$= \rho_c$	Marginally bounded	Expands forever
< 0	$< \rho_c$	Open	Expands forever

14.13.1 Density Parameter Ω

Obviously, the critical density sets a natural scale for the density of the universe. So some astronomers prefer to characterize the behavior of the universe in terms of the density parameter Ω, which is defined by

$$\Omega = \rho/\rho_c. \tag{14.33}$$

In general, Ω is a function of time because both ρ and ρ_c depend on time. The present value of the density parameter is denoted Ω_0.

We can express the density parameter Ω in terms of s and H. Let us go back to Equation 14.32:

$$\begin{aligned} H^2 &= \frac{8\pi G}{3}\rho - \frac{k}{s^2} \\ &= \frac{8\pi G}{3}\rho_c\Omega - \frac{k}{s^2} \\ &= H^2\Omega - k/s^2. \end{aligned}$$

Rearranging gives

$$\Omega - 1 = \frac{k}{s^2H^2}\ . \tag{14.34}$$

Then, we can summarize the possibilities:

1. Open universe: $0 < \Omega < 1$, $k < 0$, $\rho < \rho_c$
2. Marginally bounded universe: $\Omega = 1$, $k = 0$, $\rho = \rho_c$
3. Closed universe: $\Omega > 1$, $k > 0$, $\rho > \rho_c$

We see that the case of $\Omega = 1$ is very special because, in that case, k must equal zero, and because k is a fixed constant, Equation 14.34 becomes $\Omega = 1$ for all time. That is true independent of the type of matter the universe contains.

14.13.2 Deceleration Parameter q_0

Another way of determining the future of the universe is to measure the rate at which the cosmological expansion has been gradually slowing down. This slowing can be measured because it causes the relationship between red shift and distance for extremely remote galaxies to deviate from the direct proportion specified by Hubble's law.

Now, suppose that we were to measure the red shifts of galaxies several billion light-years from the Earth. The light from these galaxies has taken billions of years to arrive at our telescope, so what our measurements will reveal is how fast the universe was expanding billions of years ago. If the universe was expanding faster then than it is today, our data will deviate slightly from the straight-line Hubble's law. Astronomers denote the amount of deceleration with the deceleration parameter q_0.

To introduce this deceleration parameter q_0, let us consider a Taylor expansion of the scale factor about the present time:

$$s(t) = s(t_0) + \dot{s}(t_0)(t - t_0) + \frac{1}{2}\ddot{s}(t_0)(t - t_0)^2 + \dots .$$

Dividing through by $s(t_0)$,

$$s(t)/s(t_0) = 1 + H_0(t - t_0) - \frac{q_0}{2} H_0^2 (t - t_0)^2 + \dots$$

where

$$q_0 \equiv -\frac{\ddot{s}(t_0)}{s(t_0)} \frac{1}{H_0^2} . \tag{14.35}$$

The larger the value of q_0, the more rapid the deceleration.

Now, from the acceleration equation 14.28 and the definition of critical density (Equations 14.32 and 14.35), we find

$$q_0 = \frac{4\pi G}{3} \rho \frac{3}{8\pi G \rho_c} = \frac{1}{2}\Omega_0. \tag{14.36}$$

We see that a measurement of q_0 would immediately tell us Ω_0.

The curves in Figure 14.26 display the deviation from the straight-line Hubble's law for different rates of deceleration of the universe. The case $q_0 = 0$ corresponds to no deceleration at all. This is possible if the universe is completely empty and thus has no gravity to slow down the expansion. The case of $q_0 = 1/2$ corresponds to a marginally bounded universe. If q_0 is between 0 and 1/2, the universe is unbounded and will continue to expand forever. Such a universe contains matter at less than the critical density. If q_0 is greater than 1/2, the universe is bounded and is filled with matter having a density greater than ρ_c.

In principle, it should be possible to determine the nature of the universe by measuring the red shifts and distances of many remote galaxies and then plotting the data on a Hubble diagram such as the one in Figure 14.26. If the data points fall above the $q_0 = 1/2$ line, the universe is bounded. If the data points fall between the $q_0 = 0$ and $q_0 = 1/2$ lines, the universe is unbounded. Unfortunately, making these observations of remote galaxies is extremely difficult. The data points of the past measurements are scattered about the $q_0 = 1/2$ line. This scatter prevents astronomers from knowing conclusively whether the universe is bounded or unbounded. Nevertheless, because the data lie close to the $q_0 = 1/2$ line, astronomers suspect that the universe might be marginally bounded.

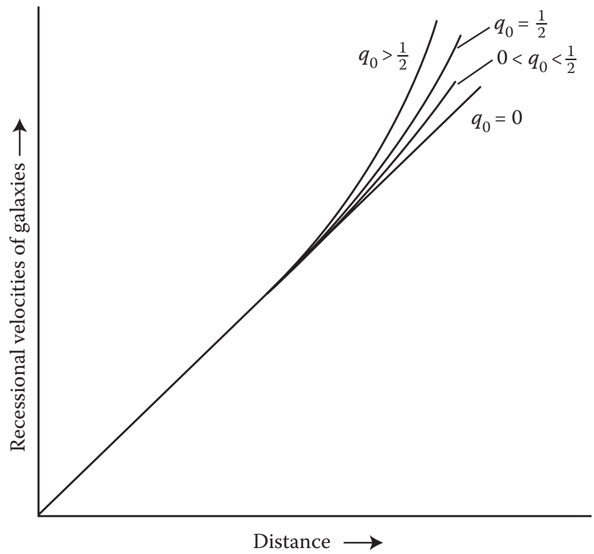


FIGURE 14.26 Deceleration diagram, showing the relationship between recessional velocity and distance for various q_0.

14.13.3 An Accelerating Universe?

Type la supernovae may be a way out of the previously mentioned problem. These supernovae are thought to be more reliable standard candles and have been detected with red shifts almost as great as $z = 1$. In 1998, the first convincing measurements of q_0 was made by two teams of astronomers studying such distant supernovas. Rather surprisingly, the result is that they are farther away than would be expected on the basis of their red shifts and looked about 15% to 20% dimmer than expected. Thus, the expansion of the universe over the past few billion years is speeding up instead of slowing down, and $q_0 < 0$. According to the standard Big Bang, the universe has expanded ever since its explosive birth, but gravity has gradually slowed the expansion. Even if the universe grows forever, the theory predicts it should do so at a steadily decreasing rate. After the initial discovery in 1998, these observations were corroborated by several independent sources. 2011 Nobel Prize in Physics were awarded to Saul Perimutter and Adam Riess of the US and Brian Schmidt of Australia for their discovery. Models attempting to explain accelerating expansion include some form of dark energy. The simplest explanation for dark energy is that it is Einstein's cosmological constant or vacuum energy. It is beyond the scope of this text to discuss these interesting topics.

14.14 MICROWAVE BACKGROUND RADIATION

What is the evidence for the Big Bang? It is the small remnant of radiation left over from the Big Bang, known as the microwave (or cosmic) background radiation.

In the late 1940s, George Gamow and colleagues pointed out that if the universe began with a hot Big Bang, as they thought likely, the blackbody radiation emitted at that time should still be present. The universe has expanded so much since the Big Bang that all short-wavelength photons today have wavelengths that are so stretched that they have become long-wavelength, low-energy photons. This cosmic radiation would look like the radiation emitted by a blackbody at a very low temperature. Gamow predicted a current temperature of about 5 K. At the time Gamow made this prediction, equipment capable of detecting such radiation was not available, so no observation was attempted. But by the early 1960s, Robert Dicke at Princeton University had arrived independently at a prediction of such radiation at 10 K by a different route. And he began designing an antenna to detect this microwave radiation. Finally, Arno Penzias and Robert Wilson of Bell Laboratories (Figure 14.27) detected this small ancient radiation and found the corresponding temperature to be approximately 3 K.

Since those pioneering days, scientists have made many measurements of the intensity of the cosmic background radiation at a variety of wavelengths. The most accurate measurements come from the Cosmic Background Explorer (COBE) satellite, which was placed in orbit in 1989 (Figure 14.28). Data from COBE's spectrometer (Figure 14.28) shows that this ancient radiation has a blackbody spectrum with a temperature of 2.73 K. This radiation field fills all of space, so it is known as the cosmic microwave background.

This cosmic background radiation was released about half a million years after the expansion began at a time when hydrogen atoms had cooled to a temperature of 3000 K, enough to make the universe transparent. Since then, the radiation has cooled to a temperature of 2.73 K today (Figure 14.29).

You may wonder how the cosmic background radiation has a blackbody spectrum while the space is expanding. Let us digress a moment to take a look at the effect of expansion on the spectrum of the cosmic radiation.

We may visualize the effect of the adiabatic expansion of the universe in this way. Imagine a box made of perfectly reflecting mirrors, and blackbody radiation from a hot source is directed to the box. Next, the box is closed so that there can be no leakage of the radiation. The radiation will be trapped and bounce back and forth between the walls. Now, let the walls of the box be slid outward so that the volume of the box increases. As radiation strikes the moving walls, it undergoes Doppler shifts to the red, so the wavelengths of the entire radiation increase. But the wavelengths increase in

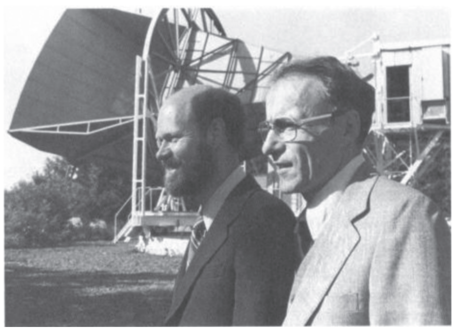


FIGURE 14.27 Arno Penzias and Robert Wilson.

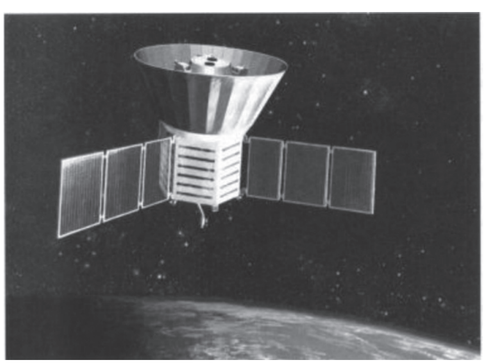


FIGURE 14.28 Cosmic Background Explorer (COBE) satellite.

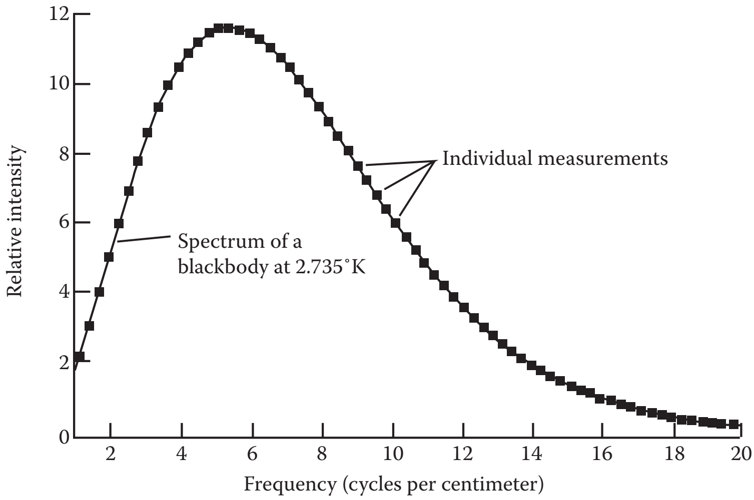


FIGURE 14.29 Data from COBE's spectrometer (a blackbody spectrum with a temperature of 2.73 K).

such a way that the distribution among them still corresponds to the radiation curve for a blackbody. The effect of the moving walls, as the box becomes larger, is to change the radiation from that corresponding to one temperature to that for a blackbody at a lower temperature.

At any given wavelength, the cosmic background radiation is extremely isotropic on the small scale. In directions that differ by a few minutes of arc, any fluctuation in its intensity is less than one part in 10,000. On the other hand, a large-scale anisotropy in the cosmic background radiation has now been established in the sense that it is slightly hotter in one direction than in the opposite direction in the sky. This is a result of our motion through space (Figure 14.30). If we move toward a blackbody, its radiation is Doppler shifted to shorter wavelengths and resembles that from a slightly hotter blackbody. When we move away from it, the radiation appears like that from a slightly cooler blackbody.

The uniformity of the radiation tells us that at an age of less than a million years, the universe was extremely uniform in density. But at least some density variations had to be present to allow matter to gravitationally clump up to form galaxies and super-clusters of galaxies. The isotropy of the microwave background radiation, therefore, puts interesting constraints on theories of supercluster, cluster, and galaxy formation.

COBE's Differential Microwave Radiometer, a set of very sensitive and stable radio receivers, was designed to analyze this problem by mapping the cosmic background far more precisely than

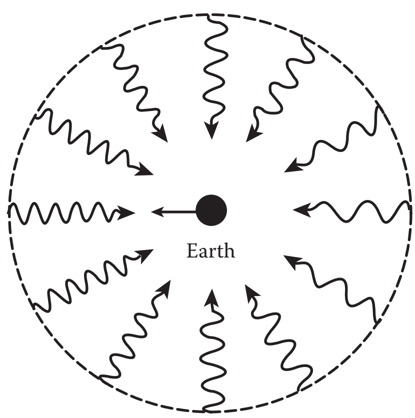


FIGURE 14.30 Our motion through the microwave background.

is possible from the Earth's surface. The data showed that COBE had detected the nonuniformity in the microwave background, amounting to approximately 30 millionths of a Kelvin. This may be sufficient to seed the formation of large-scale structures in the universe, especially if there is a great deal of nonluminous or "dark matter" present in the universe. This invisible matter, whose nature is not yet known, supposedly provided an added gravitational force needed to pull the gas together into galaxies within a reasonable time. We will explore the dark matter problem in the following section.

We can only give account of the evolution of the early universe from Planck's time. Let us see why. According to modern quantum theory, particles have wavelike properties; their positions are somewhat uncertain (because of Heisenberg's uncertainty principle). A particle of mass m typically cannot be precisely located with a distance λ_c (known as the Compton wavelength) given by

$$\lambda_c = h/mc$$

where h is Planck's constant.

The Compton wavelength is a measure of how fuzzy the physical world is at the quantum level. Quantum effects limit our ability to locate a particle of size smaller than its Compton wavelength, which is a characteristic distance over which the particle exists.

On the other hand, general relativity tells us that events that occur within the Schwarzschild radius of an object of mass m are hidden from the outside world. This limiting distance L is approximately

$$L \cong Gm/c^2.$$

Consider the infant universe as a black hole whose mass is so tiny that its Schwarzschild radius is smaller than its Compton wavelength. Because of quantum-mechanical uncertainty, we cannot be sure that the singularity is inside the event horizon. Indeed, there is a probability that the singularity spends some time outside the event horizon. This means that general relativity breaks down. The limiting case of the smallest black hole that does not suffer from this breakdown is obtained by equating L and λ_c. We can solve the resulting equation for mass, known as the Planck mass m_P:

$$m_P = \sqrt{hc/G} = 5.46 \times 10^{-8} \text{ kg}.$$

The limiting distance corresponding to $m = m_P$ is called the Planck length L_P:

$$L_P = (Gh/c^3)^{1/2} \approx 10^{-35} \text{ m}.$$

The time for light to travel across the Planck length is called the Planck time t_P:

$$t_p = L_P/c = Gh/c^5 = 1.35 \times 10^{-43} \text{ s}.$$

Because the theory of general relativity breaks down for times earlier than the Planck time, no one knows how to describe the universe before Planck time. Relativistic space–time is no longer a continuum, and quantum gravity is needed.

Today, the universe is matter-dominated. But the very early universe was actually dominated by radiation. To see this, let us go back to blackbody radiation; its energy density u is

$$u = aT^4.$$

Using Einstein's mass-energy relationship $E = mc^2$, we can convert this energy density into an equivalent mass density ρ_r

$$\rho_r = aT^4/c^2$$

where a is the radiation density constant

$$a = 4\sigma/c$$

and σ is the Stefan–Boltzmann constant = $(5.6697 \times 10^{-8}\ \text{W/m}^2 \cdot \text{K}^4)$. Hence,

$$a = \frac{4 \times (5.6697 \times 10^{-8}\,\text{W/m}^2 \cdot \text{K}^4)}{2.998 \times 10^8\,\text{m/s}} = 7.564 \times 10^{-16}\ \text{W/m}^3 \cdot \text{K}^4.$$

Then ρ_r is equal to

$$\rho_r = \frac{(7.564 \times 10^{-16})(2.7)^4}{2.998 \times 10^8\,\text{m/s}} = 4.5 \times 10^{-31}\ \text{kg/m}^3.$$

This density is much less than that of luminous matter, $\rho_m \approx 4 \times 10^{-28}$ kg/m^3, so the universe is now dominated by matter. Let us trace back by considering the shrinking universe. The matter density increases with the scale length as

$$\rho_m \propto s^{-3}.$$

The radiation density ρ_r increases as T^4. Now, the wavelength of a photon is proportional to s, $\lambda \propto s$. The energy of photon $E = hv = hc/\lambda \propto s^{-1}$, and so $T = E/k \propto s^{-1}$. Finally, we have

$$\rho_r \propto s^{-4}.$$

At some time in the past, the energy density of radiation exceeds that of matter and the universe was radiation-dominated.

14.15 DARK MATTER

To end this chapter, let us discuss briefly the dark matter problem. We saw in Section 14.13 that the dividing line that decides the fate of the universe is the critical mass density ρ_c; if the mass density is greater, the universe is destined to re-collapse, and if not, the universe will expand forever. The visible or "baryonic" matter (matter made of protons and neutrons) that radiates and that we can therefore see as stars, galaxies, and hot gases between stars and between galaxies accounts for approximately 10% of mass in the universe. The rest is in the form of dark matter, matter we cannot see. Recently, physicists and cosmologists have come to believe that the baryonic matter only accounts for about 4% to 5% of the mass in the universe, dark matter makes up 25%, and the rest is something called "dark energy." Some try to associate dark energy with the cosmological constant, which Einstein introduced into his equations for general relativity in order to counteract the force of gravity and therefore to maintain a static universe, and which he abandoned when the universe was shown to be not static, but expanding. Physicists have, if anything, less understanding of dark energy than of dark matter. So we shall not discuss dark energy.

Because dark matter does not shine, it can only be observed gravitationally through dynamic motion or lensing. Dark matter not only decides the fate of the universe; it is also the backbone of galaxy formation.

The distribution of matter during the early universe must not have been perfectly uniform. If it had been, it would still have to be absolutely uniform today; there would now be only a few atoms per cubic meter of space with no stars and no galaxies. Consequently, there must have been slight lumpiness, or density fluctuations, in the distribution of matter in the early universe. Through the action of gravity, these fluctuations eventually grew to become the galaxies and clusters of galaxies that we see today throughout the universe.

Our understanding of how gravity can amplify density fluctuations dates back to 1902, when British physicist James Jeans solved the problem of how the region of higher density gravitationally attracts nearby matter and gas mass. As this happens, however, the pressure of the gas inside these regions will also increase, which can make these regions expand and disperse. The question then becomes the following: Under what conditions does gravity overwhelm gas pressure so that a permanent object can form?

Jeans proved that an object will grow from a density fluctuation provided that the fluctuation extends over a distance that exceeds the so-called Jeans length L_J:

$$L_J = \sqrt{\pi kT/(mG\rho)}$$

where k = Boltzmann constant = 1.38×10^{-23} J/K and T = temperature of the gas (in Kelvin); m = mass of a single particle in the gas; ρ = average density of matter in the gas; and G = universal constant of gravitation = 6.67×10^{-11} N · m^2/kg^2.

Density fluctuations that extend across a distance larger than the Jeans length tend to grow, and fluctuations smaller than L_J tend to disappear.

Observations made with COBE of the nonuniformity in the cosmic background radiation show fluctuations of about 1 part in 10^5 at the time when hydrogen atoms started to form. Such small fluctuations are far too small to form the large structures in the universe we see today. Gravity is not strong enough to grow galaxies from such small fluctuations within reasonable time. Model calculations indicate that, for density fluctuations in the very early universe to collapse to form the galaxies we see today, those fluctuations must be at least 0.2% greater than the average density. If normal matter in the early universe had been clumped at this scale, the variations in the cosmic background radiation today would be at least 30 times larger than what is observed. To solve this discrepancy, we turn once again to dark matter for help.

The nature of the unseen dark matter is not known yet. This does not prevent physicists from hypothesizing different types of dark matter in the hope of explaining the large structure that we see. They distinguish between two basic types of dark matter: the hot dark matter (moving at the speed of light) and the cold dark matter (moving slowly, significantly less than the speed of light). Neutrinos are an example of hot dark matter, and examples of cold dark matter include weakly interacting massive particles (WIMPs) as well as other even more speculative exotic particles.

There are indications, from laboratory experiments and astronomical events (supernova 1987A), that neutrinos may have masses of 10 to 40 eV (for comparison, the electron mass is 0.511 MeV). There are expected to be approximately as many neutrinos and antineutrinos left over from the early universe as there are photons in the microwave radiation background; a neutrino mass of 10 or more electron volts would therefore mean that it is neutrinos rather than nuclear particles that provide most of the mass density of the universe. This will certainly close the universe. However, streaming neutrinos tend to break up small mass concentrations but would leave large ones more or less untouched. Such a selective demolition of certain mass concentrations in the early universe would occur long before radiation decoupling and gravitational collapse. This would mean that the neutrinos had destroyed every nucleus around which galaxies of less than a certain size could condense. Thus, neutrinos (in fact, including other hot dark matter) do not have favorable properties for structure formation.

Cold dark matter avoids this difficulty; small groups of mass come together first, and these small aggregates gather to form large structures. The cold dark matter predicts that galaxies would be created in a rather restricted mass range: from approximately 1/1000 (10^{-1}) to about 10,000 (10^4) times as massive as the Milky Way—not bigger or smaller. It is interesting to note that almost all known galaxies have masses within this range. However, the results of recent red shift surveys and the discovery of the voids and filaments created serious problems for cold dark matter as the ultimate constituent of the structure of the universe.

If neither hot nor cold dark matter can provide all the answers, cosmologists may have to turn to more complex cosmological models, perhaps incorporating both types of dark matter simultaneously or including new physical processes in their simulations.

Dead stars, large planets, and black holes, once thought to be good candidates for dark matter, are now considered unlikely. WIMPs are currently one of the leading suspects. Physicists theorize that the tiny weighty particles (estimated to be 50 times heavier than a proton) originated during the Big Bang, but they only interact weakly with the protons and neutrons of the visible universe. A dozen experiments worldwide are based on the assumption that occasionally, a WIMP might smack into normal matter. The challenge has been to differentiate them from other particles that zip through the cosmos.

Recently, physicists working on DAMA (the Italian Dark Matter Experiment) announced that they possibly have found the elusive particles. The DAMA is a mile underground and uses ultracold detectors that emit flashes of light whenever a particle collides with sodium iodide atoms. The DAMA experiment could possibly differentiate WIMPs from charged particles; it could not distinguish the elusive mystery matter from ordinary neutrons. Because the DAMA experiment is a mile underground, it is shielded from most, but not all, stray neutrons.

Another suspect candidate for dark matter is the so-called massive compact halo objects (MACHOs). They have been detected by gravitational lensing of stars in the Large Magellanic Cloud (LMC). We look from our position in the disk of the galaxy toward LMC stars, which lie outside the galactic halo. If there are invisible compact objects there, and they pass extremely close to the line of sight, then gravitational lensing of the LMC star can occur. This micro-lensing event is very rare, so one has to monitor millions of stars in the LMC every few days for a period of years. After monitoring the brightness of a star in the LMC, astronomers see a temporary brightening that lasted for nearly a month. The favored explanation is gravitational lensing, rather than variability of the star, for several reasons. First, the brightening only happened once, rather than periodically. Second, the brightening is the same in both red and blue light, whereas variable stars brighten differently at different wavelengths. And finally, the symmetric shape of the light curves matches that expected if an invisible gravitational lens were to pass in front of the star. The mass of these invisible objects is estimated as a little less than a solar mass. However, they appear to have insufficient density to completely explain the galactic halo.

Dark matter, what it is made of and exactly how much there is, is still a topic of great interest and a topic of hot pursuit for cosmologists.

PROBLEMS

1. Find the potential at a point outside of, and on the axis of symmetry of, a homogeneous circular disk of radius a, small thickness t, and mass σ per unit volume.
2. Find the gravitational force between a uniformly thin spherical shell of mass M and a particle of mass m located a distance r from its center.
3. The force $\vec{F}$ is specified to satisfy the conditions under which potential energy exists. The force is

$$\vec{F} = F_0(x + by)\hat{i} + F_0 b(x + by)\hat{j}$$

where F_0 and b are constants. Find the potential energy function $V(r)$.
4. A narrow tunnel is bored through a large sphere of mass M along a diameter. Show that the motion of a particle of mass m in the tunnel will oscillate about the midpoint of the tunnel with a period τ given by

$$\tau = (3\pi/G\rho)^{1/2}$$

where ρ is the density of the large sphere.

5. Compute the perihelic motion of a planet if Newton's law of gravitation is replaced by

$$F = \frac{GmM}{r^{2+\delta}}, \quad \delta \ll 1.$$

What value of δ would be necessary to account for the observed advance of Mercury's perihelion? Could this same value of δ account for the advance of the perihelia of other planets? Marshal the best arguments you can for or against such an alteration of Newton's law of gravitation.

6. A galaxy has a recession speed of 13,000 km/s. What is its distance in Mpc?
7. A galaxy at a distance of 300 Mpc has a recession speed of 21,000 km/s. If its recession speed has been constant over time, how long ago was the galaxy adjacent to our galaxy, the Milky Way?
8. Given Hubble constant H_0 = 65 km/s/Mpc, calculate the Hubble time.
9. Calculate the energy density u_r for the cosmic background radiation T = 2.7 K, where $u_r = aT^4$ and $a = 4\sigma/c$ and σ is the Stefan–Boltzmann constant (= 5.6697×10^{-8} W/m$^2 \cdot$ K^4). Convert this energy density to an equivalent mass density ρ_r.
10. A celestial object S is gravitationally lensed by a galaxy G, which is exactly along the line of sight between the object S and the observer O with distances d_1 and d_2 from S and O, respectively. Using small angle approximation, find an expression for the angular size δ (Figure 14.31).
11. Planck's law for the intensity of blackbody radiation is

$$I_\lambda = \frac{2hc^2/\lambda^5}{e^{hc/\lambda kT} - 1}.$$

As the universe expands with a scale factor (radius) $s(t)$, the intensity varies as $I_\lambda \propto s^5$ while the wavelength goes as $\lambda \propto s$.

(a) Show that $T \propto s^{-1}$ if the blackbody formula is to remain valid.

(b) At what wavelength does the blackbody curve reach a maximum for the observed 2.7 K background radiation?

REFERENCE

Chow, T.L. *Gravity, Black Holes, and the Very Early Universe*, Springer, New York, 2007.

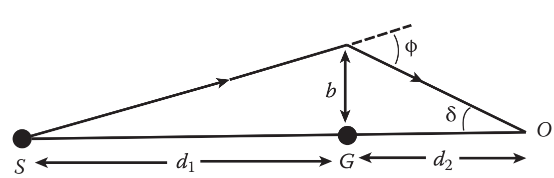


FIGURE 14.31 Celestial object S is gravitational lensed by a galaxy G.

15 Hamilton–Jacobi Theory of Dynamics

The Hamilton–Jacobi (H-J) theory has some advantages in the analysis of periodic motions and is one of the methods developed for more practical problem solving. Before the advent of the modern quantum theory, Bohr's atomic theory was treated in terms of the H-J theory. The H-J theory also played an important role in optics and in shaping up and deriving the Schrodinger equations.

15.1 CANONICAL TRANSFORMATION AND H-J EQUATION

The H-J equation can be derived by many means; the one from canonical transformation theory probably is the simplest. We noted in Chapter 5 that there is some advantage in using canonical transformation. If the equations of motion are simpler in the set of new variables Q_j and P_j than in the old set q_j and p_j, we then have a clear gain. The simplest possible situation would be one in which the new Hamiltonian K vanishes; then the canonical equations in the new system would be

$$\dot{Q}_i = \frac{\partial K}{\partial P_i} = 0 \quad \text{and} \quad \dot{P}_i = -\frac{\partial K}{\partial Q_i} = 0.$$

Thus, the new variables Q_j and P_j would be constant, say, α_I and β_I, respectively.

To find the transformation, we use the fact that $K = 0$. From Equation 5.28, this yields

$$H(q_j, \partial F_1/\partial q_j, t) + \frac{\partial F_1}{\partial t} = 0. \tag{15.1}$$

Now, the total time derivative of $F_1(q_j, Q_j, t)$ is

$$\frac{\mathrm{d}F_1}{\mathrm{d}t} = \frac{\partial F_1}{\partial t} + \sum_j \left[\frac{\partial F_1}{\partial q_j}\dot{q}_j + \frac{\partial F_1}{\partial Q_j}\dot{Q}_j\right]$$

$$= \frac{\partial F_1}{\partial t} + \sum_j \left[p_j \dot{q}_j + \frac{\partial F_1}{\partial Q_j}\cdot 0\right].$$

In the last step, we have used the fact that $Q_j = \alpha_j$ and $\partial F_1/\partial q_j = p_j$. Combining this result with Equation 5.53, we have

$$\frac{\mathrm{d}F_1}{\mathrm{d}t} = \sum_j p_j \dot{q}_j - H = L$$

or

$$F_1 = \int_{t_1}^{t_2} L\,\mathrm{d}t. \tag{15.2}$$

This shows that the generating function F_1 is none other than the action S that appeared in Hamilton's principle:

$$S = \int_{t_1}^{t_2} L\,\mathrm{d}t. \tag{15.3}$$

The action S is also known as Hamilton's principal function. With F_1 replaced by S, Equation 15.1 becomes the H-J equation:

$$H(q_1, q_2, \ldots, q_n, \partial S/\partial q_1, \partial S/\partial q_2, \ldots, \partial S/\partial q_n, t) + \frac{\partial S}{\partial t} = 0. \tag{15.4}$$

Note that we also have

$$p_i = \partial S/\partial q_i. \tag{15.5}$$

Equation 15.4 is known as the H-J equation, a partial differential equation for the Hamilton's principal function S, which is also the generator of a contact transformation in which all the coordinates and momenta are constant. The generator S is supposed to be a function of q_i and Q_i. Because Q_i are constant, they will be a function only of q_i. Q_i are constants of integration found when Equation 15.4 is integrated. We will use a simple example later illustrating the way the constants of integration enter into the solution of the problem. At the moment, we can only say that because S is a function of the n variables q_i and the time, there will be $(n + 1)$ constants of integration. However, as S appears in the H-J equation only in the form of its derivatives, the solution is specified only within an undetermined constant. So one will be an additive constant and will play no rule in the theory. Omitting this one and denoting the remaining n constants by $\alpha_1, \alpha_2, \ldots, \alpha_n$, they are the n constant variables Q_i. The value of P_i can be found from Equation 6.28. Because P_i are constants, we get

$$-\frac{\partial S}{\partial Q_i} = -\frac{\partial S}{\partial \alpha_i} = P_i = -\beta_i \ (i = 1, 2, \ldots, n). \tag{15.6}$$

This relationship (Equation 15.6) enables us to express q_i as a function of the time in terms of $2n$ constants of integration. These constants are then fixed by the initial conditions of the problems. In short, we have a complete solution.

For a conservative system with stationary constraints, the time is not contained explicitly in the function H, and H is a constant, energy E of the system. Then from Equation 15.4, we have

$$\partial S/\partial t = -H(q_i, \partial S/\partial q_i) = -E.$$

That is, the dependence of S on t is expressed by the term $-\alpha_1 t$. And the function S now breaks up into terms, one of which depends only on q_i and the other only on t:

$$S(q_i, t) = S_0(q_i) - Et. \tag{15.7}$$

The function S_0 is often called the contracted action or characteristic function. Inserting Equation 15.7 into Equation 15.4 with H not a function of t, we obtain the H-J equation for the function S_0:

$$H\left(q_i, \frac{\partial \mathrm{S}_0}{\partial \mathrm{q_i}}\right) = \alpha_1 = E. \tag{15.8}$$

Equation 15.8 may be regarded as a partial differential equation for determining S_0 in which the new Hamiltonian is equal to one of the new (constant) coordinates $Q_i = E$. This solution of Equation 15.8 will depend on $n - 1$ additional constants of integration, one of which will be additive. Altogether, it will depend on n nonadditive constants, including the energy; these will correspond to the n constants Q_i. The n constants P_i can again be introduced via Equation 5.28:

$$-\frac{\partial S_0}{\partial Q_i} = -\frac{\partial S_0}{\partial \alpha_i} = P_i = -\beta_i, i = 2,\ldots,n. \tag{15.9}$$

Because the constant Q_1 corresponds to the energy E, we have, from Equation 15.7,

$$-\frac{\partial S}{\partial E} = -\frac{\partial S_0}{\partial E} + t = -\beta_1 \text{ or } \frac{\partial S_0}{\partial E} = t + \beta_1. \tag{15.10}$$

In this way, the equations of coordinates are again expressed as functions of the time in terms of the $2n$ constants (E; $\alpha_2, \ldots, \alpha_n$; $\beta_1, \ldots, \beta_n$). We now summarize the recipe for solving a given problem by means of Equation 15.8, the H-J equation when H is not a function of the time.

(1) Solve Equation 15.6 for S_0 in terms of E and $(n - 1)$, other nonadditive constants of integration: all terms containing q_1 or $\partial S_0/\partial q_i$ are put on one side of Equation 15.8; the other side is a different function of the remaining variables. If the equality is to hold with all the variables independent of one another, each side must be equal to a separation constant, say, α_2:

$$f_1(q_1, \partial S_0/\partial q_1, \alpha_1) = g_1(q_2,\ldots,q_n, \partial S_0/\partial q_2,\ldots,\partial S_0/\partial q_n) = \alpha_2.$$

In this way, we obtain two equations:

$$f_1(q_1, \partial S_0/\partial q_1, \alpha_1) = \alpha_2 \tag{15.11}$$

$$g_1(q_2,\ldots,q_n, \partial S_0/\partial q_2,\ldots,\partial S_0/\partial q_n) = \alpha_2. \tag{15.12}$$

Equation 15.11 is an ordinary differential equation for that part of S_0 and may be solved by the usual methods, giving

$$S_0 = S_1(q_1, \alpha_1, \alpha_2).$$

Repeating the separation process on Equation 15.12, we get another ordinary differential equation for that of S_0, which depends only on q_2. Solving this ordinary differential equation, we obtain

$$\mathrm{S}_0 = S_2(q_2, \alpha_1, \alpha_2, \alpha_3)$$

and so on. Adding all these solutions together, we have for the solution of the H-J equation

$$S_0 = S_1(q_1, \alpha_1, \alpha_2) + S_2(q_2, \alpha_1, \alpha_2, \alpha_3).$$

(2) We set

$$\partial S_0/\partial \alpha_i = \beta_i, i = 2, \ldots, n. \tag{15.13a}$$

(3) We set

$$\partial S_0/\partial\alpha_i = t + \beta_i \ (\alpha_1 = E). \tag{15.13b}$$

(4) Now, S_0 is a function of $q_1,\ldots, q_n$ as well as of $\alpha_1, \ldots, \alpha_n$, and the derivatives of S_0 with respect to the α's are functions of the same arguments. So we have, in Equations 15.13a and 15.13b, a set of n relations between $q_1,\ldots, q_n$, t, and the $2n$ constants $\alpha_1,\ldots, \alpha_n$, and $\beta_1,\ldots, \beta_n$. We may solve these for each q_i, obtaining

$$q_i = q_i(\alpha_1,\ldots,\alpha_n, \beta_1,\ldots,\beta_n, t).$$

The problem has been solved.

In order to acquire some feeling and better understanding of the method of solving the H-J equation, we will offer some illustrative simple examples using the recipe developed above.

Example 15.1

Using the H-J methods to describe the motion of a particle of mass m in the xy-plane, assuming $x = y = 0$ and $\dot{x} = \dot{y} = v_0$ at $t = 0$.

The use of such a powerful method as that of the H-J equation is not quite necessary for such a simple problem; its solution is well known. But an example with uncomplicated algebra and familiar physics will help the reader to gain a better understanding of the procedures employed.

Now the Hamiltonian H of the particle is

$$H = \frac{1}{2m}\left(p_x^2 + p_y^2\right) + mgy = \alpha_1.$$

Replacing p_x and p_y in the Hamiltonian H by $\partial S/\partial x$ and $\partial S/\partial y$, respectively, we obtain the H-J equation:

$$\frac{1}{2m}\left[\left(\frac{\partial S}{\partial x}\right)^2 + \left(\frac{\partial S}{\partial y}\right)^2\right] + mgy = \alpha_1.$$

Because the Hamiltonian H does not contain time explicitly, we assume a separable S:

$$S = S_x(x) + S_y(y).$$

Accordingly, the H-J equation becomes

$$\frac{1}{2m}\left[\left(\frac{\partial S_x}{\partial x}\right)^2 + \left(\frac{\partial S_y}{\partial y}\right)^2\right] + mgy = \alpha_1.$$

Separating the x and y terms, we find

$$\partial S_x/\partial x = [2m(\alpha_1 - mgy) - (\partial S_x/\partial y^2)]^{1/2} = \alpha_2$$

from which we obtain

$$S_x = \alpha_2 x$$

$$\partial S_y/\partial y = \left[2m(\alpha_1 - mgy) - \alpha_2^2\right]^{1/2}.$$

The second equation gives

$$S_y = -\frac{1}{3m^2g}\left[2m(\alpha_1 - mgy) - \alpha_2^2\right]^{3/2}.$$

Adding S_x and S_y together, we obtain the action function S:

$$S = S_x(x) + S_y(y) = \alpha_2 x - \frac{1}{3m^2g}\left[2m(\alpha_1 - mgy) - \alpha_2^2\right]^{3/2}.$$

Now, by Equation 15.13, we have

$$\partial S/\partial\alpha_1 = t + \beta_1 = -\frac{1}{mg}\left[2m(\alpha_1 - mgy) - \alpha_2^2\right]^{1/2}$$

$$\partial S/\partial\alpha_2 = \beta_2 = x + \frac{\alpha_2}{m^2g}\left[2m(\alpha_1 - mgy) - \alpha_2^2\right]^{1/2}.$$

Solving these two equations simultaneously for x

$$x = \frac{\alpha_2}{m}(t + \beta_1) + \beta_2$$

and so

$$\dot{x} = \alpha_2/m.$$

From the initial conditions at $t = 0$, we have

$$0 = (\alpha_1/m)\beta_1 + \beta_2 \text{ and } v_0 = \alpha_1/m$$

and so

$$x = v_0 t.$$

For y, we have

$$y = -\frac{g(t + \beta_1)^2}{2} + \frac{(2m\alpha_1 - \alpha_2^2)}{2m^2g}$$

and

$$\dot{y} = -g(t + \beta_1).$$

Applying the initial conditions at $t = 0$, we finally obtain

$$y = v_0 t - \frac{1}{2}gt^2.$$

Example 15.2: The Central Force Motion

The power of the method becomes more convincing if we apply it to central force motion, where an electron of charge $-e$ revolves about an atomic nucleus of charge Ze.

As the mass of the nucleus is much greater than the mass of the electron, we may consider the electron evolving about a stationary nucleus without making any appreciable error. Employ polar coordinates r and θ in the plane of motion to specify the position of the electron relative to the nucleus (Figure 15.1). The kinetic energy of the electron is

$$T = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right)$$

and its potential energy is

$$V = -\,Ze^2/r.$$

Then the Lagrangian L is

$$L = \frac{1}{2}m\left(\dot{r}^2 + r^2\dot{\theta}^2\right) - Ze^2/r$$

and

$$p_r = \partial L/\partial \dot{r} = m\dot{r}$$

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

The Hamiltonian H is

$$H = \frac{1}{2m}\left(p_r^2 + \frac{1}{r^2}p_\theta^2\right) - \frac{Ze^2}{r}\ .$$

Replacing p_r and p_θ by $\partial S/r$ and $\partial\ \partial S/\partial\theta$, respectively, we obtain the H-J equation:

$$\frac{1}{2m}\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{2mr^2}\left(\frac{\partial S}{\partial \theta}\right)^2 + V(r) + \frac{\partial S}{\partial t} = 0. \qquad (15.14)$$

You may note that we use Equation 15.4, not Equation 15.8. This is equivalent to setting $\partial S/\partial t = -\alpha_1$ in Equation 15.14.

Because the Hamiltonian H does not contain time explicitly, we assume a separable Hamilton's principal function S:

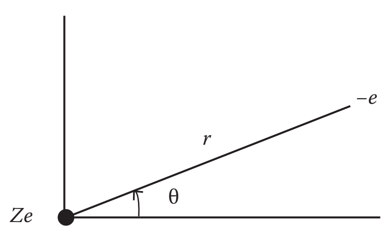


FIGURE 15.1 Motion of an electron in the field of a nucleus.

$$S = S_1(r) + S_2(\theta) + S_3(t).$$

Substituting this into Equation 15.14 gives, after rearrangement,

$$\frac{dS_3}{dt} = \frac{1}{2m}\left[\left(\frac{dS_1}{dr}\right)^2 + \frac{1}{r^2}\left(\frac{dS_2}{d\theta}\right)^2\right] - \frac{Ze^2}{r} \equiv -\alpha_1$$

which gives us two equations:

$$\frac{1}{2m}\left[\left(\frac{dS_1}{dr}\right)^2 + \frac{1}{r^2}\left(\frac{dS_2}{d\theta}\right)^2\right] - \frac{Ze^2}{r} \equiv -\alpha_1 \qquad (15.15)$$

$$dS_3/dt = -\alpha_1. \qquad (15.16)$$

Integrating Equation 15.16, we have

$$S_3 = -\alpha_1 t\theta. \qquad (15.17)$$

We can separate the variables in Equation 15.15 again:

$$\frac{dS_2}{d\theta} = \left[2m\alpha_1 r^2 + 2me^2Zr - r^2\left(\frac{dS_1}{dr}\right)^2\right]^{1/2} = \alpha_2.$$

Thus,

$$\frac{dS_2}{d\theta} = \alpha_2, \; S_2 = \alpha_2\theta. \qquad (15.18)$$

$$\frac{dS_1}{dr} = \left[2m\alpha_1 + \frac{2me^2Z}{r} - -\frac{\alpha_2^2}{r^2}\right]^{1/2}$$

$$S_1 = \int\left[2m\alpha_1 + \frac{2me^2Z}{r} - -\frac{\alpha_2^2}{r^2}\right]^{1/2} dr. \qquad (15.19)$$

Equation 15.18 tells us that the angular momentum p_θ is a constant that is expected because θ is a cyclic coordinate. In fact, whenever a coordinate, such as θ in this example, does not appear explicitly in the H-J equation, we can solve for the conjugate momentum and perform at least one separation of variables.

Adding together the solution, we obtain the function S:

$$S = \int\left[2m\alpha_1 + \frac{2me^2Z}{r} - -\frac{\alpha_2^2}{r^2}\right]^{1/2} dr + \alpha_2\theta - \alpha_1 t. \qquad (15.20)$$

Applying Equations 15.13a and 15.13b, we have

$$\frac{dS}{d\alpha_1} = \int \frac{mr\,dr}{(-\alpha_2 + 2mZe^2r + 2m\alpha_1 r^2)^{1/2}} - t = \beta_1 \qquad (15.21)$$

$$\frac{dS}{d\alpha_2} = \theta - \alpha_2 \int \frac{dr}{r\left(-\alpha_2^2 + 2mZe^2 r + 2m\alpha_1 r^2\right)^{1/2}} = \beta_2. \tag{15.22}$$

We shall evaluate only Equation 15.22, which is the equation of the orbit, with the help of the following standard integral:

$$\int \frac{dx}{x\sqrt{a+bx+cx^2}} = \frac{1}{\sqrt{-a}} \sin^{-1}\left(\frac{bx+2a}{x\sqrt{b^2-4ac}}\right), a < 0.$$

Putting in the values of the constants appearing in Equation 15.22 and performing the integration, we obtain

$$\frac{1}{r} = \frac{me^2Z}{\alpha_2^2} + \sqrt{\frac{m^2e^2Z^2}{\alpha_2^4} + \frac{2m\alpha_1}{\alpha_2^2}} \cos(\theta - \beta_2 + \pi/2).$$

This equation can be further written in the form

$$r = \frac{\varepsilon l}{1 - \varepsilon \cos(\theta - \beta_2 + \pi/2)} \tag{15.23}$$

where

$$\varepsilon l = \alpha_2^2/me^2Z, \quad \varepsilon = \sqrt{1 + 2\alpha_1\alpha_2^2/(me^4Z^2)}. \tag{15.24}$$

Equation 15.23 is the equation of a conic section with eccentricity ε. The eccentricity shows that the orbit is elliptic, parabolic, or hyperbolic according to the energy α_1 being negative, zero, or positive. In the first case, we see that, from Equation 15.23, the major axis of the ellipse is

$$2a = r_{\max} + r_{\min} = \frac{2\varepsilon l}{1-\varepsilon^2}.$$

Substituting εl and ε^2 from Equation 15.24 into the above equation and putting E for the negative energy $-\alpha_1$, we obtain

$$2a = e^2Z/E$$

showing that all elliptical orbits having the same major axes have the same energy irrespective of their eccentricities.

15.2 ACTION AND ANGLE VARIABLES

Many physical problems are of a multiple periodic nature. For example, in the central force motion discussed in Example 15.2, the radius vector r goes through a cycle from minimum to maximum to minimum again as the angle θ varies from 0 to 2π. In the harmonic oscillator, the position coordinate is a periodic function of the time; the momentum, too, oscillates between minimum and maximum values.

The H-J theory is particularly effective in handling such periodic systems. By a periodic motion of the system we mean that the projection of the representative point in phase space on any (p_k, q_k) plane is simply periodic in a one-dimensional sense. That is, the path in the (p, q) plane is a closed curve.

In order to study this cyclic motion, we introduce a special type of variable called the action variable or phase integral. This is defined by the equation

$$J_k = \oint p_k \, dq_k, \; k = 1, 2, \ldots, n \tag{15.25}$$

where the integration is to be performed over one complete cycle of the q_k involved. This amounts to defining J_k for each degree of freedom as the area in its partial phase space bounded by the path of the representative point.

Because, by definition,

$$p_k = \frac{\partial S(q_k, \alpha_1, \alpha_2, \ldots \alpha_n)}{\partial q_k},$$

we can write the action variable associated with the coordinate q_k:

$$J_k = \oint \frac{\partial S(q_k, \alpha_1, \ldots \alpha_n)}{\partial q_k} \mathrm{d}q_k. \tag{15.26}$$

But q_k is only the variable of integration and does not appear in the final result. Hence, each J_k is a function of the α_k, the constants in the solution of the H-J equation:

$$J_k = J_k(\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Each set (p_k, q_k) of momenta and coordinates is independent. Therefore, the various J_k are independent, and we have n equations from which to solve for the α_k. Thus,

$$\left.\begin{array}{l} \alpha_1 = \alpha_1(J_1, J_2, \ldots, J_n) \\ \alpha_2 = \alpha_2(J_1, J_2, \ldots, J_n) \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ \alpha_n = \alpha_n(J_1, J_2, \ldots, J_n) \end{array}\right\}. \tag{15.27}$$

Among α_k's, there is energy E, the separation constant appearing at the very first step:

$$H = \alpha_1 = H(J_1, J_2, \ldots, J_n).$$

Now, with the aid of Equation 15.27, we can eliminate α_k's and write S as a function of the coordinates and of these new quantities, the action variable. That is,

$$S = S(q_1, \ldots, q_n, J_1, \ldots, J_n). \tag{15.28}$$

Now we choose J_k to be the new momenta in the canonical transformation. Then, by Equation 5.32a and 5.32b,

$$p_k = \frac{\partial S}{\partial q_k}, \quad Q_k = \frac{\partial S}{\partial J_k} \equiv w_k \tag{15.29}$$

where we introduce the notation w_k, called the angle variable, to designate the new coordinate. Furthermore, by Equation 5.32c, we have the new Hamiltonian K:

$$K(J_1, J_2, \ldots, J_n) = H(J_1, J_2, \ldots, J_n) = \alpha_1 \tag{15.30}$$

and Hamilton's equation for the new set of canonical variables (w_k, J_k) are

$$\dot{w}_k = \frac{\partial K}{\partial J_k}, \quad \dot{J}_k = -\frac{\partial K}{\partial w_k}, k = 1, 2, \ldots, n. \tag{15.31}$$

Because K is a function of J_k alone, so $\partial J_k/\partial w_k = 0$, and J_k is constant in time. Then, $\partial K/\partial J_k$ is constant in time; we denote this constant by ν_k. The first equation of Equation 15.31 gives, after integration,

$$w_k = \nu_k t + \beta_k \tag{15.32}$$

where β_k is a constant of integration. The constants ν_k, thus far undefined physically, will depend upon J_k through the partial derivatives $\partial K/\partial J_k$.

To examine the physical significance of ν_k, we compute first the change in a given angle variable, say, w_1, resulting from a complete cycle of variation in any one coordinate, say, q_k, while all the other coordinates are fixed. This will be

$$\Delta w_1 = \oint \frac{\partial w_1}{\partial q_k} \mathrm{d}q_k.$$

But, by Equation 15.29, $w_1 = \partial S/\partial J_1$. Hence,

$$\Delta w_1 = \oint \frac{\partial^2 S}{\partial q_k \partial J_1} \mathrm{d}q_k$$

or, reversing the order of the integration and partial differentiation,

$$\Delta w_1 = \frac{\partial}{\partial J_1} \oint \frac{\partial S}{\partial q_k} \mathrm{d}q_k = \frac{\partial}{\partial J_1} \oint p_k \, \mathrm{d}q_k. \tag{15.33}$$

We note that the last integral in Equation 15.33 is J_k. Therefore,

$$\Delta w_1 = \frac{\partial}{\partial J_1} \oint p_k \, dq_k = \frac{\partial J_k}{\partial J_1} = \delta_{1k} = \begin{cases} 1, & \text{if} \quad k = 1 \\ 0, & \text{if} \quad k = 2, \ldots, n \end{cases}$$

so that w_1 increases by unity when q_1 goes through a complete cycle but is not changed by the variation in any other coordinate.

From Equation 15.32, it follows that $\Delta w_1 = \nu_1(\Delta t)$, where Δt is the time required for q_1 to complete its cycle. Hence, ν_1 must be interpreted as the frequency of the motion in the usual reciprocal time sense. In a similar way, each ν_k is found to be the frequency of the motion in the corresponding q_k.

The power of the action-angle method now becomes apparent. We can compute the frequencies of the periodic motions directly without finding the variations of the coordinates with time. We need only express $\alpha_1 = H = K$ in terms of $J_1, J_2, \ldots, J_n$. Then calculate the frequency $\nu_k = \partial K/\partial J_k$ for each k.

Example 15.3: Simple Harmonic Oscillator (SHO) in Angle and Action Variable

The Hamiltonian of an SHO is

$$H = \frac{p^2}{2m} + \frac{1}{2}kq^2 = E = \alpha_1$$

in which k is the spring constant of the oscillator. Solving for momentum p, we obtain

$$p = \sqrt{2m(\alpha_1 - kq^2/2)}$$

from which follows the action variable

$$J = \oint p\,\mathrm{d}q = 4\int_0^{\pi/2} \sqrt{2m(\alpha_1 - kq^2/2)}\mathrm{d}q = 2\pi\alpha_1\sqrt{m/k}. \qquad (15.34)$$

Solving for α_1, we obtain

$$\alpha_1 = \frac{1}{2\pi}\sqrt{\frac{k}{m}}J = K.$$

Hence, by Equation 15.31,

$$\dot{w} = \frac{\partial K}{\partial J} = \frac{1}{2\pi}\sqrt{\frac{k}{m}} = \nu$$

and we recognize that it is the frequency of the oscillations.

The H-J theory has played a pivotal role in the transition from classical to quantum mechanics. The first quantum thoughts of Planck and Einstein concerned the nature of light. Bohr linked this wave-particle nature of light with Rutherford's atomic model, a tiny heavy nucleus surrounded by orbiting electrons. But electron orbits cannot be stable because of the electromagnetic waves emitted by the accelerated orbiting electrons. Bohr temporarily solved this problem by postulating in 1913 that the electrons can only orbit around the nucleus in allowed orbits without radiation; along these stationary orbits, the angular momentum of the electrons is "quantized," that is, it satisfies the simple relationship

$$mvr = n\hbar$$

where n is an integer and labels the orbits with $n = 1$ for the innermost one; $\hbar$ is the Planck constant h divided by 2π.

A. Sommerfeld and W. Wilson, independently, generalized Bohr's quantum conditions to include elliptical orbits. According to them, the motion is limited to such orbits for which the action variables become an integral multiple of h:

$$J_k = \oint p_k\,\mathrm{d}q_k = n_k h,\; n_k = 1, 2, 3, \ldots$$

Combining Sommerfeld–Wilson quantum conditions with the result (Equation 15.34), we find

$$J = \oint p\,\mathrm{d}q = 2\pi E\sqrt{m/k} = nh$$

from which we obtain the energy E of the stationary orbits:

$$E = n\hbar\omega$$

where $\omega = (k/m)^{1/2}$ is the frequency of the oscillator. The quantum mechanical result is $(n + \frac{1}{2})\hbar\omega$, and it differs by the "zero point" energy $\frac{1}{2}\hbar\omega$ that is related to the uncertainty principle. The next example illustrates quantization of the elliptical orbits in Bohr's atomic theory.

Example 15.4: Quantization of the Elliptical Orbits in Bohr's Atomic Theory

The Bohr theory of the hydrogen atom pictures the electron moving about the nucleus in prescribed circular orbits. Sommerfeld and Wilson generalized this to include elliptical orbits. Action variables therefore play an important part in this mathematical model. This is central force motion, and the central force is the attractive Coulomb force between the positive nucleus and the negative electrons. Using plane polar coordinates r and θ, we have

$$J_r = 2\int_{r_1}^{r_2} p_r\,\mathrm{d}r, \quad J_\theta = \int_0^{2\pi} p_\theta\,\mathrm{d}\theta$$

where r_1 is the minimum and r_2 is the maximum value of r during a cycle. We have treated the central force motion of the electron in Example 15.2 and have found

$$p_r = \sqrt{2m[E + Ze^2/r - \alpha_2^2/2mr^2}, \quad p_\theta = \alpha_2.$$

Note that E and α_1 are the same. Hence,

$$J_r = 2\int_{r_1}^{r_2} \sqrt{2m[E + Ze^2/r - \alpha_2^2/2mr^2}\,dr, \quad J_\theta = 2\pi\alpha_2$$

where r_1 and r_2 are the roots of

$$2m[E + Ze^2/r - \alpha_2^2/2mr^2 = 0.$$

Integration (using no. 187 of Peirce's Table of Integrals) yields

$$J_r = \sqrt{(2\pi^2 mZ^2e^4/(-E)} - 2\pi\alpha_2 = \sqrt{(2\pi^2 mZ^2e^4/(-E)} - J_\theta.$$

Solving for E, we obtain

$$E = K = -\frac{2\pi^2 mZ^2e^4}{(J_r + J_\theta)^2}.$$

The frequencies are

$$\nu_r = \frac{\partial K}{\partial J_r} = \frac{4\pi^2 mZ^2e^4}{(J_r + J_\theta)^2} = \frac{\partial K}{\partial J_\theta} = \nu_\theta. \qquad (15.35)$$

The two frequencies involved are equal to one another; the system is said to be degenerate. Degeneracy occurs when a particle undergoes motion around a closed orbit, here an ellipse for negative total energy.

By the Bohr postulates, the action variables (the phase integrals) are quantized; that is, they can be only integral multiples of Planck's constant h. Hence, we write

$$J_r = kh \text{ and } J_\theta = lh \tag{15.36}$$

where k and l are positive integers. From Equation 15.35, we see that the frequencies are also quantized, depending on the principal quantum number $n = k + l$:

$$\nu = \frac{4\pi^2 mZ^2 e^4}{h^3 n^3}.$$

The energy likewise is quantized:

$$E = -\frac{2\pi^2 mZ^2 e^4}{h^2 n^2}.$$

An important property of the canonical transformation is the invariance of the action variables (the phase integrals). The magnitudes of the variables J_k remain the same in any system of canonical coordinates we choose to use. Furthermore, the action variable is an adiabatic invariant. We remind the reader that an adiabatic invariant is a quantity that remains constant if the parameters that describe the system are changed extremely slowly (adiabatically, to use a thermodynamic term), even though they may ultimately be greatly changed. We shall not prove this important property; interested readers can consult the book of Landau and Lifshitz (1969).

Adiabatic invariance played an important part in the development of quantum physics. Ehrenfest showed that adiabatic invariants are quantizable. Thus, the action variables gained great importance. The notion of adiabatic invariance has been also profitably used in other areas of physics. In plasma physics, for example, the adiabatic invariance of J for a charged particle moving in a magnetic field means that the magnetic flux through the circular orbit of the particle will not change in an adiabatic varying magnetic field. As the particle moves to regions of a higher field, it describes circles of progressively decreasing radii to keep the flux constant; the component of velocity perpendicular to the magnetic field thus progressively increases, and the component of velocity in the field direction decreases. If the magnetic field is strong enough, this longitudinal velocity can be reduced to zero and the particles turn back. Thus, by arranging a cylindrical field, which is weaker in the central region than at either of the ends, we obtain a confinement of the charged particles. This is the "magnetic bottle" technique used in thermal nuclear fusion to confine high-temperature plasma.

15.3 INFINITESIMAL CANONICAL TRANSFORMATIONS AND TIME DEVELOPMENT OPERATOR

Consider an infinitesimal canonical transformation in which the new Q's and P's are only slightly different from the old values:

$$Q_k = q_k + \delta q_k,\ P_k = p_k + \delta p_k \tag{15.37}$$

where δq_k and δp_k are the infinitesimal changes in q_k and p_k, respectively. What is the generating function for this infinitesimal canonical transformation? Recall that when discussing examples of canonical transformations (Chapter 5), it was shown that the function $F_2 = \sum_k q_k P_k$ (Equation 5.33a) generates the identity transformation. Therefore, we may write the generating function as

$$F = \sum_k q_k P_k + \varepsilon G(q, P, t) \tag{15.38}$$

where ε is an infinitesimal parameter independent of q_k and p_k, and G is assumed to be a twice differentiable arbitrary function. Then, from Equations 5.32a and 5.32b, the new variables are given by

$$Q_k = \frac{\partial F}{\partial P_k} = q_k + \varepsilon\frac{\partial G}{\partial P_k}, \quad p_k = \frac{\partial F}{\partial q_k} = P_k + \varepsilon\frac{\partial G}{\partial q_k}, \text{ and } K = H + \varepsilon\frac{\partial G}{\partial t}.$$

That is,

$$\delta q_k = \varepsilon\frac{\partial G}{\partial P_k}, \quad \delta p_k = \varepsilon\frac{\partial G}{\partial q_k}. \tag{15.39}$$

Note that if we replace P_k by $p_k + \delta p_k$ in $G(q, P, t)$ and then perform Taylor's expansion on $G(q, p_k + \delta p_k, t)$, the first term is simply $G(q, P, t)$, and the second order is of order ε. Also, note that

$$\frac{\partial G}{\partial P_k} = \sum_j \left(\frac{\partial G}{\partial p_j}\right)\left(\frac{\partial p_j}{\partial P_k}\right) = \sum_j \left(\delta_{jk} + \frac{\partial^2 G}{\partial P_k \partial q_j}\right)\frac{\partial G}{\partial p_j}.$$

Inserting these into Equation 15.39 and neglecting terms of order ε^2, we obtain

$$\delta q_k = \varepsilon\frac{\partial}{\partial p_k}G(q_k, p_k), \quad \delta p_k = -\varepsilon\frac{\partial}{\partial p_k}G(q_k, p_k). \tag{15.40}$$

Now, if the physical system evolves in time through an interval δt, then with the aid of the canonical equations, we have

$$\delta q_k = \dot{q}_k \delta t = \frac{\partial H}{\partial p_k}\delta t, \quad \delta p_k = \dot{p}_k \delta t = -\frac{\partial H}{\partial q_k}\delta t.$$

A quick comparison of these equations with Equation 15.40 shows that the parameter ε can be identified with δt and G with the Hamiltonian H. Thus, the motion of the system in a time interval dt can be described by an infinitesimal canonical transformation generated by the Hamiltonian. Correspondingly, the motion of the system in a finite time interval from t_0 to t is represented by a succession of infinitesimal canonical transformations. As the result of two canonical transformations applied one after the other is equivalent to a single canonical transformation, we see that any two points on a given trajectory in phase space are connected by a canonical transformation. Hence, the motion of a mechanical system can thus be regarded as the continuous unfolding of a canonical transformation generated by the Hamiltonian of the system. From this result follows an important theorem, which can be stated as follows: any invariants during the motion of a system are also invariants under canonical transformation, and conversely, invariants of canonical transformations are invariants of the motion of the system.

In this light, the result of Equation 5.45 in Chapter 5

$$\frac{\mathrm{d}F}{\mathrm{d}t} = [F, H] + \frac{\partial F}{\partial t}$$

is not surprising. The evolution of the system motion is determined by H, and this relationship shows how a particular F develops in time through the influence of H; as a consequence, Equation 5.45 is sometimes called the unfolding theorem. It turns out that there is a neat, practical way of solving dynamical problems by means of this theorem. Expanding $q(t)$ and $p(t)$ as functions of time about their initial values q_0 and p_0, we have

$$q(t) = q_0 + \left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)_{t=0} t + \left(\frac{\mathrm{d}^2 q}{\mathrm{d}t^2}\right)_{t=0} \frac{t^2}{2!} + \ldots$$

$$p(t) = p_0 + \left(\frac{\mathrm{d}p}{\mathrm{d}t}\right)_{t=0} t + \left(\frac{\mathrm{d}^2 p}{\mathrm{d}t^2}\right)_{t=0} \frac{t^2}{2!} + \ldots .$$

Now according to Equation 5.45,

$$\frac{\mathrm{d}q}{\mathrm{d}t} = [q, H] \text{ and } \frac{\mathrm{d}^2 q}{\mathrm{d}t^2} = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\mathrm{d}q}{\mathrm{d}t} = \big[[q,\ H], H\big]$$

and so on. We have similar results for the p-derivative. It follows that

$$q(t) = q_0 + \left(\frac{\mathrm{d}q}{\mathrm{d}t}\right)_{t=0} t + \left(\frac{\mathrm{d}^2 q}{\mathrm{d}t^2}\right)_{t=0} \frac{t^2}{2!} + \ldots$$

$$= q_0 + [q, H]_{t=0} + \frac{1}{2}\big[[q, H],\ H\big]_{t=0} t^2 + \ldots .$$

Defining the operator T by

$$T \equiv \frac{\mathrm{d}}{\mathrm{d}t} = [q,\ H] + \frac{\partial}{\partial t} = \sum_i \left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial p_i}\frac{\partial}{\partial p_i}\right) + \frac{\partial}{\partial t}, \tag{15.41}$$

it follows that

$$q(t) = q_0 + t(T_0 q_0) + \frac{t^2}{2!}\left(T_0^2 q_0\right) + \ldots\ = \left(e^{tT_0}\right) q_0 \tag{15.42}$$

where the subscripts zero indicate that the derivatives must finally be evaluated at time zero. $p(t)$ can be treated in the same way. The exponential e^{tT_0} is known as the time development operator; it guides the evolution of $q(t)$ in time from its initial value q_0.

The time development operator is very useful in seeking solutions as illustrated in the following simple example.

Example 15.5

Consider a harmonic oscillator; its Hamiltonian is

$$H = p^2/2m + \frac{1}{2} m\omega^2 q^2.$$

Find its solution $q(t)$ by means of the time development operator.

Solution:

The operator ***T*** has the form

$$T = \left(\frac{\partial H}{\partial p_i}\frac{\partial}{\partial q_i} - \frac{\partial H}{\partial p_i}\frac{\partial}{\partial p_i}\right) + \frac{\partial}{\partial t} = \frac{p}{m}\frac{\partial}{\partial q} - m\omega^2 q\frac{\partial}{\partial p}.$$

Then,

$$(Tq)_{t=0} = p_0/m,\ (T^2q)_{t=0} = -\omega^2 q,\ (T^3q)_{t=0} = -\omega^2 p/m,\ (T^4q)_{t=0} = -\omega^4 q$$

and so on. Substituting these into Equation 15.42, we obtain

$$\begin{aligned} q(t) &= q_0 + t\frac{p_0}{m} + \frac{t^2}{2!}(-\omega^2 q_0) + \frac{t^3}{3!}\left(-\omega^2\frac{p_0}{m}\right) + \ldots \\ &= q_0\left(1 - \frac{\omega^2 t^2}{2!} + \ldots\right) + \frac{p_0}{m\omega}\left(\omega t - \frac{\omega^3 t^3}{3!} + \ldots\right) \\ &= q_0 \cos\omega t + \frac{p_0}{m\omega}\sin\omega t \end{aligned}$$

with a similar development for $p(t)$.

15.4 H-J THEORY AND WAVE MECHANICS

There is a formal analogy between the trajectory of a particle in a conservative force field and the path followed by a light ray in a region of space where the index of refraction does not vary appreciably over a distance equal to the wavelength of the light in the medium. Under these circumstances, the wave nature of light is not exhibited by the light wave; its propagation through the medium is accurately described by geometric optics; that is, it is described by the propagation of the points on its wave fronts along the ray trajectories normal to the wave fronts. Study of this formal analogy between particle motion in a conservative force field and geometric optics will lead us to guess a transition from classical mechanics to wave mechanics.

We know that light bends when it passes obliquely from a medium of one index of refraction into a medium having a different index of refraction. This bending is described by Snell's law, which states that

$$n_1 \sin\theta_1 = n_2 \sin\theta_2$$

where θ_1 is the angle of incidence and θ_2 is the angle of refraction; n is the index of refraction of the medium as shown in Figure 15.2. For the motion of a particle in a conservative force field, there exists a formula analogous to Snell's law; it is the H-J equation given by Equations 15.4 and 15.5. If the system is conservative, H does not contain time t explicitly.

The action S can be separated as given by Equation 15.7:

$$s(q_k, t) = S_0(q_k) - E_t \tag{15.43}$$

and so

$$p_k = \partial S/\partial q_k = \partial S_0/\partial q_k,\ \text{that is, } \vec{p} = \nabla S_0. \tag{15.44}$$

The H-J equation reduces to the form given by Equation 15.8. If the conservative field is described by $V(r)$, Equation 15.8 becomes

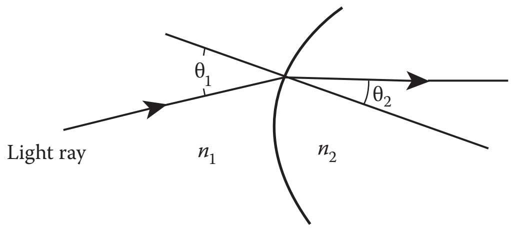


FIGURE 15.2 Snell's law of refraction.

$$\frac{p^2}{2m} + V(r) = \frac{1}{2m}(\nabla S_0)^2 + V(r) = E. \tag{15.45}$$

Solving for ∇S_0,

$$\left|\nabla S_0\right| = \sqrt{2m(E - V)}. \tag{15.46}$$

We can establish the analogy a little further. Consider a particle moving from one region into another region where potential energy changes from V_1 to V_2. A change in potential energy of the particle entails a change in its velocity (Figure 15.3). The change in velocity only occurs in the component of the velocity normal to the surface of discontinuity. But the tangential component remains unchanged, so that

$$V_1 \sin\theta_1 = V_2 \sin\theta_2.$$

The magnitude of the velocity vector in the two regions may be obtained from Equation 15.45, which yields the equation

$$\sqrt{E - V_1}\,\sin\theta_1 = \sqrt{E - V_2}\,\sin\theta_2.$$

It should be clear that the quantity $(E - V)^{1/2}$ plays the same role as the index of refraction.

In these considerations, we have limited to a system of one particle for ease in discussion. But most of the results hold for a many-particle system. Then, we will deal with system points in configuration space, and the motion of a system may be represented by a continuous curve in configuration space. In our one-particle system, this curve will be the actual trajectory of the particle in ordinary space. S_0 = constant represents a set of surfaces, and Equation 15.44 implies that the trajectory of

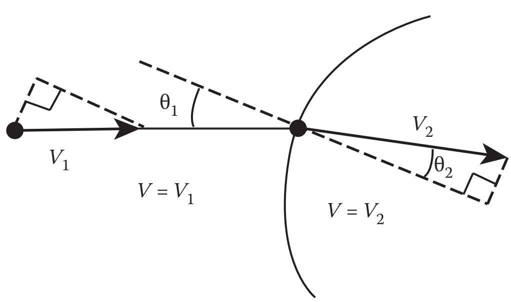


FIGURE 15.3 Law of refraction for the motion of a particle in a conservative force field.

the particle is everywhere normal to such surfaces. This is suggestive of the relationship between wave surfaces and rays in optics. Suppose that the particle motion is, in fact, associated with some form of wave motion in this way. If the wave behavior is represented by a wave function ψ obeying an equation similar to the scalar wave equation in optics, then,

$$\nabla^2\psi - \frac{n^2}{v_0^2}\frac{\partial^2\psi}{\partial t^2} = 0 \tag{15.47}$$

where allowance has been made for a variation of wave velocity from point to point by the incorporation of a refractive index, n, which is a continuous function of position. Confining attention to a single frequency ω,

$$\nabla^2\psi - \frac{n^2\omega^2}{v_0^2}\psi = 0. \tag{15.48}$$

The general solution of this equation is of the form

$$\psi = \psi_0(q_k)\exp i(k_0 f(q_k) - \omega t) \tag{15.49}$$

where $k_0 = 2\pi/\lambda_0 = \omega/v_0$, and ψ_0 is real. We limit our discussion to the case where the variation of the index of refraction is small in a single wavelength. This may be shown to require the condition that

$$(\nabla f)^2 = n^2. \tag{15.50}$$

Surfaces of constant phase are given by

$$k_0 f(q_k) - \omega t = \phi(q_k, t) = \text{constant}. \tag{15.51}$$

This is analogous with Equation 15.43, and it is thus possible to identify surfaces of constant S, which may also be plotted in configuration space and will coincide instantaneously with various S_0 surfaces as shown in Figure 15.4 as wave surfaces of constant phase. From this identification,

$$S = a\varphi,\ S_0 = ak_0 f,\ E = a\omega \tag{15.52}$$

where a is a constant. Equations 15.46, 15.50, and 15.52 now give, as a value for the index of refraction,

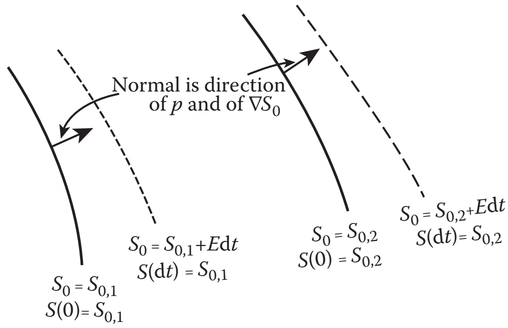


FIGURE 15.4 Sketch of wave surfaces of constant phase.

$$n = |\nabla f| = \frac{1}{ak_0}|\nabla S_0| = \frac{v_0}{\omega a}\sqrt{2m(E-V)}. \tag{15.53}$$

The wave equation 15.48 thus becomes

$$\nabla^2\psi - \frac{2m(E-V)}{a^2}\psi = 0. \tag{15.54}$$

Putting the proportionality constant $a = \hbar (= h/2\pi)$, this equation is recognized as Schrödinger's wave equation for the single particle in a conservative force field.

In the light of the above discussions, can we say that classical mechanics contains within it the seeds of quantum mechanics?

PROBLEMS

1. Solve the Hamilton–Jacobi equation for motion of a particle in a potential $U(r)$ for the following cases:
   a. $U(r) = kx^2/2$ (a one-dimensional harmonic oscillator)
   b. $U(r) = k/r - Fz$ (a combination of a uniform field and a Coulomb field)
2. Solve the problem of a falling body near the surface of the Earth with the H-J method.
3. Describe the motion of a projectile of mass m in the xy-plane by using the Hamilton–Jacobi equation. The initial conditions are $x = y = 0$ and $\dot{x} = \dot{y} = v_0$ at time zero.
4. Solve the problem of a particle in the potential field $V = (m/2)(\omega^2 r^2 + \omega_z^2 z^2)$ with the Hamilton–Jacobi method.
5. Evaluate the integral J_r appearing in Example 15.4.
6. Solve the problem of a damped harmonic oscillator with the H-J method (Moore).
7. Solve the problem of the simple pendulum in action-angle variables assuming small vibration.
8. Consider a two-dimensional harmonic oscillator with unequal spring constants. Find the following:
   a. The Hamiltonian for this oscillator
   b. The momenta and the Hamilton–Jacobi equation
   c. The frequencies, using the method of action-angle variables.
9. Derive Equation 15.8 from Hamilton's principle (Page).

REFERENCE

Landau, L.D., and Lifshitz, E.M. *Mechanics*, Pergamon Press, Oxford, UK, 1969.

16 Introduction to Lagrangian and Hamiltonian Formulations for Continuous Systems and Classical Fields

The dynamic systems we have discussed up to this point have their number of degrees of freedom finite, and they are hence countable. There are some mechanical problems, however, which involve continuous systems as, for example, the problem of a vibrating elastic solid. Here, each point of the continuous solid partakes in the oscillations, and the complete motion can only be described by specifying the position coordinates of all points. The coordinates and momenta of discrete mechanics are replaced by field quantities, that is, functions or fields defined over space and time, which describe the dynamics of the system. It is not difficult to modify the previous formulations of discrete mechanics so as to handle such problems. The most direct method is to approximate the continuous system by one containing discrete particles and then examine the change in the equations describing the motion as the continuous limit is approached.

The study of field theories has been developed as a logical extension to that of continuous material systems. In this chapter, it is proposed to give a brief account of the essential features of such theories.

The best-known example is the electromagnetic field. This may be described in terms either of the electric and magnetic field strengths or of the scalar and vector potential functions; in either case, the quantities concerned are continuously variable functions of space and time. This form of description is based ultimately upon observations of the motions of ordinary material particles postulated to carry electric charges. The idea of a continuous field is introduced in order to avoid the concept of "action at a distance" between the particles. The sources of the field are the charges residing on the particles. The idea is refined and idealized to the extent that the field is considered to exist in some form even in the absence of the particles. The properties of the electromagnetic field are summarized in the set of differential relations known as Maxwell's equations. These will usually be referred to as the field equations.

It is assumed that fields are associated with other types of fundamental particles in the same way that the electromagnetic field is associated with photons. Different types of particles possess different fields that are described by different field equations involving one or more field variables. It is also assumed that the equations must be invariant to Lorentz transformations and thus conform to the relativistic requirement that all the basic laws of nature shall take the same form in all frames of reference. This is already the case with Maxwell's equations, although the development of electromagnetic theory preceded that of special relativity. However, we will not discuss the Lorentz invariance of field equations; it is beyond our scope here.

16.1 VIBRATION OF LOADED STRING

As an example of the transition from a discrete to a continuous system, we consider the problem of an arbitrary number of similar oscillators coupled together. This study will lead us to a description of the oscillations of a continuous medium and naturally into an analysis of wave motions.

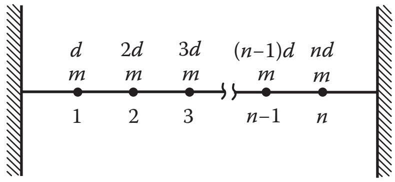


FIGURE 16.1 Loaded string.

For simplicity, consider an approximately one-dimensional distribution: a light elastic string that is fixed at both ends and loaded with a given number n of identical particles, each of mass m, equally spaced a distance d apart along the string. The particles are labeled from 1 to n as shown in Figure 16.1. If we displace the particles and allow them to vibrate vertically in a plane, then we have a set of n-coupled oscillators. Why do we choose transverse oscillations? The treatment for transverse vibrations is quite similar to that for longitudinal vibrations, but transverse oscillations are easier to visualize than longitudinal oscillations, so we choose to study the transverse oscillations.

If we limit the transverse displacements of the particles so they are sufficiently small, then we can ignore any increase in the tension of the string as the particles oscillate. We use as our generalized coordinates the transverse displacements $q_1, q_2, \ldots, q_n$ (see Figure 16.2 for clarity; only three particles are shown). Either Newton's second law or Lagrange's equation can be used to obtain the equations of motion of the particles. We choose to use Lagrange's equations. Now, the kinetic energy of the system is

$$T = \frac{1}{2}m\left(\dot{q}_1^2 + \dot{q}_2^2 + \ldots + \dot{q}_n^2\right). \tag{16.1}$$

For the potential energy of the system, we must first calculate the distance between the kth and $(k + 1)$th particles. In equilibrium, the distance between them is d, but when they are displaced it is

$$d' = d + \delta\, d = [d^2 + (q_{k+1} - q_k)^2]^{1/2} \cong d + \frac{(q_{k+1} - q_k)^2}{2d} + \ldots .$$

The increase in the length of the string connecting the two particles is then approximately

$$\delta d = (q_{k+1} - q_k)^2/2d.$$

This also applies to the sections of the string at each end if we set $q_0 = q_{n+1} = 0$. If S is the tension in the string, the work done against the tension in increasing the length of the string by this amount is $S\,\delta d$:

$$S\delta d = S(q_{k+1} - q_k)^2/2d.$$

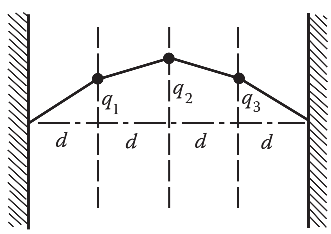


FIGURE 16.2 Transverse motion of a loaded string.

Thus, adding the contributions from each piece of the string, we find that the total potential energy of the system is

$$\begin{aligned} U &= \frac{1}{2}K\left[q_1^2 + (q_2 - q_1)^2 + \ldots + (q_n - q_{n-1})^2 + q_n^2\right] \\ &= \frac{1}{2}K\sum_k (q_{k+1} - q_k)^2 \end{aligned} \tag{16.2a}$$

where

$$K = S/d. \tag{16.2b}$$

The Lagrangian function of the system is

$$L = T - U = \frac{1}{2}\sum_k \left[m\dot{q}_k^2 - K(q_{k+1} - q_k)^2\right]. \tag{16.3}$$

The Lagrange's equations of motion then give

$$\ddot{q}_k = \omega_0^2(q_{k-1} - 2q_k + q_{k+1}),\ k = 1, 2, \ldots, n \tag{16.4}$$

where

$$\omega_0 = \sqrt{K/m} = \sqrt{S/md} \tag{16.5}$$

and remember that $q_0 = 0$ and $q_{n+1} = 0$.

To solve the preceding system of n equations for normal modes of oscillations, we assume that all q_k's oscillate sinusoidally with a common frequency:

$$q_k(t) = A_k e^{i\omega t} \tag{16.6}$$

where the amplitude of vibration A_k can be complex.

Substitution of the preceding solution into Equation 16.4 yields the following recursion formula for the amplitude:

$$(-\omega^2 + 2\omega_0^2)A_k - \omega_0^2(A_{k-1} + A_{k+1}) = 0,\ k = 1, 2, \ldots, n. \tag{16.7}$$

The boundary condition ($q_0 = q_{n+1} = 0$) requires that $A_0 = 0$ and $A_{n+1} = 0$.

Equation 16.7 is a system of n linear homogeneous algebraic equations with the n unknowns A_1, A_2, …, A_n. A nontrivial solution is given by setting the determinant of its coefficients equal to zero:

$$\begin{vmatrix} -\omega^2 + 2\omega_0^2 & -\omega_0^2 & 0 & \ldots & 0 \\ -\omega_0^2 & -\omega^2 + 2\omega_0^2 & -\omega_0^2 & \ldots & 0 \\ 0 & -\omega_0^2 & -\omega^2 + 2\omega_0^2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & -\omega^2 + 2\omega_0^2 \end{vmatrix} = 0.$$

This determinant is of the nth order, and there are n values of ω. Its solution is quite complicated for large values of n. We shall not pretend to solve it but will try to find the n values of ω by working directly with the recursion relation (Equation 16.7). To this purpose, let us assume that the amplitude of particle k is of the form

$$A_k = A\sin(k\phi). \tag{16.8}$$

Here φ is some angle. If a similar equation is used to define the amplitudes of the adjacent particles $k - 1$ and $k + 1$, the recursion relation Equation 16.7 yields

$$(-\omega^2 + 2\omega_0^2)\sin(k\phi) - \omega_0^2[\sin(k-1)\phi + \sin(k+1)\phi] = 0$$

which easily reduces to

$$\omega^2 = \omega_0^2(2 - 2\cos\phi) = 4\omega_0^2\sin^2(\phi/2)$$

or

$$\omega = 2\omega_0\sin(\phi/2). \tag{16.9}$$

Equation 16.7 gives the normal frequencies in terms of the angle φ, which remains to be determined. This we can do by applying the boundary conditions that $A_k = 0$ for $k = 0$ and $k = n + 1$. The former condition is automatically satisfied; the latter requires that $(n + 1)\varphi$ be equal to any integral multiple of

$$(n + 1)\varphi = N\pi,\ N = 1, 2, 3,\ldots. \tag{16.10}$$

Then, the normal frequencies are given by Equation 16.7:

$$\omega_N = 2\omega_0\sin\frac{N\pi\ k}{n+1} \tag{16.11}$$

and the amplitudes for the normal modes are also determined:

$$A_{kN} = A_N\sin\frac{N\pi\ k}{n+1}. \tag{16.12}$$

Here, the value of $k = 1, 2,\ldots, n$ denotes a particular particle in the loaded string, and the value of $N = 1, 2,\ldots$ refers to the normal mode in which the system is vibrating. The different normal modes are illustrated in Figure 16.3 by plotting the amplitudes for the case of three particles.

The actual displacement of the kth particle when the entire string is oscillating in the Nth mode is given by

$$q_{kN}(t) = A_N\sin\frac{N\pi\ k}{n+1}\cos(\omega_N t). \tag{16.13}$$

The general type of motion is a superposition of all normal modes. How many modes are there? We saw that there are just two normal modes for two coupled oscillators. With n linear oscillators, there are only n independent normal modes. We can show this clearly with the help of Equations 16.10 and 16.11. The point is that beyond $N = n$ these equations do not describe any

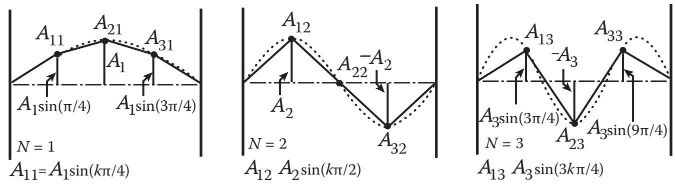


FIGURE 16.3 Three normal modes of a loaded vibrating string.

physically new situations. As we go from $N = 1$ to $N = n$, we find n different distinctive frequencies. At $N = n + 1$, $\omega_{n+1} = 2\omega_0 = \omega_{max}$, but $A_{k(n+1)} = 0$; thus, it does not correspond to a possible motion. For $N = n + 2$, Equation 16.11 gives

$$\omega_{n+2} = 2\omega_0 \sin\frac{(n+2)\pi}{2(n+1)} = 2\omega_0 \sin\left(\pi - \frac{n\pi}{2(n+1)}\right) = 2\omega_0 \sin\frac{n\pi}{2(n+1)} = \omega_n.$$

Similarly, $\omega_{n+3} = \omega_{n-1}$, and so on. And the relative amplitudes of the particles in normal mode also repeat themselves; for example,

$$A_{k,n+2} = A_{n+2} \sin\frac{k(n+2)\pi}{n+1} = A_{n+2} \sin\left(2k\pi - \frac{kn\pi}{n+1}\right) = -A_{n+2} \sin\frac{kn\pi}{n+1} = A_{k,n}.$$

Similarly, $A_{n+3} = A_{n-1}$, and so on.

Having found the number of modes, we can now write down the general type of motion that is of the form

$$q_k = \sum_{N=1}^{n} A_N \sin\frac{N\pi k}{n+1} \cos(\omega_N t - \varepsilon_N). \tag{16.14}$$

The values of A_N and ε_N are determined by the initial conditions.

The normal coordinates of the system $\eta_1, \eta_2, \ldots$ can be introduced according to the scheme described in Equation 8.37 so that

$$q_k = \sum_{N=1}^{n} \eta_N \sin\frac{N\pi k}{n+1}. \tag{16.15}$$

Associated with each normal coordinate η_N is a frequency ω_N given by Equation 16.11. Equation 16.15 is similar to Equation 8.38 except that the quantities C_{jN} are replaced by $\sin[N\pi k/(n + 1)]$, which satisfy the following relationship:

$$\sum_{N=1}^{n} \sin\frac{N\pi k}{n+1} \sin\frac{N\pi s}{n+1} = \frac{1}{2}(n+1)\delta_{ks}, \; k, s = 1, 2, \ldots, n \tag{16.16}$$

which is a relationship similar to the ortho-normalization conditions for the C_{Nk}.

It is interesting to note that as the number n of particles becomes very large compared to the mode number N, the normal frequencies display an integral harmonic relationship. To see this, we let the number n of particles increase but let the spacing d between neighboring particles decrease so that the length of the loaded string remains constant at $D = (n + 1)d$. The total mass of the string, $M = nm$, also remains constant. Thus, as the number n of particles increases, the mass of an individual particle decreases. Now as n becomes very large, we can replace the sine term in Equation 16.11 by the argument

$$\sin\frac{N\pi}{2(n+1)} \cong \frac{N\pi}{2(n+1)}$$

and Equation 16.11 reduces to

$$\begin{aligned}\omega_N &= 2\omega_0 \frac{N\pi}{2(n+1)} = \sqrt{\frac{S}{md}}\frac{N\pi}{n+1} = \sqrt{\frac{S}{\rho}}\frac{N\pi}{(n+1)d} \\ &= \sqrt{\frac{S}{\rho}}\frac{N\pi}{D} = N\frac{\pi}{D}\sqrt{\frac{S}{\rho}} = N\omega_1\end{aligned} \tag{16.17}$$

where the linear density

$$\rho = m/d \quad \text{and} \quad \omega_1 = (\pi/D)\sqrt{S/\rho}. \tag{16.18}$$

ω_1 can be considered to be the fundamental frequency. The other different normal frequencies can then be regarded as the second harmonic, the third harmonic, and so on. We stated at the beginning of this section that the treatment for longitudinal oscillations is quite similar to that for transverse oscillations. We now examine this briefly. Instead of a loaded string, we consider a linear chain: particles of mass m connected by springs, of spring constant k, along a straight line. The springs have the length a if unstrained. If the springs are compressed or stretched, longitudinal oscillations will result. If the displacements of the masses from their equilibrium positions are denoted by $z_1, z_2, \ldots, z_n$ (Figure 16.4), the equation of motion of the pth particle is

$$m\frac{d^2z_p}{dt^2} = k(z_{p+1} - z_p) - k(z_p - z_{p-1})$$

or

$$\frac{d^2z_p}{dt^2} = \omega_0^2(z_{p-1} - 2z_p + z_{p+1})$$

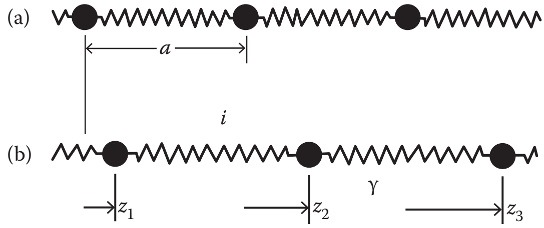


FIGURE 16.4 (a) Spring-coupled masses in equilibrium; (b) spring-coupled masses after small longitudinal displacement.

which has the same form as Equation 16.4. Therefore, mathematically, all the results we have obtained for the transverse oscillations of the loaded string are also valid for the longitudinal oscillations of the linear chain.

The linear chain just described may seem like a very artificial system, but it can be applied to the study of lattice vibrations with rich rewards. For example, a line of atoms in a crystal is surprisingly well represented by such a rudimentary model.

16.2 VIBRATING STRINGS AND THE WAVE EQUATION

16.2.1 WAVE EQUATION

Suppose now that we allow the number of masses to become infinite so that we have a continuous string. As in the last section, we shall let $n \to \infty$, but at the same time let $m \to 0$ and $d \to 0$ in such a way that both the linear mass density $\rho = m/d$ and the total length of the string $D = (n + 1)d$ remain constant. Furthermore, we introduce a new quantity $x = kd$ to specify the distance along the string. Then $N\pi k/(n + 1)$ can be rewritten as

$$\frac{N\pi\, k}{n+1} = \frac{N\pi(kd)}{(n+1)d} = \frac{N\pi\, x}{D}$$

and Equation 16.15 becomes

$$q(t) = \sum_{N=1}^{\infty} \eta_N(t) \sin\frac{N\pi\, x}{D} \tag{16.19}$$

where the sum runs from $N = 1$ to $N = \infty$ because the number of particles is infinite so that there is an infinite set of normal modes. The normal frequencies ω_N also display an integral harmonic relationship and are given by the same equation, Equation 16.17.

It is easy to show that the kinetic energy T and the potential energy U of a continuous string are given by the following equations:

$$T = \sum_N \frac{D\rho\, \dot{\eta}_N^2}{4}, \quad U = \sum_N \frac{D\rho\, \omega_N \eta_N^2}{4}. \tag{16.20}$$

We leave the proof to the reader as a homework problem.

What is the equation of motion for a continuously vibrating string? To answer this question, we return to the loaded string and rewrite its equation of motion, Equation 16.4, as

$$\frac{m}{d}\ddot{q}_k = \frac{S}{d}\left(\frac{q_{k+1} - q_k}{d}\right) - \frac{S}{d}\left(\frac{q_k - q_{k-1}}{d}\right). \tag{16.21}$$

Now as $d \to 0$, we can write

$$\frac{q_{k+1} - q_k}{d} = \frac{q(x+d) - q(x)}{d} = \left.\frac{\partial q}{\partial x}\right|_{x+d/2}$$

and

$$\frac{q_k - q_{k-1}}{d} = \frac{q(x) - q(x-d)}{d} = \left.\frac{\partial q}{\partial x}\right|_{x-d/2}.$$

Consequently, the limiting value of the right-hand side of Equation 16.4 is equal to the second derivative multiplied by d:

$$\lim_{d\to 0}\frac{1}{d}\left(\frac{q_{k+1} - q_k}{d} - \frac{q_k - q_{k-1}}{d}\right) = \lim_{d\to 0}\frac{\partial q/\partial x\big|_{x+d/2} - \partial q/\partial x\big|_{x-d/2}}{d} = \left.\frac{\partial^2 q}{\partial x^2}\right|_{x=kd}.$$

The equation of motion can therefore be written in the form

$$\frac{\partial^2 q}{\partial t^2} = \frac{Sd^2}{m}\frac{\partial^2 q}{\partial x^2} \quad \text{or} \quad \frac{\partial^2 q}{\partial t^2} = \frac{S}{\rho}\frac{\partial^2 q}{\partial x^2}.$$

It is obvious that the dimensions of S/ρ are the dimensions of a squared velocity. If we write $S/\rho = v^2$, the above (or last) equation becomes

$$\frac{\partial^2 q}{\partial t^2} = v^2\frac{\partial^2 q}{\partial x^2}. \tag{16.22}$$

This is the one-dimensional wave equation, a well-known differential equation of mathematical physics.

If the mass of the string is not uniformly distributed, $\rho = \rho(x)$, the tension of the string will also be a function of x, $S = S(x)$. In this case, we should write, as we move to the limit on the right-hand side of Equation 16.4,

$$\lim_{d\to 0}\frac{S(x+d/2)\partial q/\partial x\big|_{x+d/2} - S(x-d/2)\partial q/\partial x\big|_{x-d/2}}{d} = \frac{\partial}{\partial x}\left(S\frac{\partial q}{\partial x}\right)$$

and the wave equation takes the following complicated form:

$$\frac{\partial}{\partial x}\left(S\frac{\partial q}{\partial x}\right) - \rho\frac{\partial^2 q}{\partial t^2} = 0. \tag{16.23}$$

This equation is known as the homogeneous Sturm–Liouville equation and is of considerable interest in many quantum mechanical problems. We shall not pursue its solution but refer interested readers to textbooks on quantum mechanics or mathematical physics.

We now return to Equation 16.22. Its general solutions represent traveling disturbances of some sort. It is easy to verify that a very general type of solution is of the form

$$q = f(x + vt) \text{ or } q = f(x - vt). \tag{16.24}$$

The function $f(x + vt)$ represents a disturbance traveling to the left with a velocity v, and $f(x - vt)$ represents the propagation of a disturbance to the right.

16.2.2 Separation of Variables

One of the simplest and most common methods of solving a partial differential equation such as Equation 16.22 is the method of separation of variables, namely, to seek a solution that is a product of the functions of each of the independent variables. In our case, we assume

$$q(x,t) = X(x)\Phi(t). \tag{16.25}$$

As indicated, X is a function of x only, and Φ is a function of t only. Substitution of Equation 16.25 into Equation 16.22 yields

$$v^2\Phi\frac{\mathrm{d}^2X}{\mathrm{d}x^2} = X\frac{\mathrm{d}^2\Phi}{\mathrm{d}t^2}$$

or, dividing through by the product $X\Phi$,

$$\frac{v^2}{X}\frac{\mathrm{d}^2X}{\mathrm{d}x^2} = \frac{1}{\Phi}\frac{\mathrm{d}^2\Phi}{\mathrm{d}t^2}. \tag{16.26}$$

Now, the left-hand side of Equation 16.26 is a function of x alone, and the right-hand side is a function of t alone. This is impossible unless each side is a constant. For reasons that will be immediately obvious, we let this constant be $-\omega^2$. Then, we have

$$\frac{\mathrm{d}^2X}{\mathrm{d}x^2} + \frac{\omega^2}{v^2}X = 0, \quad \frac{\mathrm{d}^2\Phi}{\mathrm{d}t^2} + \omega^2\Phi = 0. \tag{16.27}$$

These equations are of a familiar form (the first one is often called the Helmholtz equation), and their solutions are

$$X(x) = e^{\pm i(\omega/v)x}, \; \Phi(t) = e^{\pm i\omega t}. \tag{16.28}$$

By multiplying X by Φ and an arbitrary complex constant, using any of the four possible combinations of the signs, we obtain a solution of Equation 16.22. The arbitrary complex constant gives the solution an arbitrary phase and amplitude. We can take the real part or add the solution and its conjugate to get a real solution. The solution we have just found is of the form

$$q(x,t) = A\sin\omega\,(x/v \pm t) \text{ or } q(x,t) = A\cos\omega\,(x/v \pm t).$$

This represents a wave of angular frequency ω traveling along the $-x$-axis (if we have the positive sign) or the $+x$-axis (if we have the negative sign) at a velocity v.

16.2.3 Wave Number and Phase Velocity

It is customary to define a new quantity k by the following relationship:

$$k \equiv \omega/v \tag{16.29}$$

where k is called the wave number (i.e., the number of wavelengths per unit length) or the propagation constant and has dimensions L^{-1}. The wavelength X is given by $v/t = 2\pi v/\omega$; hence, the relationship between k and λ is

$$k = 2\pi/\lambda. \tag{16.29a}$$

In terms of the wave number k, Equation 16.28 takes the form

$$X(x) = e^{\pm ikx},\ \Phi(t) = e^{\pm i\omega t} = e^{\pm ikvt} \tag{16.30}$$

and the solution $q(x,t)$ of Equation 16.22 is a linear combination of the following terms:

$$\begin{aligned} e^{ik(x+vt)} &= e^{i(kx+\omega t)}, \quad & e^{-ik(x+vt)} &= e^{-i(kx+\omega t)} \\ e^{ik(x-vt)} &= e^{i(kx-\omega t)}, \quad & e^{-ik(x-vt)} &= e^{-i(kx-\omega t)}. \end{aligned}$$

The argument of the exponential is called the phase φ of the wave

$$\varphi = kx \pm \omega t. \tag{16.31}$$

If we move our viewpoint along the x-axis at such a velocity that the phase of the wave at every point is the same, then we will see a stationary wave of the same shape. The velocity $W = \mathrm{d}x/\mathrm{d}t$ with which we must move is called the phase velocity of the wave, and it is easy to show that this phase velocity is equal to v, the velocity with which the waveform propagates. As φ = constant, so

$$\mathrm{d}\varphi = k\mathrm{d}x \pm \omega\,\mathrm{d}t = 0$$

and

$$W = \frac{\mathrm{d}x}{\mathrm{d}t} = \left|\frac{\omega}{k}\right| = v. \tag{16.32}$$

16.2.4 Group Velocity and Wave Packets

The wave equation, Equation 16.22, is linear, so the superposition principle holds for its solution. This means that we can build up general solutions by making linear combinations of known solutions. Let us first consider a simple case where we have two almost equal solutions to the wave equation denoted by Q_1 and Q_2, each of which has the same amplitude but differs slightly in frequency and wavelength:

$$Q_1(x, t) = A\cos(\omega t - kx)$$

$$Q_2(x, t) = A\cos[(\omega + \Delta\omega)t - (k + \Delta k)x]$$

where $\Delta\omega \ll \omega$ and $\Delta k \ll k$. Each represents a pure sinusoidal wave extending to infinity along the x-axis. Together they give a resultant wave Q.

$$\begin{aligned} Q &= Q_1 + Q_2 \\ &= A\{\cos(\omega t - kx) + \cos[(\omega + \Delta\omega)t - (k + \Delta k)x]\}. \end{aligned} \tag{16.33}$$

Using the trigonometric identity

$$\cos A + \cos B = 2\cos\frac{A+B}{2}\cos\frac{A-B}{2},$$

Equation 16.33 becomes

$$Q = 2\cos\frac{2\omega\, t - 2kx + \Delta\omega\, t - \Delta kx}{2}\cos\frac{-\Delta\omega\, t + \Delta kx}{2}$$
$$\cong 2\cos\frac{1}{2}(\Delta\omega\, t - \Delta kx)\cos(\omega\, t - kx). \tag{16.34}$$

This represents an oscillation of the original frequency ω but with a modulated amplitude as shown in Figure 16.5. A given segment of the wave system, such as *AB*, can be regarded as a "wave packet" and moves with a velocity V_g (not determined yet). Because $\Delta k \ll k$, this segment contains a large number of oscillations of the primary wave that moves with the phase velocity *V*. The velocity V_g with which the modulated amplitude propagates is called the group velocity and can be determined by the requirement that the phase of the modulated amplitude be constant. Thus,

$$V_g = \frac{\mathrm{d}x}{\mathrm{d}t} = \frac{\Delta\omega}{\Delta k} \to \frac{\mathrm{d}\omega}{\mathrm{d}k}\,. \tag{16.35}$$

The modulation of the wave is repeated indefinitely in the case of superposition of two almost equal waves. However, the technique of Fourier analysis developed in Chapter 8 can be used to demonstrate that any isolated packet of oscillatory disturbance of frequency ω can be described in terms of a combination of infinite trains of frequencies distributed around ω, and the velocity of the packet is given by Equation 16.35. Thus, if we first superpose a system of *n* waves, then we have

$$\psi(x,t) = \sum_{j=1}^{n} A_j e^{i(\omega_j t - k_j x)} \tag{16.36}$$

where the factor A_j denotes the amplitudes of the individual waves. As *n* approaches infinity, the frequencies become continuously distributed. Thus, we can replace the summation in Equation 16.36 with integration, as we did in Chapter 8, and obtain

$$\psi(x,t) = \int_{-\infty}^{\infty} A(k)e^{i(\omega\, t - kx)}\,\mathrm{d}k \tag{16.37}$$

where the amplitude factor $A(k)$ is often called the distribution function of the wave. For $\psi(x, t)$ to represent a wave packet traveling with a characteristic group velocity, it is necessary that the range of propagation vectors *k* included in the superposition be very small. Thus, we assume that the spectral distribution function $A(k)$ is nonzero only for a small range of values about a particular k_0 of *k*. This condition is

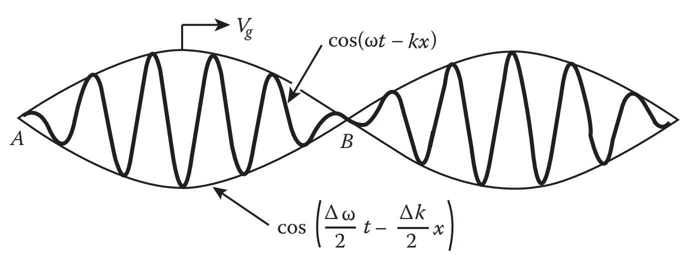


FIGURE 16.5 Superposition of two waves.

$$A(k) \neq 0,\ k_0 - \varepsilon < k < k_0 + \varepsilon,\ \varepsilon \ll k_0. \tag{16.38}$$

The behavior in time of the wave packet is determined by the way in which the angular frequency ω depends upon the wave number k: $\omega = \omega(k)$ (known as the dispersion relationship). If ω varies slowly with k, then $\omega(k)$ can be expanded in a power series about k_0:

$$\begin{aligned}\omega(k) &= \omega(k_0) + \left.\frac{d\omega}{dk}\right|_0 (k - k_0) + \ldots \\ &= \omega(k_0) + \omega'(k - k_0) + \ldots\end{aligned} \tag{16.39}$$

where

$$\omega_0 = \omega(k_0) \quad \text{and} \quad \omega' = \left.\frac{d\omega}{dk}\right|_0$$

and the subscript zero means evaluated at $k = k_0$. We next rewrite the argument of the exponential in Equation 16.37 as

$$\begin{aligned}\omega\, t - kx &= (\omega_0 t - k_0 x) + \omega'(k - k_0)t - (k - k_0)x \\ &= (\omega_0 t - k_0 x) + (k - k_0)(\omega' t - x)\end{aligned}$$

and Equation 16.37 becomes, upon substitution of the preceding result into it,

$$\psi(x,t) = e^{i(k_0 x - \omega_0 t)} \int_{k_0-\varepsilon}^{k_0+\varepsilon} A(k) e^{i(k-k_0)(\omega' t - x)}\, dk. \tag{16.40}$$

Here, the integral depends on x and t only in the combination $x - (d\omega/dk)|_0 t$, and we denote it by $B(x - \omega' t)$. Then, Equation 16.40 becomes

$$\psi(x,t) = B(x - \omega' t) e^{i(k_0 x - \omega_0 t)} \tag{16.41}$$

which is in the form of a product of an envelope function B and a plane wave. The requirement of constant phase for the amplitude (the envelope function B) now leads to

$$V_g = \frac{dx}{dt} = \omega' = \left.\frac{d\omega}{dk}\right|_0. \tag{16.42}$$

The group velocity is of great physical importance. Because every wave train has a finite extent, what we usually observe is the motion of a group of waves except in very rare cases where we follow the motion of an individual wave crest.

In a dispersing medium, wave speed varies with wavelength. Thus, the group velocity differs from the phase velocity in a dispersing medium. It is very easy to show this. Because $k = \omega/v$, we first rewrite Equation 16.42 as

$$\frac{1}{V_g} = \left.\frac{dk}{d\omega}\right|_0$$

from which we obtain

$$\frac{1}{V_g} = \left.\frac{\mathrm{d}k}{\mathrm{d}\omega}\right|_0 = \left.\frac{\mathrm{d}(\omega/v)}{\mathrm{d}\omega}\right|_0 = \frac{v_0 - (\omega\ \mathrm{d}v/\mathrm{d}\omega)\big|_0}{v_0^2}$$

or

$$V_g = \frac{v_0}{1-(\omega_0/v_0)(\mathrm{d}v/\mathrm{d}\omega)\big|_0} \tag{16.43}$$

only for a non-dispersing medium, where $v = W$ = constant, $V_g = v_0 = W$. The subscript zero still means "evaluated" at $k = k_0$ (or, equivalently, at $\omega = \omega_0$).

The concept of a wave packet played an important role in the development of quantum mechanics. In 1925, de Broglie proposed a hypothesis that the same dualism of wave and corpuscle that is present in light and other electromagnetic radiation may also occur in matter. According to this proposal, a moving material particle will have a wave, called the matter wave or de Broglie wave, associated with it. Instead of individual waves, however, de Broglie suggested that we can think of the particle inside a wave packet with the entire packet traveling at the particle's velocity (the term "wave packet" is actually due to Erwin Schrödinger). de Broglie demonstrated that the group velocity of the wave packet is equal to the particle's velocity. Guided by an idea based on the special theory of relativity, de Broglie found that the wavelength λ of the matter wave associated with a moving particle of momentum p is $\lambda = \hbar/p$, where $\hbar$ is Planck's constant divided by 2π. This tells us at once why de Broglie waves only manifest themselves at the atomic level. On the macroscopic level, all dimensions are so enormous, compared to the de Broglie wavelength, that wave aspects are totally undetectable. At the atomic and subatomic levels, the dimensions become comparable to the de Broglie wavelength, and the wave aspects dominate. The matter waves idea inspired Schrödinger to formulate, develop, and complete the wave mechanics (one of the mathematical formalisms of quantum mechanics).

It should be emphasized that although the same mathematical techniques are used to study de Broglie's matter wave packets and the classical wave packets, the physical content of the matter waves is totally different from that of classical waves, such as waves on a string. It is beyond the scope of this book to discuss this further here.

16.3 CONTINUOUS SYSTEMS AND CLASSICAL FIELDS

16.3.1 Lagrangian Formulation

We saw in the previous section that a discrete loaded string can be converted into a continuous one by taking the limit $d \to 0$. As a result, the integer index i identifying the particular mass point becomes the continuous position coordinate x, and the variable q_i gets replaced by the variable $q(x)$. The quantity

$$\frac{q_{i+1}-q_i}{d} = \frac{q(x+d)-q(x)}{d} \to \frac{\mathrm{d}q}{\mathrm{d}x} \tag{16.44}$$

as d, now playing the role of dx, approaches zero. Also in the same limit, $m/d \to \rho$ (the mass per unit length of the continuous string) and

$$d \to x \sum (\ )d \to \int (\ )\mathrm{d}x. \tag{16.45}$$

The Lagrangian function of Equation 16.3 becomes

$$L = \frac{1}{2}\int\left[\rho\,\dot{q}^2 - K\left(\frac{\mathrm{d}q}{\mathrm{d}x}\right)^2\right]\mathrm{d}x = \int L_d\,\mathrm{d}x \tag{16.46}$$

where

$$L_d = \frac{1}{2}\left[\rho\,\dot{q}^2 - K\left(\frac{\mathrm{d}q}{\mathrm{d}x}\right)^2\right] \tag{16.47}$$

which is the Lagrangian density of the system.

It is important to note that the variable x is not a generalized coordinate; it only serves the role of a continuous index previously played by the discrete integer i. In the discrete case, each value of i was a different one of the generalized coordinate q_i. Now, for each value of x, there is a generalized coordinate $q(x)$. We should write the generalized coordinate as $q(x, t)$ because q is also a function of time t. Subsequently, it will be necessary to differentiate functions that are dependent on x and t through their explicit dependence on q and its derivatives. The more general convention to be used then is

$$\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial}{\partial t} + \frac{\mathrm{d}q}{\mathrm{d}t}\frac{\partial}{\partial q} + \frac{\mathrm{d}q,x}{\mathrm{d}t}\frac{\partial}{\partial q,x} + \frac{\mathrm{d}\dot{q}}{\mathrm{d}t}\frac{\partial}{\partial \dot{q}},\ q,x = \frac{\partial q}{\partial x}.$$

For the one-dimensional continuous string, the generalized coordinate can be distinguished by one continuous index x: $q(x)$. For a three-dimensional continuous system, three continuous indices, say, x, y, z, would distinguish the generalized coordinates, so we write $q(x, y, z, t)$. And the corresponding Lagrangian density would be

$$\iiint L_d\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z. \tag{16.48}$$

From Equation 16.46, we see that L_d for the elastic string is a function of $\dot{q} = \partial q/\partial t$ and $\partial q/\partial x$, where both x and t are to be treated on an equal footing. We therefore replace the total differential symbols by the partial differential ones. In some cases, L_d can be a function of q itself and may also involve x and t in an explicit manner. That is,

$$L_d = L_d(q, \partial q/\partial x, \partial q/\partial t, x, t).$$

Similarly, for any general three-dimensional continuous system, the Lagrangian density of the system will have the following functional form:

$$L_d = L_d(q, \partial q/\partial x, \partial q/\partial y, \partial q/\partial z, \partial q/\partial t, x, y, z, t). \tag{16.49}$$

It is important to note from the preceding discussion that it is the Lagrangian density £, rather than the Lagrangian L itself, that becomes the relevant function in describing the motion of the continuous system. As we saw in Chapters 4 and 5, the Lagrangian and Hamiltonian formalisms are based upon the possibility of constructing a Lagrangian L such that L fulfills Hamilton's principle:

$$\delta\int_{t_1}^{t_2} L\,\mathrm{d}t = 0.$$

In view of Equation 16.46, this may be written in terms of the Lagrangian density L_d as

$$\delta\int\iiint L_d\,\mathrm{d}x\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}t = \delta\int\iiint L_d\,\mathrm{d}V\,\mathrm{d}t = 0 \tag{16.50}$$

where dV is a volume element. As in Chapters 4 and 5, the variation of the integral is carried out by varying the path of integration in such a way that keeps the endpoints of the path fixed and varies the values of q at all other points, but keeping x, y, z, and t fixed. Note that the variation of q is taken to be zero not only at the endpoints but also on the surface of the volume of integration.

As we shall show a little later, Hamilton's principle is fulfilled if the Lagrangian density L_d satisfies the following differential equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_d}{\partial \dot{q}} + \sum_{i=1}^{3}\frac{\partial}{\partial x_i}\frac{\partial L_d}{\partial(\partial q/\partial x_i)} - \frac{\partial L_d}{\partial q} = 0 \tag{16.51}$$

where we have made the change of notation (x, y, z) (x_1, x_2, x_3). Equation 16.51 is a modified Lagrange's equation for a continuous system. To derive it, we follow the familiar procedures of variational analysis of Chapter 4 by labeling the varied paths with parameter α and remembering that

$$\delta \to d\alpha\frac{\partial}{\partial\alpha}. \tag{16.52}$$

Now the variation of L_d can be written as

$$\delta L_d = \frac{\partial L_d}{\partial q}\delta q + \frac{\partial L_d}{\partial \dot{q}}\delta\dot{q} + \sum_{i=1}^{3}\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\delta\left(\frac{\partial q}{\partial x_i}\right) \tag{16.53}$$

and Hamilton's principle (Equation 16.50) gives

$$\int\iiint\left[\frac{\partial L_d}{\partial q}\delta q + \frac{\partial L_d}{\partial \dot{q}}\delta\dot{q} + \sum_{i=1}^{3}\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\delta\left(\frac{\partial q}{\partial x_i}\right)\right]\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}t = 0. \tag{16.54}$$

Integrating the second term by parts, we obtain

$$\int_1^2\frac{\partial L_d}{\partial\dot{q}}\delta\dot{q}\mathrm{d}t = \left[\frac{\partial L_d}{\partial q}\delta q\right]_1^2 - \int_1^2\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L_d}{\partial\dot{q}}\right)\delta q\mathrm{d}t = -\int_1^2\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L_d}{\partial\dot{q}}\right)\delta q\mathrm{d}t \tag{16.55}$$

where conditions $(\delta q)_1 = (\delta q)_2 = 0$ have been used. Similarly, we simplify the integrals involving the spatial derivatives of q:

$$\begin{aligned}\int\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\delta\left(\frac{\partial q}{\partial x_i}\right)\mathrm{d}x_i &= \int\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\frac{\partial}{\partial x_i}(\delta q)\,\mathrm{d}x_i\\ &= \left[\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\delta q\right] - \int\frac{\mathrm{d}}{\mathrm{d}x_i}\left(\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\right)\delta q\mathrm{d}x_i\\ &= -\int\frac{\mathrm{d}}{\mathrm{d}x_i}\left(\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\right)\delta q\mathrm{d}x_i.\end{aligned} \tag{16.56}$$

Again, conditions $(\delta q)_1 = (\delta q)_2 = 0$ have been used. In Equation 16.56, we have written a total time derivative with respect to x_i in order to emphasize the fact that the derivative involves not only the explicit dependence of £ on x_i but also the implicit dependence on x_i through q. Substituting Equations 16.55 and 16.56 into Equation 16.54, we obtain

$$\int\iiint \delta q\left[\frac{\partial L_d}{\partial q}-\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_d}{\partial \dot{q}}-\sum_{i=1}^{3}\frac{\mathrm{d}}{\mathrm{d}x_i}\frac{\partial L_d}{\partial(\partial q/\partial x_i)}\right]\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}t = 0. \tag{16.57}$$

Now, the variations $q(x_1, x_2, x_3, t)$ are completely arbitrary, so Equation 16.57 is satisfied if and only if the integrand vanishes identically:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_d}{\partial \dot{q}}+\sum_{i=1}^{3}\frac{\mathrm{d}}{\mathrm{d}x_i}\frac{\partial L_d}{\partial(\partial q/\partial x_i)}-\frac{\partial L_d}{\partial q}=0 \tag{16.58}$$

which is identical to Equation 16.51.

The reader who has been carefully following the details up to now might be amazed as to why a continuous system with an infinite number of degrees of freedom has only one equation of motion, while a discrete system with n degrees of freedom has n Lagrange's equations of motion. We should remember that the generalized coordinate q is a function of continuous parameters x_1, x_2, x_3, and t, so we can say that Equation 16.58 furnishes a separate equation of motion for each set of values of x_i. In other words, for a continuous system, we have a field equation, and the generalized coordinate q is usually called the field variable.

We can generalize the previous results to more complicated continuous systems described by more than one generalized coordinate. For example, consider the vibrations of a solid. A particle will now undergo displacements along all three directions x_1, x_2, x_3. So there will be three generalized coordinates $q_i(x_1, x_2, x_3, t)$ $i = 1, 2, 3$. Accordingly, the Lagrangian density L_d becomes a function of all the generalized coordinates and their space–time derivatives:

$$L_d = L_d\left(q_i, \dot{q}_i, \partial q_i/\partial x_i, x_i, t\right). \tag{16.59}$$

For each of the generalized coordinates q_i, there will be an equation of motion of the form obtained in Equation 16.58:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L_d}{\partial \dot{q}_k}+\sum_{i=1}^{3}\frac{\mathrm{d}}{\mathrm{d}x_i}\frac{\partial L_d}{\partial(\partial q_k/\partial x_i)}-\frac{\partial L_d}{\partial q_k}=0. \tag{16.60}$$

It should be noted that q_i need not be displaced as it was in the continuous string. For example, for a Lagrangian description of the electromagnetic field, we use the scalar potential φ and the three components of the vector potential $\vec{A}$ as the four generalized coordinates q_i, $i = 1, 2, 3, 4$, and the q_i are usually called the field variables.

16.3.2 Hamiltonian Formulation

The Hamiltonian formulation for continuous systems follows the same scheme as was given in Chapter 5. For the discrete case, conjugated to each generalized coordinate q_i, there is a generalized momentum:

$$p_i = \frac{\partial L}{\partial \dot{q}_i} \tag{16.61}$$

which for the loaded string is

$$p_i = \frac{\partial L}{\partial \dot{q}_i} = d\frac{\partial L_i}{\partial \dot{q}_i} \tag{16.62}$$

where

$$L_i = \frac{1}{2}\left[\frac{m}{d}\dot{q}_i^2 - Kd\left(\frac{q_{i+1} - q_i}{d}\right)^2\right]. \tag{16.63}$$

The Hamiltonian for the system can be defined in terms of p_i and L_i:

$$H = \sum_i p_i\dot{q}_i - L_i = \sum_i d\left(\frac{\partial L_i}{\partial \dot{q}_i}\dot{q}_i - L_i\right) \tag{16.64}$$

in the limit as $d \to 0$, $q_i \to q(x)$, and

$$L = \sum_i L_i \to \int L_d\,\mathrm{d}x. \tag{16.65}$$

Equation 16.64 becomes

$$H = \int \mathrm{d}x\left(\frac{\partial L_d}{\partial \dot{q}}\dot{q} - L_d\right). \tag{16.66}$$

But

$$p_i = d(\partial L_i/\partial \dot{q}_i) \to dx(\partial L_d/\partial \dot{q}_i)$$

vanishes in the continuous limit. So we have to find a way out. One way is to define a momentum density π conjugate to q:

$$\pi = \partial L_d/\partial \dot{q}_i \tag{16.67}$$

which remains finite. With the introduction of π, Equation 16.66 becomes

$$H = \int \mathrm{d}x(\pi)\dot{q} - L_d \equiv \int \mathrm{d}x\,\hbar \tag{16.68}$$

where

$$\hbar = \pi\dot{q} - \pounds. \tag{16.69}$$

It is important to bear in mind that in the Hamiltonian theory, q_i and π_i are all treated as equally independent variable functions of x_i and t.

16.3.3 Conservation Laws

Much of the formal development of Hamiltonian formalism, conservation theorems, Poisson brackets, and so forth can also be carried out for continuous systems. In this section, we state a number of conservation laws.

In the Hamiltonian theory, when the Lagrangian L does not depend explicitly on time t, then the conservation of energy is expressed by the equation $\mathrm{d}H/\mathrm{d}t = 0$. We expect that the application of this conservation law to the total Hamiltonian equation 16.68, which represents the total energy of the field (the system), will lead to the interpretation of the Hamiltonian density $\hbar$ as an energy density of the system, for which a continuity equation holds:

$$\partial\hbar/\partial t + \nabla\cdot\vec{S} = 0 \tag{16.70}$$

where the quantity S_i is defined as

$$S_i = \sum_j \frac{\partial L_d}{\partial(\partial q_j/\partial x_i)}\dot{q}_j. \tag{16.71}$$

It is easy to show that

$$\begin{aligned}\frac{\partial\hbar}{\partial t} + \sum_i \frac{\partial S_i}{\partial x_i} &= \sum_k \dot{q}_k\left(\frac{\partial}{\partial t}\frac{\partial L_d}{\partial\dot{q}_k} + \sum_{i=1}^{3}\frac{d}{dx_i}\frac{\partial L_d}{\partial(\partial q_k/\partial x_i)} - \frac{\partial L_d}{\partial q_k}\right)\\ &+\left[\sum_k\left(\frac{\partial L_d}{\partial q_k}\dot{q}_k + \frac{\partial L_d}{\partial\dot{q}_k}\ddot{q}_k + \sum_{i=1}^{3}\frac{\partial L_d}{\partial(\partial q_k/\partial x_i)}\frac{\partial\dot{q}_k}{\partial x_i}\right) - \frac{\partial L_d}{\partial t}\right] = 0.\end{aligned} \tag{16.72}$$

The first term vanishes because of Equation 16.51, and the second does the same if £ is not an explicit function of t.

Equation 16.70 is an energy density continuity equation (a density conservation law), with $\vec{S}$ as a vector representing the flow of energy and $\hbar$ as the energy density of the system. Note that $\vec{S}$ is not uniquely defined in Equation 16.71 because Equation 16.71 is satisfied by any vector $\vec{S}' = \vec{S} + \nabla\times\vec{A}$, where $\vec{A}$ is an arbitrary vector. We usually take $\nabla\times\vec{A} = 0$.

If L_d were to depend explicitly on t, the second term in Equation 16.72 would not vanish. Thus, the equation expresses a transfer of energy between our system and some other coexistent system.

On the same pattern, let us define a momentum density vector of the field (system):

$$G_i = -\sum_k \frac{\partial L_d}{\partial\dot{q}_k}\frac{\partial q_k}{\partial x_i} \tag{16.73}$$

and a stress tensor

$$T_{ik} = \sum_l \frac{\partial L_d}{\partial(\partial q_l/\partial x_i)}\frac{\partial q_l}{\partial x_k} - L_d\delta_{ik} \tag{16.74}$$

where the familiar Kronecker delta symbol is 1 or 0 depending on the subscripts. We shall demonstrate that momentum is conserved:

$$\frac{\partial G_k}{\partial t}+\sum_i \frac{\partial T_{ik}}{\partial x_i}=0 \tag{16.75}$$

provided L_d is not an explicit function of x_k. The stress tensor T_{ik} characterizes the flux of momentum through the system.

To show Equation 16.75, let us calculate the time derivative of G_k:

$$\begin{aligned}\frac{\mathrm{d}G_k}{\mathrm{d}t}&=-\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i \frac{\partial L_d}{\partial \dot{q}_i}\frac{\partial q_i}{\partial x_k}\right]=-\sum_i\left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial L_d}{\partial \dot{q}_i}\right)\frac{\partial q_i}{\partial x_k}+\frac{\partial L_d}{\partial \dot{q}_i}\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial q_i}{\partial x_k}\right]\\&=-\sum_i\left[\dot{\pi}_i\frac{\partial q_i}{\partial x_k}+\frac{\partial L_d}{\partial \dot{q}_i}\frac{\partial \dot{q}_i}{\partial x_k}\right].\end{aligned} \tag{16.76}$$

From Equations 16.60 and 16.67, we obtain

$$\dot{\pi}_i=\frac{\partial L_d}{\partial q_i}-\sum_i\frac{\mathrm{d}}{\mathrm{d}x_k}\frac{\partial L_d}{\partial(\partial q_i/\partial x_k)}\,.$$

Using this equation in Equation 16.76, we obtain

$$\begin{aligned}\frac{\mathrm{d}G_k}{\mathrm{d}t}&=-\frac{\mathrm{d}L_d}{\mathrm{d}x_k}+\frac{\partial L_d}{\partial x_k}+\sum_{i,j}\frac{\mathrm{d}}{\mathrm{d}x_j}\left(\frac{\partial L_d}{\partial(\partial q_i/\partial x_j)}\frac{\partial q_i}{\partial x_k}\right)\\&=\frac{\partial L_d}{\partial x_k}+\sum_j\frac{\mathrm{d}}{\mathrm{d}x_j}\left(\sum_i\frac{\partial L_d}{\partial x_k}+\sum_{i,j}\left(\frac{\partial L_d}{\partial(\partial q_i/\partial x_j)}\right)\frac{\partial q_i}{\partial x_k}-\delta_{jk}L_d\right).\end{aligned} \tag{16.77}$$

If L_d is not an explicit function of x_k, then $\partial L_d/\partial x_k = 0$, and Equation 16.77 reduces to Equation 16.75, and we have the proof.

Similarly, we can define the angular momentum density for the field, compared to the usual definition of the angular momentum:

$$M_{ij}=x_iG_j-x_jG_i$$

and it proceeds in much the same way that we might obtain the conservation law for this angular momentum density.

16.4 SCALAR AND VECTOR OF FIELDS

To end this chapter, we briefly discuss types of fields, especially scalar and vector fields.

16.4.1 Scalar Fields

The simplest type of field is that having a single real scalar field variable φ. Consider a simple Lagrangian density L_d that has the form

$$L_d=\sum_\mu\left(\frac{\mathrm{d}\phi}{\mathrm{d}x_\mu}\right)^2+k^2\phi^2\ (\mu=1,2,3,4).$$

From Equation 16.51, we find that the associated equation of motion is

$$\sum_{\mu}\frac{d^2\phi}{dx_{\mu}^2}-k^2\phi=0.$$

That is,

$$(\Box - k^2)\varphi = 0$$

where

$$\Box=\nabla^2-\frac{1}{c^2}\frac{d^2}{dt^2}$$

is the d'Alembertian operator.

In quantum theory, such a field can be shown to describe the wave aspects of neutral particles having a proper (rest) mass $m = kh/2\pi c$. Neutral pions (π mesons) are thought to be associated with a related type of pseudo-scalar field differing only in that φ changes sign under space inversion, that is, $\phi(\vec{r},t)=-\phi(-\vec{r},t)$. The effect of this modification is apparent only in interaction with other fields.

It has been shown that a complex field variable $\varphi = \varphi_1 + i\varphi_2$ is associated with electrically charged particles.

16.4.2 Vector Fields

A vector field has as field variables the components of a four-vector. The electromagnetic field is an example of this type of field.

The electromagnetic field in the absence of matter is described by Maxwell's equations for free space with charge and current densities everywhere zero:

$$\nabla\cdot\vec{B}=0 \quad \nabla\times\vec{E}=-\frac{1}{c}\frac{d\vec{B}}{dt} \tag{16.78}$$

$$\nabla\cdot\vec{E}=0 \quad \nabla\times\vec{B}=\frac{1}{c}\frac{d\vec{E}}{dt}\ . \tag{16.79}$$

The magnetic field is perpendicular to the electric field, and they are always in phase. So one field determines the other; we need only discuss one of them explicitly, and we choose the electric field. This leaves Equation 16.79 as the field equations proper.

The electromagnetic fields $\vec{E}$ and $\vec{B}$ can be expressed in terms of electromagnetic potentials $\vec{A}$ and φ:

$$\bar{B}=\nabla\times\vec{A} \quad \vec{E}=-\nabla\phi-\frac{1}{c}\frac{d\vec{A}}{dt}\ .$$

The four components (A_1, A_2, A_3, $A_4 = i\varphi$) may be identified as the components of a four-vector from the relationship (the Lorentz gauge condition)

$$\nabla\cdot\vec{A}+\frac{1}{c}\frac{d\phi}{dt}=0.$$

Using this relationship, Equation 16.79 may be rewritten in the relativistically invariant form:

$$\sum_{\mu} \frac{d}{dx_\mu}\left(\frac{dA_\mu}{dx_\nu} - \frac{dA_\nu}{dx_\mu}\right) = 0.$$

Alternatively, if we define an antisymmetric second-order tensor $F_{\mu\nu}$ by

$$F_{\mu\nu} = \frac{dA_\mu}{dx_\nu} - \frac{dA_\nu}{dx_\mu},$$

the field equations take the form

$$\sum_{\mu} \frac{dF_{\mu\nu}}{dx_\mu} = 0.$$

It can be shown that these equations may be deduced from Hamilton's principle:

$$\delta \int L_d d^4x = \delta \int \iint L_d \, dx_1 \, dx_2 \, dx_3 \, dx_4 = 0$$

with

$$L_6 = \sum_{\mu}\sum_{\nu} (F_{\mu\nu})^2 = \sum_{\mu,\nu}\left(\frac{dA_\mu}{dx_\nu} - \frac{dA_\nu}{dx_\mu}\right)^2.$$

Other types of vector field have also been considered. In particular, it has been shown that, as in the scalar case, complex field components imply that the associated particles are electrically charged. A further generalization, based on nonclassical considerations, is that the particles associated with any vector field have an intrinsic angular momentum of amount $h/2\pi$ (unit spin). This is in contrast to scalar fields, which are associated with particles of zero spin.

PROBLEMS

1. Suppose that two waves with slightly different frequencies and wavelengths are propagating together through a medium. For simplicity, let us assume that their amplitude and initial phases are equal. Then they can be represented as

$$\psi_1 = A\cos(\omega t - kx) \; \psi_2 = A\cos[(\omega + d\omega)t - (k + dk)x]$$

where $dk \ll k$, and $d\omega \ll \omega$.

(a) Neglecting second-order differentials but not first-order differentials, show that the resultant wave is given by

$$\psi = \psi_1 + \psi_2 \cong 2\cos\left(\frac{d\omega}{2}t - \frac{dk}{2}x\right)\cos(\omega t - kx).$$

This represents an oscillation of the original frequency ω but with a modulated amplitude.

(b) Determine the group velocity v_g by inspection of the preceding equations.

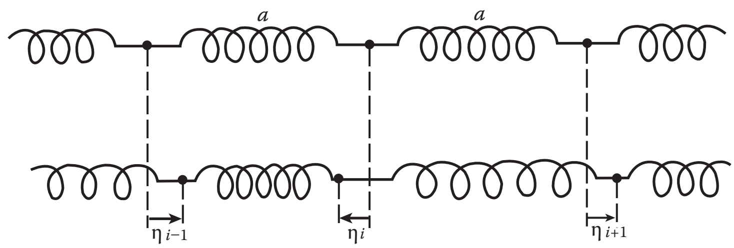


FIGURE 16.6 Infinitely long elastic rod undergoing small longitudinal vibrations.

2. Consider an infinitely long elastic rod undergoing small longitudinal vibrations, that is, oscillatory displacements of the particles of the rod parallel to the axis of the rod. A system composed of discrete particles that approximates the continuous rod is an infinite chain of equal mass points spaced a distance a apart and connected by uniform massless springs having force constants k (Figure 16.6). The mass points can move only along the length of the chain. Show that the Lagrangian for

$$L = \frac{1}{2}\sum_i a\left[\frac{m}{a}\dot{\eta}_i^2 - ka\left(\frac{\eta_{i+1}-\eta_i}{a}\right)^2\right]$$

is the system and the resulting Lagrange equations of motion for the coordinates η_i are

$$\frac{m}{a}\ddot{\eta}_i - ka\left(\frac{\eta_{i+1}-\eta_i}{a^2}\right) + ka\left(\frac{\eta_i-\eta_{i-1}}{a^2}\right) = 0.$$

3. From Equation 16.69, we can write the modified Hamilton's principle for a continuous system as

$$\delta\int\iiint\left\{\sum_k \pi_k\dot{q}_k - \hbar\right\}\mathrm{d}x_1\mathrm{d}x_2\mathrm{d}x_3\mathrm{d}t = 0.$$

Obtain Hamilton's equations of motion for a continuous system.

4. Show that if ψ and ψ^* are taken as two independent field variables, the Lagrangian density

$$L_d = \frac{h^2}{8\pi^2 m}\nabla\psi\cdot\nabla\psi^* + V\psi^*\psi + \frac{h}{2\pi i}(\psi^*\psi - \psi\psi^*)$$

leads to the Schrödinger equation

$$-\frac{\hbar^2}{8\pi^2 m}\nabla^2\psi + V\psi = \frac{ih}{2\pi}\frac{\partial\psi}{\partial t}$$

and its complex conjugate.

What are the canonical momenta? Obtain the Hamiltonian density corresponding to L_d.

Appendix 1: Vector Analysis and Ordinary Differential Equations

Many mathematical techniques will be introduced as the need arises, but vector analysis and ordinary differential equations (ODEs) are the basic techniques set forth as preliminaries to the main discussion.

A1.1 ELEMENTS OF VECTOR ALGEBRA

Vectors are quantities possessing magnitude and direction, and their combination with each other has a commutative property. A vector quantity will be denoted by boldface type (as $\boldsymbol{A}$) or with an arrow over the symbol (as $\vec{A}$), and its magnitude will be represented by the symbol itself (as A). The corresponding letter with a circumflex over it, such as $\hat{A} = \boldsymbol{A}/A$, denotes a unit vector in the direction of the vector $\boldsymbol{A}$. A vector can be specified by its components and the unit vectors along the coordinate axes. In a rectangular Cartesian coordinate system, an arbitrary vector $\boldsymbol{A}$ can be expressed as

$$\vec{A} = A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3 = \sum_{i=1}^{3} A_i\hat{e}_i \tag{A1.1}$$

where $\hat{e}_i (i = 1,2,3)$ are unit vectors along the rectangular axes x_i ($x_1 = x$, $x_2 = y$, $x_3 = z$). We also use the component triplet (A_x, A_y, A_z) or (A_1, A_2, A_3) as an alternative designation for vector $\boldsymbol{A}$:

$$\boldsymbol{A} = (\mathrm{A}_x, \mathrm{A}_y, \mathrm{A}_z). \tag{A1.2}$$

It is assumed that the reader is familiar with basic vector algebra. However, for the reader who is in need of review, a brief summary on vector algebra is given below.

A1.1.1 Equality of Vectors

Two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ are equal if, and only if, their respective components are equal:

$$\boldsymbol{A} = \boldsymbol{B} \text{ or } (A_1, A_2, A_3) = (B_1, B_2, B_3)$$

is equivalent to the three equations

$$A_1 = B_1, A_2 = B_2, A_3 = B_3.$$

Geometrically, equal vectors are parallel and have the same length but do not necessarily have the same position.

A1.1.2 Vector Addition

The addition of two vectors is defined by the equation

$$\boldsymbol{A} + \boldsymbol{B} = (A_1, A_2, A_3) + (B_1, B_2, B_3) = (A_1 + B_1, A_2 + B_2, A_3 + B_3).$$

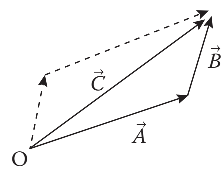


FIGURE A1.1 Vector addition.

That is, the sum of two vectors is a vector whose components are sums of the components of the two given vectors.

We can add two nonparallel vectors by the graphical method as shown in Figure A1.1. To add vector ***B*** to vector ***A***, shift ***B*** parallel to itself until its tail is at the head of ***A***. The vector sum ***A*** + ***B*** is a vector ***C*** drawn from the tail of ***A*** to the head of ***B***. The order in which the vectors are added does not affect the result.

A1.1.3 Multiplication by Scalar

If c is a scalar, then $c\boldsymbol{A} = (cA_1, cA_2, cA_3)$. Geometrically, the vector $c\boldsymbol{A}$ is parallel to ***A*** and is c times the length of ***A***. When $c = -1$, the vector $-\boldsymbol{A}$ is one whose direction is the reverse of that of ***A***, but both have the same length. That is, subtraction of vector ***B*** from vector ***A*** is equivalent to adding $-\boldsymbol{B}$ to ***A***:

$$\boldsymbol{A} - \boldsymbol{B} = \boldsymbol{A} + (-\boldsymbol{B}).$$

We see that vector addition has the following properties:

(a) $\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$ (commutativity)
(b) $(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$ (associativity)
(c) $\boldsymbol{A} + \boldsymbol{0} = \boldsymbol{0} + \boldsymbol{A} = \boldsymbol{A}$
(d) $\boldsymbol{A} + (-\boldsymbol{A}) = \boldsymbol{0}$

A1.1.4 Multiplication of Two Vectors

We now turn to multiplication of two vectors. Note that division by a vector is not defined; expressions such as $k/\boldsymbol{A}$ or $\boldsymbol{B}/\boldsymbol{A}$ are meaningless. There are three ways of multiplying two vectors:

A1.1.4.1 Scalar (Dot or Inner) Product

The scalar product of two vectors ***A*** and ***B*** is defined, in geometrical language, as the product of their magnitude and the cosine of the (smaller) angle between them as shown in Figure A1.2. Obviously it is a real number:

$$\boldsymbol{A} \cdot \boldsymbol{B} \equiv AB \cos\theta, \ (0 \le \theta \le \pi) \tag{A1.3}$$

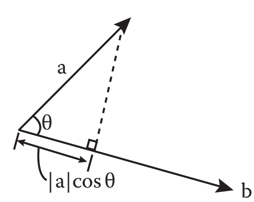


FIGURE A1.2 Scalar product.

It is clear from the definition (Equation A1.3) that the scalar product is commutative

$$\boldsymbol{A} \cdot \boldsymbol{B} = \boldsymbol{B} \cdot \boldsymbol{A} \tag{A1.4}$$

and the product of a vector with itself gives the square of the length of the vector:

$$\boldsymbol{A} \cdot \boldsymbol{A} = A^2 = A_x^2 + A_y^2 + A_z^2. \tag{A1.5}$$

If $\boldsymbol{A} \cdot \boldsymbol{B} = 0$ and neither $\boldsymbol{A}$ nor $\boldsymbol{B}$ is a null (zero) vector, then $\boldsymbol{A}$ is perpendicular to $\boldsymbol{B}$.

We can get a simple geometric interpretation of the scalar product from an inspection of Figure A1.4:

($B\cos\theta)A$ = projection of $\boldsymbol{B}$ onto $\boldsymbol{A}$ multiplied by the magnitude of $\boldsymbol{A}$
($A\cos\theta)B$ = projection of $\boldsymbol{A}$ onto $\boldsymbol{B}$ multiplied by the magnitude of $\boldsymbol{B}$.

If only the components of $\boldsymbol{A}$ and $\boldsymbol{B}$ are known, it would then not be practical to calculate $\boldsymbol{A} \cdot \boldsymbol{B}$ from definition A1.3. But, in this case, we can calculate $\boldsymbol{A} \cdot \boldsymbol{B}$ in terms of the components

$$\boldsymbol{A} \cdot \boldsymbol{B} = \left(A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3\right) \cdot \left(B_1\hat{e}_1 + B_2\hat{e}_2 + B_3\hat{e}_3\right). \tag{A1.6}$$

The right-hand side has nine terms, all involving the product $\hat{e}_i \cdot \hat{e}_j$. Fortunately, the angle between each pair of unit vectors is 90°, and from Equations A1.3 and A1.5, we find that

$$\hat{e}_i \cdot \hat{e}_j = \delta_{ij}, \quad i, j = 1,2,3 \tag{A1.7}$$

where δ_{ij} is the Kronecker delta symbol:

$$\delta_{ij} = \begin{cases} 0, & \text{if} \quad i \neq j \\ 1, & \text{if} \quad i = j \end{cases}. \tag{A1.8}$$

After using Equation A1.7 to simplify the resulting nine terms on the right-hand side of Equation A1.6, we obtain

$$\boldsymbol{A} \cdot \boldsymbol{B} = A_1B_1 + A_2B_2 + A_3B_3 = \sum_{i=1}^{3} A_iB_i. \tag{A1.9}$$

A1.1.4.2 Vector (Cross or Outer) Product

The vector product of two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ is a vector and is written as

$$\boldsymbol{C} = \boldsymbol{A} \times \boldsymbol{B}. \tag{A1.10}$$

As shown in Figure A1.3, the two vectors $\boldsymbol{A}$ and $\boldsymbol{B}$ form two sides of a parallelogram. We define $\boldsymbol{C}$ to be perpendicular to the plane of this parallelogram with a magnitude equal to the area of the parallelogram. And we choose the direction of $\boldsymbol{C}$ along the thumb of the right hand when the fingers rotate from $\boldsymbol{A}$ to $\boldsymbol{B}$ (angle of rotation less than 180°).

$$\vec{C} = \vec{A} \times \vec{B} = AB\sin\theta\ \hat{e}_C, \quad (0 \leq \theta \leq \pi). \tag{A1.11}$$

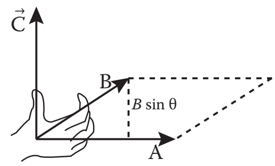


FIGURE A1.3 Vector product.

From the definition of the vector product and following the right hand rule, we can see immediately that

$$\boldsymbol{A} \times \boldsymbol{B} = -\boldsymbol{B} \times \boldsymbol{A}. \tag{A1.12}$$

Hence, the vector product is not commutative. If $\boldsymbol{A}$ and $\boldsymbol{B}$ are parallel, then

$$\boldsymbol{A} \times \boldsymbol{B} = 0 \text{ (if } \boldsymbol{A} \text{ and } \boldsymbol{B} \text{ are parallel).} \tag{A1.13a}$$

In particular,

$$\boldsymbol{A} \times \boldsymbol{A} = 0. \tag{A1.13b}$$

In vector components, we have

$$\mathrm{A} \times \mathrm{B} = \left(A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3\right) \times \left(B_1\hat{e}_1 + B_2\hat{e}_2 + B_3\hat{e}_3\right). \tag{A1.14}$$

Using the following relations:

$$\begin{aligned} &\hat{e}_i \times \hat{e}_i = 0, && i = 1,\ 2,\ 3 \\ &\hat{e}_1 \times \hat{e}_2 = \hat{e}_3,\quad \hat{e}_2 \times \hat{e}_3 = \hat{e}_1, && \hat{e}_3 \times \hat{e}_1 = \hat{e}_2, \end{aligned} \tag{A1.15a}$$

Equation A1.14 becomes

$$\boldsymbol{A} \times \boldsymbol{B} = (A_2B_3 - A_3B_2)\hat{e}_1 + (A_3B_1 - A_1B_3)\hat{e}_2 + (A_1B_2 - A_2B_1)\hat{e}_3. \tag{A1.15b}$$

This can be written as an easily remembered determinant of third order:

$$\boldsymbol{A} \times \boldsymbol{B} = \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \end{vmatrix}. \tag{A1.16}$$

The noncommutative nature of the cross product of two vectors now appears as a consequence of the fact that interchanging two rows of a determinant changes its sign. Similarly, the vanishing of the cross product of two vectors in the same direction appears as a consequence of the fact that a determinant vanishes if one of its rows is a multiple of another.

The vector resulting from the vector product of two vectors is called a pseudo vector, and the ordinary vectors are sometimes called axial vectors. Thus, in Equation A1.10, $\boldsymbol{C}$ is a pseudo vector and $\boldsymbol{A}$ and $\boldsymbol{B}$ are axial vectors. On an inversion of coordinates, axial vectors change sign and a pseudo vector does not change sign.

We can rewrite the nine equations implied by Equation A1.15 in terms of permutation symbols ε_{ijk}:

$$\hat{e}_i \times \hat{e}_j = \varepsilon_{ijk}\hat{e}_k \tag{A1.17}$$

where ε_{ijk} is defined by

$$\varepsilon_{ijk} = \begin{cases} +1 & \text{if } (i,\ j,\ k) \text{ is an even permutation of } (1, 2, 3) \\ -1 & \text{if } (i,\ j,\ k) \text{ is an odd permutation of } (1, 2, 3) \\ 0 & \text{otherwise (e.g., if two or more indices are equal).} \end{cases} \tag{A1.18}$$

It follows immediately that

$$\varepsilon_{ijk} = \varepsilon_{kij} = \varepsilon_{jki} = -\varepsilon_{jik} = -\varepsilon_{kji} = -\varepsilon_{ikj}.$$

There is a very useful identity relating the ε_{ijk} and the Kronecker delta symbol:

$$\sum_{k=1}^{3} \varepsilon_{mnk}\varepsilon_{ijk} = \delta_{mi}\delta_{nj} - \delta_{mj}\delta_{ni} \tag{A1.19a}$$

$$\sum_{j,k} \varepsilon_{mjk}\varepsilon_{njk} = 2\delta_{mn} \qquad \sum_{i,j,k} \varepsilon_{ijk}^2 = 6. \tag{A1.19b}$$

Using permutation symbols, we can now write the vector product $\boldsymbol{A} \times \boldsymbol{B}$ as

$$\boldsymbol{A} \times \boldsymbol{B} = \left(\sum_{i=1}^{3} A_i\hat{e}_i\right) \times \left(\sum_{j=1}^{3} B_j\hat{e}_j\right) = \sum_{i,j}^{3} A_iB_j\left(\hat{e}_i \times \hat{e}_j\right) = \sum_{i,j,k}^{3}\left(A_iB_j\varepsilon_{ijk}\right)\hat{e}_k.$$

Thus, the kth component of $\boldsymbol{A} \times \boldsymbol{B}$ is

$$(\boldsymbol{A} \times \boldsymbol{B})_k = \sum_{i,j} A_iB_j\varepsilon_{ijk} = \sum_{i,j} \varepsilon_{kij}A_iB_j.$$

If $k = 1$, we obtain the usual geometrical result:

$$(\boldsymbol{A} \times \boldsymbol{B})_1 = \sum_{i,j} \varepsilon_{1ij}A_iB_j = \varepsilon_{123}A_2B_3 + \varepsilon_{132}A_3B_2 = A_2B_3 - A_3B_2.$$

A1.1.4.3 Tensor Product

The product of $\boldsymbol{A}$ and $\boldsymbol{B}$ together without a dot or cross between, that is, $\boldsymbol{AB}$, is called a dyadic. We see that, in general, this product has nine terms and that each term consists of a scalar coefficient multiplied by a pair of unit vectors. As we will not use this product in this book, we shall not discuss it further.

A1.1.4.4 Triple Scalar Product $\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$

We now briefly discuss the scalar $\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C})$. This scalar represents the volume of the parallelepiped formed by the coterminous sides $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$ because

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = ABC \sin\theta \cos\alpha = hS = \text{volume}$$

where S is the area of the parallelogram with sides $\boldsymbol{B}$ and $\boldsymbol{C}$, and h is the height of the parallelogram (Figure A1.4).

Now,

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \left(A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3\right) \cdot \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}$$

$$= A_1(B_2C_3 - B_3C_2) + A_2(B_3C_1 - B_1C_3) + A_3(B_1C_2 - B_2C_1)$$

so that

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix}. \tag{A1.20}$$

The exchange of two rows (or two columns) changes the sign of the determinant but does not change its absolute value (this will be proved in the chapter on determinants). Using this property, we find

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \begin{vmatrix} A_1 & A_2 & A_3 \\ B_1 & B_2 & B_3 \\ C_1 & C_2 & C_3 \end{vmatrix} = -\begin{vmatrix} C_1 & C_2 & C_3 \\ B_1 & B_2 & B_3 \\ A_1 & A_2 & A_3 \end{vmatrix} = \boldsymbol{C} \cdot (\boldsymbol{A} \times \boldsymbol{B}).$$

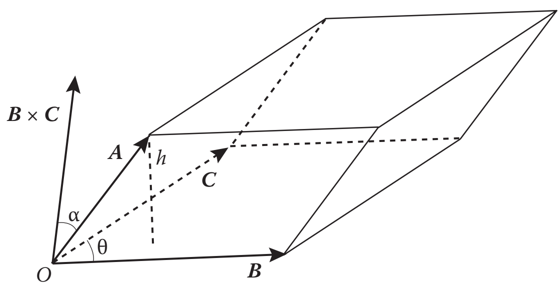


FIGURE A1.4 Triple scalar product.

That is, the dot and the cross may be interchanged in the triple scalar product:

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = (\boldsymbol{A} \times \boldsymbol{B}) \cdot \boldsymbol{C}. \tag{A1.21}$$

In fact, as long as the three vectors appear in cyclic order, $\boldsymbol{A} \to \boldsymbol{B} \to \mathbf{C} \to \boldsymbol{A}$, then the dot and cross may be inserted between any pairs:

$$\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B} \cdot (\boldsymbol{C} \times \boldsymbol{A}) = \boldsymbol{C} \cdot (\boldsymbol{A} \times \boldsymbol{B}). \tag{A1.22}$$

It should be noted that the scalar resulted from the triple scalar product changes sign on an inversion of coordinates. For this reason, the triple scalar product is sometimes called a pseudo scalar.

A1.1.4.5 Triple Vector Product

The triple product $\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C})$ is a vector because it is the vector product of two vectors: $\boldsymbol{A}$ and $\boldsymbol{B} \times \boldsymbol{C}$. This vector is perpendicular to $\boldsymbol{B} \times \boldsymbol{C}$, and so it lies in the plane of $\boldsymbol{B}$ and $\boldsymbol{C}$. If $\boldsymbol{B}$ is not parallel to $\boldsymbol{C}$, $\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = x\boldsymbol{B} + y\boldsymbol{C}$. Now dot both sides with $\boldsymbol{A}$, and we obtain $x(\boldsymbol{A} \cdot \boldsymbol{B}) + y(\boldsymbol{A} \cdot \boldsymbol{C}) = 0$ because $\boldsymbol{A}\cdot[\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C})] = 0$. Thus,

$$x/(\boldsymbol{A} \cdot \boldsymbol{C}) = -y/(\boldsymbol{A} \cdot \boldsymbol{B}) \equiv \lambda \quad (\lambda \text{ is a scalar})$$

and so

$$\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = x\boldsymbol{B} + y\boldsymbol{C} = \lambda[\boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B})].$$

We now show that $\lambda = 1$. To this purpose, let us consider the special case when $\boldsymbol{B} = \boldsymbol{A}$ and dot the last equation with $\boldsymbol{C}$:

$$\boldsymbol{C} \times [\boldsymbol{A} \times (\boldsymbol{A} \times \boldsymbol{C})] = \lambda[(\boldsymbol{A} \cdot \boldsymbol{C})^2 - \boldsymbol{A}^2\boldsymbol{C}^2]$$

or, by an interchange of dot and cross,

$$-(\boldsymbol{A} \cdot \boldsymbol{C})^2 = \lambda[(\boldsymbol{A} \cdot \boldsymbol{C})^2 - \boldsymbol{A}^2\boldsymbol{C}^2].$$

In terms of the angles between the vectors and their magnitudes, the last equation becomes

$$-A^2C^2\sin^2\theta = \lambda(A^2C^2 \cos^2\theta - A^2C^2) = -\lambda\, A^2C^2\sin^2\theta.$$

Hence, $\lambda = 1$, and so

$$\boldsymbol{A} \times (\boldsymbol{B} \times \boldsymbol{C}) = \boldsymbol{B}(\boldsymbol{A} \cdot \boldsymbol{C}) - \boldsymbol{C}(\boldsymbol{A} \cdot \boldsymbol{B}). \tag{A1.23}$$

A1.2 CALCULUS OF VECTORS

A1.2.1 Differentiation and Integration of Vector Functions

Consider a vector function $\boldsymbol{A}$; its components are functions of a scalar variable u. Then,

$$\vec{A}(u) = A_1(u)\hat{e}_1 + A_2(u)\hat{e}_2 + A_3(u)\hat{e}_3. \tag{A1.24}$$

We shall see later that the vector may be position, displacement, velocity, and so on, and the parameter u is usually the time t.

The derivative of the vector function $\boldsymbol{A}(u)$ with respect to u is defined by the limit

$$\frac{\mathrm{d}\vec{A}}{\mathrm{d}u} = \lim_{\Delta u \to 0} \frac{\Delta \vec{A}}{\Delta u} = \lim_{\Delta u \to 0} \left(\hat{e}_1 \frac{\Delta A_1}{\Delta u} + \hat{e}_2 \frac{\Delta A_2}{\Delta u} + \hat{e}_3 \frac{\Delta A_3}{\Delta u} \right)$$

where $\Delta A_1 = A_1(u + \Delta u) - A_1(u)$, and so on. Hence,

$$\frac{\mathrm{d}\vec{A}}{\mathrm{d}u} = \hat{e}_1 \frac{\mathrm{d}A_1}{\mathrm{d}u} + \hat{e}_2 \frac{\mathrm{d}A_2}{\mathrm{d}u} + \hat{e}_3 \frac{\mathrm{d}A_3}{\mathrm{d}u}. \tag{A1.25}$$

That is, the derivative of a vector is a vector whose Cartesian components are ordinary derivatives. The usual rules of differentiation familiar in calculus can be extended to vectors with one caution that the order of factors in products may be important. If $\varphi(u)$ is a scalar function, and $\boldsymbol{A}(u)$, $\boldsymbol{B}(u)$ are vector functions, then it is easy to show that

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(\phi\vec{A}\right) = \phi\frac{\mathrm{d}\vec{A}}{\mathrm{d}u} + \frac{\mathrm{d}\phi}{\mathrm{d}u}\vec{A}$$

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(\vec{A}\cdot\vec{B}\right) = \vec{A}\cdot\frac{\mathrm{d}\vec{B}}{\mathrm{d}u} + \frac{\mathrm{d}\vec{A}}{\mathrm{d}u}\cdot\vec{B}$$

$$\frac{\mathrm{d}}{\mathrm{d}u}\left(\vec{A}\times\vec{B}\right) = \vec{A}\times\frac{\mathrm{d}\vec{B}}{\mathrm{d}u} + \frac{\mathrm{d}\vec{A}}{\mathrm{d}u}\times\vec{B}.$$

We can define higher-order derivatives in similar way. The second derivative of $\boldsymbol{A}(u)$, for example, is given by

$$\frac{\mathrm{d}^2\vec{A}}{\mathrm{d}u^2} = \frac{\mathrm{d}^2A_1}{\mathrm{d}u^2}\hat{e}_1 + \frac{\mathrm{d}^2A_2}{\mathrm{d}u^2}\hat{e}_2 + \frac{\mathrm{d}^2A_3}{\mathrm{d}u^2}\hat{e}_3.$$

Also, quite analogously to the ordinary integration of a scalar function, we define the indefinite integral of $\boldsymbol{A}(u)$ as

$$\int \vec{A}(u)\,\mathrm{d}u = \hat{e}_1 \int A_1(u)\,\mathrm{d}u + \hat{e}_2 \int A_2(u)\,\mathrm{d}u + \hat{e}_3 \int A_3(u)\,\mathrm{d}u. \tag{A1.26}$$

If there exists a vector function $\boldsymbol{B}(u)$ such that $\vec{A}(u) = \frac{\mathrm{d}}{\mathrm{d}u}\vec{B}(u)$, then

$$\int \vec{A}(u)\,\mathrm{d}u = \int \frac{\mathrm{d}}{\mathrm{d}u}\left\{\vec{B}(u)\right\}\mathrm{d}u = \vec{B}(u) + \vec{c} \tag{A1.27}$$

where $\vec{c}$ is an arbitrary constant vector independent of u. For a definite integral, we have

$$\int_{\alpha}^{\beta} \vec{A}(u)\,\mathrm{d}u = \int_{\alpha}^{\beta} \frac{\mathrm{d}}{\mathrm{d}u}\left\{\vec{B}(u)\right\}\mathrm{d}u = \vec{B}(\beta) - \vec{B}(\alpha). \tag{A1.28}$$

We now explain what the gradient, the divergence, and the curl are.

A1.2.2 The Gradient

Given a scalar field described by a scalar function $\varphi(x_1, x_2, x_3)$ that is defined and differentiable at each point with respect to the position coordinates (x_1, x_2, x_3), then the total differential corresponding to an infinitesimal change $\mathrm{d}\boldsymbol{r} = (\mathrm{d}x_1, \mathrm{d}x_2, \mathrm{d}x_3)$ is

$$\mathrm{d}\phi = \frac{\partial \phi}{\partial x_1}\mathrm{d}x_1 + \frac{\partial \phi}{\partial x_2}\mathrm{d}x_2 + \frac{\partial \phi}{\partial x_3}\mathrm{d}x_3. \tag{A1.29}$$

We can express $\mathrm{d}\varphi$ as a scalar product of two vectors:

$$\mathrm{d}\phi = \frac{\partial \phi}{\partial x_1}\mathrm{d}x_1 + \frac{\partial \phi}{\partial x_2}\mathrm{d}x_2 + \frac{\partial \phi}{\partial x_3}\mathrm{d}x_3 = (\nabla\phi)\cdot \mathrm{d}\vec{r} \tag{A1.30}$$

where

$$\nabla\phi \equiv \frac{\partial \phi}{\partial x_1}\hat{e}_1 + \frac{\partial \phi}{\partial x_2}\hat{e}_2 + \frac{\partial \phi}{\partial x_3}\hat{e}_3 \tag{A1.31}$$

is a vector field (or a vector function of the coordinates). $\nabla\varphi$ is called the gradient of the scalar function φ and is often written as grad φ.

There is a simple geometric interpretation of $\nabla\varphi$. Note that $\varphi(x_1, x_2, x_3) = c$, where c is a constant, represents a surface. Let $\boldsymbol{r} = x_1\hat{e}_1 + x_2\hat{e}_2 + x_3\hat{e}_3$ be the position vector to a point $P(x_1, x_2, x_3)$ on the surface. If we move along the surface to a nearby point $Q(\boldsymbol{r} + \mathrm{d}\boldsymbol{r})$, then $\mathrm{d}\boldsymbol{r} = \mathrm{d}x_1\hat{e}_1 + \mathrm{d}x_2\hat{e}_2 + \mathrm{d}x_3\hat{e}_3$ lies in the tangent plane to the surface at P. But as long as we move along the surface, φ has a constant value, and so $\mathrm{d}\varphi = 0$. Consequently, from Equation A1.31,

$$\mathrm{d}\boldsymbol{r} \cdot \nabla\varphi = 0. \tag{A1.32}$$

Equation A1.32 states that $\nabla\varphi$ is perpendicular to $\mathrm{d}\boldsymbol{r}$ and therefore to the surface (see Figure A1.5).

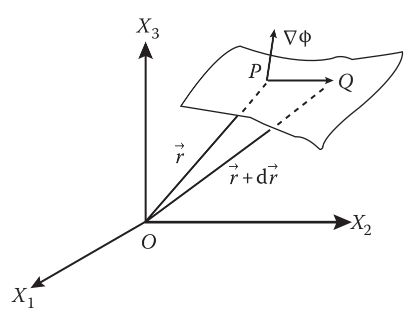


FIGURE A1.5 Geometric interpretation of $\nabla\phi$.

Now let us return to

$$d\phi = (\nabla\phi)\cdot d\vec{r}.$$

The vector $\nabla\varphi$ is fixed at any point P, so that $d\varphi$, the change in φ, will depend to a great extent on $d\boldsymbol{r}$. Consequently, $d\varphi$ will be a maximum when $d\boldsymbol{r}$ is parallel to $\nabla\varphi$. Next, because $d\boldsymbol{r}\cdot\nabla\varphi = |d\vec{r}||\nabla\phi|\cos\theta$, and $\cos\theta$ is a maximum for $\theta = 0$, thus, $\nabla\varphi$ is in the direction of maximum increase of $\varphi(x_1, x_2, x_3)$.

Example A1.1

Given that $\Phi = x^2 + y^2 + z^2$, find and describe $\nabla\Phi$.

Solution:

Obviously the surfaces of constant Φ, such as $\Phi = x^2 + y^2 + z^2 = 4$, are concentric spheres about the origin. Because $\partial\Phi/\partial x = 2x$, $\partial\Phi/\partial y = 2y$, $\partial\Phi/\partial z = 2z$, we find from Equation A1.31 that

$$\nabla\Phi = 2x\hat{i} + 2y\hat{j} + 2z\hat{k}.$$

As the components of this vector are proportional to x, y, and z, the vector is in the outward radial direction, perpendicular to the surfaces of constant Φ.

Gradients find many physical applications:

1. On a weather map, the isobars are lines of constant pressure (Φ), and $\nabla\Phi$ are the pressure gradient. Winds of high velocity are usually associated with a large pressure gradient.
2. In heat flow, isothermals are lines of constant temperature (Φ). Heat flows in the direction of the negative temperature gradient.
3. In electrostatics, equipotential surfaces (V = constant) and lines of force are mutually perpendicular, the electric field being equal to $-\nabla V$.

A1.2.3 Vector Differential Operator ∇

We denoted the operation that changes a scalar field to a vector field in Equation A1.31 by the symbol ∇ (called "del"):

$$\nabla \equiv \frac{\partial}{\partial x_1}\hat{e}_1 + \frac{\partial}{\partial x_2}\hat{e}_2 + \frac{\partial}{\partial x_3}\hat{e}_3. \tag{A1.33}$$

It is a vector differential operator, often called the gradient operator, but we prefer to call it the del operator. It will help us in the future to keep in mind that the del operator acts both as a differential operator and a vector.

We now apply the del operator vector differentiation of a vector field. There are two types of products involving two vectors, namely, the scalar and vector products. Vector differential operations on vector fields can also be separated into two types called the curl and the divergence.

A1.2.4 Divergence of a Vector

If $\boldsymbol{V}(x_1, x_2, x_3) = V_1\hat{e}_1 + V_2\hat{e}_2 + V_3\hat{e}_3$ is a differentiable vector field [i.e., it is defined and differentiable at each point (x_1, x_2, x_3) in a certain region of space], the divergence of $\boldsymbol{V}$, written $\nabla \cdot \boldsymbol{V}$ or div $\boldsymbol{V}$, is defined by the scalar product

$$\nabla \cdot \vec{V} = \left(\frac{\partial}{\partial x_1}\hat{e}_1 + \frac{\partial}{\partial x_2}\hat{e}_2 + \frac{\partial}{\partial x_3}\hat{e}_3 \right) \cdot \left(V_1\hat{e}_1 + V_2\hat{e}_2 + V_3\hat{e}_3 \right) = \frac{\partial V_1}{\partial x_1} + \frac{\partial V_2}{\partial x_2} + \frac{\partial V_3}{\partial x_3}. \quad \text{(A1.34)}$$

It is obvious that the result is a scalar field. Note the analogy with $\boldsymbol{A} \cdot \boldsymbol{B} = A_1B_1 + A_2B_2 + A_3B_3$, but also note that $\nabla \cdot \boldsymbol{V} \neq \boldsymbol{V} \cdot \nabla$ (bear in mind that ∇ is an operator). $\boldsymbol{V} \cdot \nabla$ is a scalar differential operator:

$$\vec{V} \cdot \nabla = V_1 \frac{\partial}{\partial x_1} + V_2 \frac{\partial}{\partial x_2} + V_3 \frac{\partial}{\partial x_3}.$$

The physical meaning of the divergence is best seen in connection with the flow of a fluid or of heat. If $\boldsymbol{A}$ ($=\rho \boldsymbol{v}$) represents the quantity of fluid flowing per unit time and per unit area through an area perpendicular to $\boldsymbol{A}$, A_x will be the quantity passing per unit time through the unit area at right angles to the x-axis. Consider a small rectangular element ABCDEFGH of volume $dx_1dx_2dx_3$ as in Figure A1.6. The volume per unit time flowing in across the face ABCD is given by $A_2dx_1dx_3$; the amount leaving the face EFGH per unit time is

$$\left[A_2 + \frac{\partial A_2}{\partial x_2} dx_2 \right] dx_1 dx_3.$$

So the loss of mass per unit time is

$$\frac{\partial A_2}{\partial x_2} dx_1 dx_2 dx_3.$$

Adding the net rate of flow out of all three pairs of surfaces of our small rectangular volume ABCDEFG, the total mass loss per unit time is

$$\left[\frac{\partial A_1}{\partial x_1} + \frac{\partial A_2}{\partial x_2} + \frac{\partial A_2}{\partial x_3} \right] dx_1 dx_2 dx_3 = (\nabla \cdot \vec{A}) dx_1 dx_2 dx_3$$

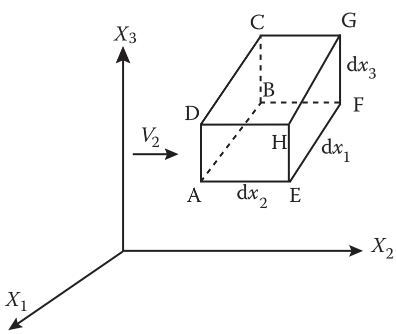


FIGURE A1.6 Divergence of a vector.

from which we see that $\nabla \cdot \vec{A}$ is the net rate of flow out per unit volume, hence the name divergence.

It is easy to show that

$$\nabla \cdot \left(f\vec{A}\right) = f\ \nabla \cdot \vec{A} + \vec{A} \cdot \nabla f \tag{A1.35}$$

where f is a scalar.

$$\nabla \cdot \left(f\vec{A}\right) = \frac{\partial}{\partial x_1}(fA_1) + \frac{\partial}{\partial x_2}(fA_2) + \frac{\partial}{\partial x_3}(fA_3)$$
$$= f\left(\frac{\partial A_1}{\partial x_1} + \frac{\partial AV_2}{\partial x_2} + \frac{\partial A_3}{\partial x_3}\right) + \left(A_1\frac{\partial f}{\partial x_1} + A_2\frac{\partial f}{\partial x_2} + A_3\frac{\partial f}{\partial x_3}\right).$$

A vector $\vec{A}$ is said to be solenoidal if its divergence is everywhere zero: $\nabla \cdot \vec{A} = 0$.

Example A1.2

Given the vector function $\vec{A} = 1\hat{i} + xyz\hat{j} + x^2y^2z^2\hat{k}$, find the general expression for $\nabla \cdot \vec{A}$ at any point and its value at the specific point (2, –1, 3).

Solution:

Taking the partial derivative, we get

$$\partial A_x/\partial x = 0,\ \partial A_y/\partial y = xz,\ \partial A_z/\partial z = 2x^2y^2z$$

and

$$\nabla \cdot \vec{A} = xz + 2x^2y^2z.$$

At the point (2, –1, 3)

$$\nabla \cdot \vec{A} = 6 + 24 = 30.$$

A1.2.5 Laplacian Operator ∇^2

The divergence of a vector field is defined by the scalar product of the operator ∇ with the vector field. What is the scalar product of ∇ with itself?

$$\nabla^2 = \nabla \cdot \nabla = \left(\frac{\partial}{\partial x_1}\hat{e}_1 + \frac{\partial}{\partial x_2}\hat{e}_2 + \frac{\partial}{\partial x_3}\hat{e}_3\right) \cdot \left(\frac{\partial}{\partial x_1}\hat{e}_1 + \frac{\partial}{\partial x_2}\hat{e}_2 + \frac{\partial}{\partial x_3}\hat{e}_3\right)$$
$$= \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2}$$

This important quantity

$$\nabla^2 = \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} \tag{A1.36}$$

is a scalar differential operator called the Laplacian, after a French mathematician of the 18th century named Laplace.

Because the Laplacian is a scalar differential operator, it does not change the vector character of the field on which it operates. Thus, $\nabla^2\varphi(\boldsymbol{r})$ is a scalar field if $\varphi(\boldsymbol{r})$ is a scalar field, and $\nabla^2[\nabla\varphi(\boldsymbol{r})]$ is a vector field because the gradient $\nabla\varphi(\boldsymbol{r})$ is a vector field.

The equation $\nabla^2\varphi = 0$ is called Laplace's equation.

A1.2.6 Curl of a Vector

If $\boldsymbol{V}(x_1, x_2, x_3)$ is a differentiable vector field, then the curl of $\boldsymbol{V}$, written $\nabla \times \boldsymbol{V}$ (or curl $\boldsymbol{V}$), is defined by the vector product

$$\text{Curl } \vec{V} = \nabla \times \vec{V} = \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ \dfrac{\partial}{\partial x_1} & \dfrac{\partial}{\partial x_2} & \dfrac{\partial}{\partial x_3} \\ V_1 & V_2 & V_3 \end{vmatrix} \tag{A1.37}$$

$$= \hat{e}_1\left(\frac{\partial V_3}{\partial x_2} - \frac{\partial V_2}{\partial x_3}\right) + \hat{e}_2\left(\frac{\partial V_1}{\partial x_3} - \frac{\partial V_3}{\partial x_1}\right) + \hat{e}_3\left(\frac{\partial V_2}{\partial x_1} - \frac{\partial V_1}{\partial x_2}\right).$$

The result is a vector field. In the expansion of the determinant, the operators $\dfrac{\partial}{\partial x_i}$ must precede V_i.

A vector $\boldsymbol{V}$ is said to be irrotational if its curl is zero: $\nabla \times \boldsymbol{V}(\boldsymbol{r}) = 0$. From this definition, we see that the gradient of any scalar field $\varphi(\boldsymbol{r})$ is irrotational.

A vector $\boldsymbol{V}$ is solenoidal (or divergence-free) if its divergence is zero. From this, one can show that the curl of any vector field $\boldsymbol{V}(\boldsymbol{r})$ is solenoidal:

$$\nabla \cdot \left(\nabla \times \vec{V}\right) = 0. \tag{A1.38}$$

We leave the proof out because it is tedious.

A vector field that has nonvanishing curl is called a vortex field, and the curl of the field vector is a measure of the vorticity of the vector field.

If $\varphi(\boldsymbol{r})$ is a scalar field and $\boldsymbol{V}(\boldsymbol{r})$ is a vector field, then one can show that

$$\nabla \times (\varphi \boldsymbol{V}) = \varphi(\nabla \times \boldsymbol{V}) + (\nabla\varphi) \times \boldsymbol{V}. \tag{A1.39}$$

To prove it, we first write

$$\nabla \times (\varphi \boldsymbol{V}) = \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ \dfrac{\partial}{\partial x_1} & \dfrac{\partial}{\partial x_2} & \dfrac{\partial}{\partial x_3} \\ \phi V_1 & \phi V_2 & \phi V_3 \end{vmatrix}$$

then notice that $\frac{\partial}{\partial x_1}(\phi V_2) = \phi \frac{\partial V_2}{\partial x_1} + \frac{\partial \phi}{\partial x_1} V_2$, so we can expand the determinant in the above equation as a sum of two determinants:

$$\nabla \times (\varphi \boldsymbol{V}) = \phi \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ \frac{\partial}{\partial x_1} & \frac{\partial}{\partial x_2} & \frac{\partial}{\partial x_3} \\ V_1 & V_2 & V_3 \end{vmatrix} + \begin{vmatrix} \hat{e}_1 & \hat{e}_2 & \hat{e}_3 \\ \frac{\partial \phi}{\partial x_1} & \frac{\partial \phi}{\partial x_2} & \frac{\partial \phi}{\partial x_3} \\ V_1 & V_2 & V_3 \end{vmatrix}$$

$$= \varphi(\nabla \times \boldsymbol{V}) + (\nabla \varphi) \times \boldsymbol{V}.$$

The physical meaning of the curl of a vector is not quite as transparent as that for the divergence. The following example from fluid flow may help us to develop a better feeling. Suppose that $\boldsymbol{V}$ represents the linear velocity v of a fluid at any point (x_1, x_2, x_3); the components of $\boldsymbol{v}$ are v_1, v_2, v_3. For the moment, let us consider v_1 only. If v_1 is a function of x_1 only, then the velocity in the x_1-axis direction will vary with x_1, but successive layers parallel to the x_1-axis will all move alike. If, however, v_1 is a function of x_2, then different layers in the x_1x_3-plane will slide over, or move relative to, one another (Figure A1.7). If $\partial v_1/\partial x_2$ is positive in a certain region, the fluid there is essentially rotating, or curling, about the x_3-axis in what we have taken to be the negative sense (rotation from x_2 to x_1). Similarly, if $\partial v_2/\partial x_1$ is positive, the fluid is rotating in a positive sense about the x_3-axis. The total rotation about the x_3-axis must then be proportional to

$$\partial v_2/\partial x_1 = -\partial v_1/\partial x_2$$

which is the x_3-component of $\nabla \times \boldsymbol{v}$.

In this manner, we see that $\nabla \times \boldsymbol{v}$ is proportional to the angular velocity of the fluid at the point considered. For instance, suppose that $v_1 = -x_2$, $v_2 = x_1$, and $v_3 = 0$; then $\nabla \times \boldsymbol{v} = \hat{e}_3(1+1) = 2\hat{e}_3$. This represents a uniform rate of rotation, or vorticity, throughout the fluid. If, on the other hand, $v_1 = x_1$, $v_2 = -x_2$, and $v_3 = 0$, there is no vorticity present in the fluid.

In advanced works on vector analysis, it is proved that any vector functions may be written as the sum of two parts, one of which is irrotational and the other solenoidal.

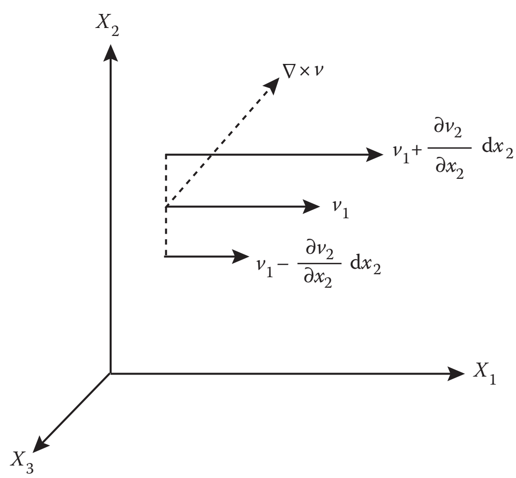


FIGURE A1.7 Curl of a vector.

We now turn to integration of vector functions. As we do not expect to use integral theorems, we limit our discussion to line integrals and surface integrals of vector functions.

A1.2.7 Vector Representation of Surfaces

A bounded plane surface has a magnitude represented by its area and a direction specified by the direction of its normal. If the surface is a part of a closed surface, the outward drawn normal is taken as positive. If the surface is not part of a closed surface, we specify the positive direction of the normal by the rule that a right-handed screw rotated in the plane of the surface in the positive sense of the periphery advances along the positive normal.

If the surface is not a plane, it may be divided into a number of elementary surfaces, each of which is a plane to any desired degree of approximation. The vector representative of the entire surface is the sum of the vectors representing its elements.

A1.2.8 Line Integrals

Let $\boldsymbol{A}(x_1, x_2, x_3)$ be a vector function of position and P_1P_2 a curve described in the sense from P_1 to P_2. Divide the curve into vector elements $\mathrm{d}\boldsymbol{\lambda}_1$, $\mathrm{d}\boldsymbol{\lambda}_2$, $\mathrm{d}\boldsymbol{\lambda}_3$, etc., and take the scalar product $\boldsymbol{A}_1\cdot \mathrm{d}\boldsymbol{\lambda}_1$ of $\boldsymbol{A}$ at the point P_1 and $\mathrm{d}\boldsymbol{\lambda}_1$, $\boldsymbol{A}_2\cdot \mathrm{d}\boldsymbol{\lambda}_2$ of $\boldsymbol{A}$ at the point C and $\mathrm{d}\boldsymbol{\lambda}_2$, and so on (Figure A1.8). The sum of these scalar products is given by

$$\sum_i \vec{A}_i \cdot \mathrm{d}\vec{\lambda}_i = \int_{P_1}^{P_2} \vec{A}\cdot \mathrm{d}\vec{\lambda}$$

and is known as the line integral of $\boldsymbol{A}$ along the curve P_1P_2.

As

$$\mathrm{d}\vec{\lambda} = \mathrm{d}x_1\hat{e}_1 + \mathrm{d}x_2\hat{e}_2 + \mathrm{d}x_3\hat{e}_3$$

$$\begin{aligned}\int_{P_1}^{P_2} \vec{A}\cdot \mathrm{d}\vec{\lambda} &= \int_{P_1}^{P_2} \left(A_1\hat{e}_1 + A_2\hat{e}_2 + A_3\hat{e}_3\right)\cdot\left(\mathrm{d}x_1\hat{e}_1 + \mathrm{d}x_2\hat{e}_2 + \mathrm{d}x_3\hat{e}_3\right)\\ &= \int_{P_1}^{P_2} A_1(x_1,x_2,x_3)\,\mathrm{d}x_1 + \int_{P_1}^{P_2} A_2(x_1,x_2,x_3)\,\mathrm{d}x_2 + \int_{P_1}^{P_2} A_3(x_1,x_2,x_3)\,\mathrm{d}x_3.\end{aligned}$$

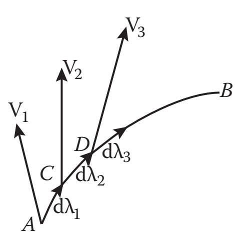


FIGURE A1.8 Line integrals.

Each integral on the right-hand side requires for its execution more than knowledge of the limits. In fact, the three integrals on the right-hand side are not completely defined because in the first integral, for example, we do not the know value of x_2 and x_3 in A_1:

$$I_1 = \int_{P_1}^{P_2} A_1(x_1, x_2, x_3)\,\mathrm{d}x_1.$$

What is needed is a statement, such as $x_2 = f(x_1)$, $x_3 = g(x_1)$, that specifies x_2, x_3 for each value of x_1 (so specify the path of integration). The integrand now reduces to $A_1(x_1, x_2, x_3) = A_1(x_1, f(x_1), g(x_1)) = B_1(x_1)$ so that the integral I_1 becomes well defined. Therefore, in general, a line integral is a path-dependent integral. However, if the scalar product $\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{\lambda}$ is equal to an exact differential, $\boldsymbol{A} \cdot \mathrm{d}\boldsymbol{\lambda} = \mathrm{d}\phi = \nabla\phi \cdot d\boldsymbol{\lambda}$, the integration depends only upon the limits and is therefore path-independent:

$$\int_{P_1}^{P_2} \vec{A} \cdot \mathrm{d}\vec{\lambda} = \int_{P_1}^{P_2} \mathrm{d}\varphi = \varphi_2 - \varphi_1.$$

We see that the line integral of the gradient of any scalar function of position ϕ around a closed curve is zero. When the path of integration is closed, the line integral is written as

$$\oint \vec{A} \cdot d\vec{\lambda} \text{ or } \oint_{\Gamma} \vec{A} \cdot d\vec{\lambda}$$

where Γ specifies the closed path.

A vector field that has the above property (path-independent) is called conservative. The curl of a conservative vector field is zero ($\nabla \times \boldsymbol{A} = \nabla \times (\nabla\phi) = 0$). A typical example of a conservative vector field in mechanics is a conservative force.

A1.2.9 Surface Integrals

Consider a surface S. Divide it into vector elements $\mathrm{d}\boldsymbol{a}_1$, $\mathrm{d}\boldsymbol{a}_2$, $\mathrm{d}\boldsymbol{a}_3$, etc. Let $\boldsymbol{A}_1$ be the value of the vector function of position $\boldsymbol{A}$ at $\mathrm{d}\boldsymbol{a}_1$, $\boldsymbol{A}_2$ its value at $\mathrm{d}\boldsymbol{a}_2$, and so on. The sum of the scalar products over the whole surface is known as the surface integral of $\boldsymbol{A}$ over the surface S:

$$\int_S \vec{A} \cdot \mathrm{d}\vec{a}$$

where $\mathrm{d}\vec{a} = \hat{n}\mathrm{d}a$, and the surface integral symbol $\int_S$ stands for a double integral over a certain surface S (Figure A1.9).

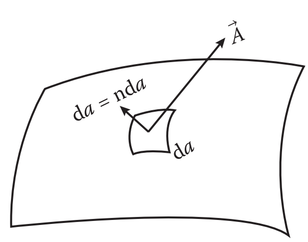


FIGURE A1.9 Surface integrals.

In rectangular coordinates, we may write

$$d\vec{a} = \hat{e}_1 da_1 + \hat{e}_2 da_2 + \hat{e}_3 da_3 = \hat{e}_1 dx_2 dx_3 + \hat{e}_2 dx_3 dx_1 + \hat{e}_3 dx_1 dx_2.$$

If a surface integral is to be evaluated over a closed surface S, the integral is written as

$$\oint_S \vec{A} \cdot d\vec{a}.$$

A1.2.10 Formulas Involving Operator ∇

We list some important formulas involving the vector differential operator ∇. In these formulas, ***A*** and ***B*** are differentiable vector field functions, and f and g are differentiable scalar field functions of position (x_1, x_2, x_3):

(1) $\nabla(fg) = f\,\nabla g + g\,\nabla f$
(2) $\nabla \times (f\vec{A}) = f\ \nabla \times \vec{A} + \nabla f \times \vec{A} \nabla \cdot (f\vec{A}) = f\nabla \cdot \vec{A} + \nabla f \cdot \vec{A}$
(3) $\nabla \times (f\vec{A}) = f\nabla \times \vec{A} + \nabla f \cdot \vec{A}$
(4) $\nabla \times (\nabla f) = 0$
(5) $\nabla \cdot (\nabla \times \vec{A}) = 0$
(6) $\nabla \cdot (\vec{A} \times \vec{B}) = (\nabla \times \vec{A}) \cdot \vec{B} - (\nabla \times \vec{B}) \times \vec{A}$
(7) $\nabla \times (\vec{A} \times \vec{B}) = (\vec{B} \cdot \nabla)\vec{A} - \vec{B}(\nabla \cdot \vec{A}) + \vec{A}(\nabla \cdot \vec{B}) - (\vec{A} \cdot \nabla)\vec{B}$
(8) $\nabla \times (\nabla \times \vec{A}) = \nabla(\nabla \cdot \vec{A}) - \nabla^2 \vec{A}$
(9) $\nabla(\vec{A} \cdot \vec{B}) = \vec{A} \times (\nabla \times \vec{B}) + \vec{B} \times (\nabla \times \vec{A}) + (\vec{A} \cdot \nabla)\vec{B} + (\vec{B} \cdot \nabla)\vec{A}$
(10) $(\vec{A} \cdot \nabla)\vec{r} = \vec{A}$
(11) $\nabla \cdot \vec{r} = 3$
(12) $\nabla \times \vec{r} = 0$
(13) $\nabla \cdot (r^{-3}\vec{r}) = 0$
(14) $d\vec{F} = (d\vec{r} \cdot \nabla)\vec{F} + \frac{\partial \vec{F}}{\partial t} dt$ ($\vec{F}$ a differentiable vector field quantity)
(15) $d\varphi = d\vec{r} \cdot \nabla\varphi + \frac{\partial \varphi}{\partial t} dt$ (φ a differentiable scalar field quantity)

A1.3 ORDINARY DIFFERENTIAL EQUATIONS

A differential equation is an equation that contains derivatives of unknown function, which express the relationship we seek. If there is only one independent variable and, as a consequence, total derivatives such as dx/dt, the equation is called an ordinary differential equation (ODE). A partial differential equation (PDE) contains several independent variables and hence partial derivatives.

The order of a differential equation is the order of the highest derivative appearing in the equation; its degree is the power of the derivative of the highest order after the equation has been rationalized, that is, after fractional powers of all derivatives have been removed. Thus, the equation

$$\frac{d^2 y}{dx^2} + 3\frac{dy}{dx} + 2y = 0$$

is of second order and first degree, and

$$\frac{d^3y}{dx^3} = \sqrt{1+(dy/dx)^3}$$

is of third order and second degree because it contains the term $(d^3y/dx^3)^2$ after it is rationalized.

A differential equation is said to be linear if each term in it is such that the dependent variable or its derivatives occur only once and only to the first power. Thus,

$$\frac{d^3y}{dx^3} + y\frac{dy}{dx} = 0$$

is not linear, but

$$x^3\frac{d^3y}{dx^3} + e^x \sin x\frac{dy}{dx} + y = \ln x$$

is linear. If in a linear differential equation there are no terms independent of y, the dependent variable, the equation is also said to be homogeneous; this would have been true for the last equation above if the "ln x" term on the right-hand side had been replaced by zero.

A very important property of linear homogeneous equations is that, if we know two solutions y_1 and y_2, we can construct others as linear combinations of them. This is known as the principle of superposition and will be proved later when we deal with such equations.

A1.3.1 First-Order Differential Equation

A first-order differential equation can always be solved, although the solution may not always be expressible in terms of familiar functions. A differential equation of the general form

$$\frac{dy}{dx} = -\frac{f(x,y)}{g(x,y)} \quad \text{or} \quad g(x,y)dy + f(x,y)dx = 0 \tag{A1.40}$$

is clearly a first-order differential equation.

A1.3.1.1 Solution by Separation of Variables

If $f(x,y)$ and $g(x,y)$ are reducible to $P(x)$ and $Q(y)$, respectively, then we have

$$Q(y)dy + P(x)dx = 0. \tag{A1.41}$$

Its solution is found at once by integrating.

The reader may notice that dy/dx has been treated as if it were a ratio of dy and dx that can be manipulated independently. Mathematicians may be unhappy about this treatment. But, if necessary, we could justify it by considering dy and dx to represent small finite changes δy and δx before we have actually gone to the limit where each becomes infinitesimal.

Example A1.3

Consider the differential equation $dy/dx = -y^2e^x$.

We can rewrite it in the following form: $-dy/y^2 = e^xdx$, which can be integrated separately giving the solution

$$1/y = e^x + c$$

where c is an integrating constant.

Sometimes when the variables are not separable, a change of variable may reduce the differential equation to one in which they are separable. The general form of differential equation amenable to this approach is

$$dy/dx = f(ax + by) \quad \text{(A1.42)}$$

where f is an arbitrary function and a and b are constants. If we let $w = ax + by$, then $bdy/dx = dw/dx - a$, and the differential equation becomes

$$dw/dx - a = bf(w)$$

from which we obtain

$$\frac{dw}{a + bf(w)} = dx$$

in which the variables are separated.

Example A1.4: Solve the Equation $dy/dx = 8x + 4y + (2x + y - 1)^2$

Let $w = 2x + y$; then $dy/dx = dw/dx - 2$, and the differential equation becomes

$$dw/dx + 2 = 4w + (w - 1)^2$$

or

$$dw/[4w + (w - 1)^2 - 2] = dx.$$

The variables are separated, and the equation can be solved.

A homogeneous differential equation that has the general form

$$dy/dx = f(y/x) \quad \text{(A1.43)}$$

may be reduced, by a change of variable, to one with variables that are separable. This can be illustrated by the following example.

Example A1.5: Solve the Equation

$$\frac{dy}{dx} = \frac{y^2 + xy}{x^2}.$$

The right-hand side can be rewritten as $(y/x)^2 + (y/x)$ and, hence, is a function of the single variable

$$v = y/x.$$

We thus use v both for simplifying the right-hand side of our equation and also for rewriting $\mathrm{d}y/\mathrm{d}x$ in terms of v and x. Now,

$$\frac{\mathrm{d}y}{\mathrm{d}x} = \frac{\mathrm{d}}{\mathrm{d}x}(xv) = v + x\frac{\mathrm{d}v}{\mathrm{d}x}$$

and our equation becomes

$$v + x\frac{\mathrm{d}v}{\mathrm{d}x} = v^2 + v$$

from which we have

$$\frac{\mathrm{d}v}{v^2} = \frac{\mathrm{d}x}{x}.$$

Integration gives

$$-\frac{1}{v} = \ln x + c \quad \text{or} \quad x = Ae^{-x/y}$$

where c and A $(= e^{-c})$ are constants.

Sometimes a nearly homogeneous differential equation can be reduced to homogeneous form, which then can be solved by variable separation. This can be illustrated by the following.

A1.3.1.2 Exact Equations

We may integrate Equation A1.40 directly if its left side is the differential du of some function $u(x,y)$, in which case, the solution is of the form

$$u(x,y) = C \tag{A1.44}$$

and Equation A1.40 is said to be exact. A convenient test to see if Equation A1.40 is exact is

$$\frac{\partial g(x,y)}{\partial x} = \frac{\partial f(x,y)}{\partial y}. \tag{A1.45}$$

To see this, let us go back to Equation A1.44, and we have

$$\mathrm{d}[u(x,y)] = 0.$$

On performing the differentiation, we obtain

$$\frac{\partial u}{\partial x}\mathrm{d}x + \frac{\partial u}{\partial y}\mathrm{d}y = 0. \tag{A1.46}$$

It is a general property of partial derivatives of any reasonable function that the order of differentiation is immaterial. Thus, we have

$$\frac{\partial}{\partial y}\left(\frac{\partial u}{\partial x}\right)=\frac{\partial}{\partial x}\left(\frac{\partial u}{\partial y}\right). \tag{A1.47}$$

Now, if our differential equation A1.40 is of the form of Equation A1.46, we must be able to identify

$$f(x, y) = \partial u/\partial x \text{ and } g(x, y) = \partial u/\partial y. \tag{A1.48}$$

Then it follows from Equation A1.47 that

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

which is Equation A1.45.

Example A1.6: Show That the Equation $x\mathrm{d}y/\mathrm{d}x + (x + y) = 0$ Is Exact and Find its General Solution

We first write the equation in standard form

$$(x + y)\mathrm{d}x + x\mathrm{d}y = 0.$$

Applying the test of Equation A1.45, we notice that

$$\frac{\partial f}{\partial y}=\frac{\partial}{\partial y}(x+y)=1 \quad \text{and} \quad \frac{\partial g}{\partial x}=\frac{\partial x}{\partial x}=1.$$

Therefore, the equation is exact, and the solution is of the form indicated by Equation A1.46. From Equation A1.48, we have

$$\partial u/\partial x = x + y,\ \partial u/\partial y = x$$

from which it follows that

$$u(x, y) = x^2/2 + xy + h(y),\ u(x, y) = xy + k(x)$$

where $h(y)$ and $k(x)$ arise from integrating $u(x,y)$ with respect to x and y, respectively. For consistency, we require that

$$h(y) = 0 \text{ and } k(x) = x^2/2.$$

Thus, the required solution is

$$x^2/2 + xy = c.$$

A1.3.1.3 Integrating Factors

If the differential equation in the form of Equation A1.40 is not already exact, sometimes it can be made so by multiplying by a suitable factor, called an integrating factor. Although an integrating factor always exists for each equation in the form of Equation A1.40, it may be troublesome to find it. However, if the equation is a linear equation of the first order, that is, if it is of the form

$$\frac{\mathrm{d}y}{\mathrm{d}x} + f(x)y = g(x) \tag{A1.49}$$

an integrating factor of the form

$$\exp\left(\int f(x)\,\mathrm{d}x\right) \tag{A1.50}$$

is always available. It is easy to verify this. Suppose that $R(x)$ is the integrating factor we are looking for. Multiplying Equation A1.49 by R, we have

$$R\frac{\mathrm{d}y}{\mathrm{d}x} + Rf(x)y = Rg(x), \quad \text{or} \quad R\mathrm{d}y + Rf(x)y\mathrm{d}x = Rg(x)\mathrm{d}x.$$

The right-hand side is integrable already; the condition that the left-hand side of Equation A1.49 be exact gives

$$\frac{\partial}{\partial y}[Rf(x)y] = \frac{\partial R}{\partial x}$$

which yields

$$\mathrm{d}R/\mathrm{d}x = Rf(x) \text{ or } \mathrm{d}R/R = f(x)\mathrm{d}x$$

integrating

$$\log\ R = \int f(x)\,\mathrm{d}x$$

from which we obtain the integrating factor R we were looking for:

$$R = \exp\left(\int f(x)\,\mathrm{d}x\right).$$

It is now possible to write the general solution to Equation A1.49. On application, the integrating factor, Equation A1.49, becomes

$$\frac{\mathrm{d}(ye^F)}{\mathrm{d}x} = g(x)e^F$$

where $F(x) = \int f(x)\,\mathrm{d}x$. The solution is clearly given by

$$y = e^{-F}\left[\int e^{F} g(x)\,\mathrm{d}x + C\right].$$

Example A1.7

Show that the equation $x\mathrm{d}y/\mathrm{d}x + 2y + x^2 = 0$ is not exact; then find a suitable integrating factor that makes the equation exact. What is the solution of this equation?

We first write the equation in the standard form:

$$(2y + x^2)\mathrm{d}x + x\mathrm{d}y = 0.$$

Then we notice that

$$\frac{\partial}{\partial y}(2y + x^2) = 2 \quad \text{and} \quad \frac{\partial}{\partial x}x = 1$$

which indicates that our equation is not exact. To find the required integrating factor that makes our equation exact, we rewrite our equation in the form of Equation A1.49:

$$\frac{\mathrm{d}y}{\mathrm{d}x} + \frac{2y}{x} = -x$$

from which we find $f(x) = 1/x$, and so the required integrating factor is

$$\exp\left(\int (1/x)\,\mathrm{d}x\right) = \exp(\ln\, x) = x.$$

Applying this to our equation

$$x^2\frac{\mathrm{d}y}{\mathrm{d}x} + 2xy + x^3 = 0 \quad \text{or} \quad \frac{\mathrm{d}}{\mathrm{d}x}(x^2y + x^4/4) = 0,$$

which integrates to

$$x^2y + x^4/4 = c$$

or

$$y = \frac{c - x^4}{4x^2}.$$

A1.3.2 Second-Order Equations with Constant Coefficients

The general form of the nth-order linear differential equation with constant coefficients is

$$\frac{\mathrm{d}^n y}{\mathrm{d}x^n} + p_1\frac{\mathrm{d}^{n-1}y}{\mathrm{d}x^{n-1}} + \ldots + p_{n-1}\frac{\mathrm{d}y}{\mathrm{d}x} + p_n y = (D^n + p_1D^{n-1} + \ldots + p_{n-1}D + p_n)y = f(x)$$

where p_1, p_2, … are constants, $f(x)$ is some function of x, and $D \equiv \frac{d}{dx}$. If $f(x) = 0$, the equation is called homogeneous; otherwise, it is called a nonhomogeneous equation. It is important to note that the symbol D is an operator, and it is meaningless unless it is applied to a function of x.

Many of the differential equations of this type that arise in physical problems are of the second order, and we shall consider in detail the solution of the equation

$$\left(\frac{d^2}{dt^2} + a\frac{d}{dt} + b\right)y = (D^2 + aD + b)y = f(t) \tag{A1.51}$$

where a and b are constants, and t is the independent variable. The solution of Equation A1.51 involves first finding the solution of the equation with $f(t)$ replaced by zero, that is,

$$\frac{d^2y}{dt^2} + a\frac{dy}{dt} + by = (D^2 + aD + b)y = 0. \tag{A1.52}$$

Equation A1.52 is called the reduced equation of Equation A1.51.

Now, if y_1 is a solution and A is any constant, then, because $DAy_1 = ADy_1$ and $D^2Ay_1 = AD^2y_1$, we have

$$(D^2 + aD + b)Ay_1 = A(D^2 + aD + b)y_1 = 0$$

and hence Ay_1 is also a solution of Equation A1.52. Next, if y_1 and y_2 are independent solutions of Equation A1.52 and A and B are any constants,

$$D(Ay_1 + By_2) = ADy_1 + BDy_2,\ D^2(Ay_1 + By_2) = AD^2y_1 + BD^2y_2$$

and hence

$$(D^2 + aD + b)(Ay_1 + By_2) = A(D^2 + aD + b)y_1 + B(D^2 + aD + b)y_2 = 0.$$

Thus, $y = Ay_1 + By_2$ is a solution of Equation A1.16 because it contains two arbitrary constants; it is the general solution. A necessary and sufficient condition for two solutions y_1 and y_2 to be linearly independent is that the Wronskian determinant of these functions does not vanish:

$$\begin{vmatrix} y_1 & y_2 \\ \frac{dy_1}{dt} & \frac{dy_2}{dt} \end{vmatrix} \neq 0.$$

Similarly, if y_1, y_2,…, y_n are n linearly independent solutions of the nth-order linear equations, then the general solution is

$$y = A_1y_1 + A_2y_2 + \ldots + A_ny_n$$

where A_1, A_2, …, A_n are arbitrary constants. This is known as the superposition principle.

Suppose that we can find one solution, say, $y_p(t)$, of Equation A1.51:

$$(D^2 + aD + b)y_p(t) = f(t). \tag{A1.51a}$$

Then on defining

$$y_c(t) = y(t) - y_p(t)$$

we find by subtracting Equation A1.51a from Equation A1.51 that

$$(D^2 + aD + b)y_c(t) = 0.$$

That is, $y_c(t)$ satisfies the corresponding homogeneous Equation A1.52, and it is known as the complimentary function $y_c(t)$ of nonhomogeneous Equation A1.51 while the solution $y_p(t)$ is called a particular integral of Equation A1.51. Thus, the general solution of Equation A1.51 is given by

$$y(t) = Ay_c(t) + By_p(t). \tag{A1.53}$$

A1.3.2.1 Finding the Complimentary Function

Clearly the complimentary function is independent of $f(t)$ and hence has nothing to do with the behavior of the system in response to the external applied influence. What it does represent is the free motion of the system. Thus, for example, even without external forces applied, the spring can oscillate because of any initial displacement and/or velocity.

Similarly, had the capacitor been charged already at $t = 0$, the circuit will subsequently display current oscillations even though there is no applied voltage.

In order to solve Equation A1.52 for $y_c(t)$, we first consider the linear first-order equation

$$a\frac{\mathrm{d}y}{\mathrm{d}t} + by = 0.$$

Separating the variables and integrating, we obtain

$$y = A\exp(-bt/a)$$

where A is an arbitrary constant of integration. This solution suggests that Equation A1.52 might be satisfied by an expression of the type

$$y = \exp(pt)$$

where p is a constant. Putting this into Equation A1.52, we have

$$\exp(pt)[p^2 + ap + b] = 0.$$

Therefore, $y = \exp(pt)$ is a solution of Equation A1.52 if

$$p^2 + ap + b = 0.$$

This is called the auxiliary (or characteristic) equation of Equation A1.52. Solve it to obtain

$$p_1 = \frac{-a + \sqrt{a^2 - 4b}}{2}, \quad p_2 = \frac{-a - \sqrt{a^2 - 4b}}{2}. \tag{A1.54}$$

We now distinguish between the cases in which the roots are real and distinct, complex, or coincident.

(i) The Case of Real and Distinct Roots ($a^2 - 4b > 0$)
In this case, we have two independent solutions $y_1 = \exp(p_1 t)$ and $y_2 = \exp(p_2 t)$, and the general solution of Equation A1.16 is a linear combination of these two:

$$y = A\exp(p_1 t) + B\exp(p_2 t) \tag{A1.55}$$

where A and B are constants.

(ii) The Case of Complex roots ($a^2 - 4b < 0$)
If the roots p_1 and p_2 of the auxiliary equation are imaginary, the solution given by Equation A1.54 is still correct. In order to give the solutions in terms of real quantities, we can use the Euler relationships to express the exponentials. If we let $r = -a/2$ and $is = \sqrt{a^2 - 4b}/2$, then

$$\exp(p_1 t) = \exp(rt)\exp(ist) = \exp(rt)[\cos st + i\sin st]$$

$$\exp(p_2 t) = \exp(rt)\exp(-ist) = \exp(rt)[\cos st - i\sin st]$$

and the general solution can be written as

$$\begin{aligned} y &= A\exp(p_1 t) + B\exp(p_2 t) \\ &= e^{rt}[(A+B)\cos\ st + i(A-B)\sin\ st] \\ &= e^{rt}[A_0\cos\ st + B_0\sin\ st] \end{aligned} \tag{A1.56}$$

with $A_0 = A + B$, $B_0 = i(A - B)$.

The solution A1.56 may be expressed in a slightly different and often more useful form by writing $B_0/A_0 = \tan\delta$. Then,

$$y = \left(A_0^2 + B_0^2\right)^{1/2} e^{rt}(\cos\delta\ \cos\ st + \sin\delta\ \sin\ st) = Ce^{rt}\cos(st - \delta) \tag{A1.57}$$

where C and δ are arbitrary constants.

(iii) The Case of Coincident Roots ($a^2 - 4b = 0$)
When $a^2 = 4b$, the auxiliary equation yields only one value for p, namely, $p = \alpha = -a/2$, and, hence, the solution $y = Ae^{\alpha t}$. This is not the general solution, as it does not contain the necessary two arbitrary constants. In order to obtain the general solution, we proceed as follows. Assume that $y = ve^{\alpha t}$, where v is a function of t to be determined. Then,

$$y' = v'e^{\alpha t} + \alpha\, ve^{\alpha t},\ y'' = v''e^{\alpha t} + 2\alpha v'e^{\alpha t} + \alpha^2 ve^{\alpha t}.$$

Substituting for y, y', and y'' in the differential equation, we have

$$e^{\alpha t}\{v'' + 2\alpha v' + \alpha^2 v + a(v' + \alpha\, v) + bv\} = 0$$

and hence

$$v'' + v'(a + 2\alpha) + v(\alpha^2 + a\alpha + b) = 0.$$

Now,

$$\alpha^2 + a\alpha + b = 0 \text{ and } a + 2\alpha = 0$$

so that

$$v'' = 0.$$

Hence, integrating

$$v = A\,t + B$$

where A and B are arbitrary constants, the general solution of Equation A1.16 is

$$y = (A\,t + B)e^{\alpha\,t} \tag{A1.58}$$

A1.3.2.2 Finding the Particular Integral

The particular integral is a solution of Equation A1.51 that takes the term $f(t)$ on the right-hand side into account. The complementary function is transient in nature, so from a physical point of view, the particular integral will usually dominate the response of the system at large times.

The method of determining the particular integral is to guess a suitable functional form containing arbitrary constants and then to choose the constants to ensure it is indeed the solution. If our guess is incorrect, then no values of these constants will satisfy the differential equation, and so we have to try a different form. Clearly this procedure could take a long time; fortunately, there are some guiding rules on what to try for the common examples of $f(t)$:

1. $f(t)$ = polynomial in t

   If $f(t)$ is a polynomial in t with the highest power t^n, then the trial particular integral is also a polynomial in t with terms up to the same power. Note that the trial particular integral is a power series in t even if $f(t)$ contains only a single term At^n.
2. $f(t) = Ae^{k\,t}$

   The trial particular integral is $y = Be^{k\,t}$.
3. $f(t) = A \sin kt$ or $A \cos kt$

   The trial particular integral is $y = A \sin kt + C \cos kt$.

   That is, even though $f(t)$ contains only a sine or cosine term, we need both sine and cosine terms for the particular integral.
4. $f(t) = A\,e^{\alpha\,t}\sin \beta t$ or $A\,e^{\alpha\,t}\cos \beta t$.

   The trial particular integral is $y = e^{\alpha\,t}(B \sin \beta t + C \cos \beta t)$.
5. $f(t)$ is a polynomial of order n in t, multiplied by $e^{k\,t}$.

   The trial particular integral is a polynomial in t with coefficients to be determined, multiplied by $e^{k\,t}$.
6. $f(t)$ is a polynomial of order n in t, multiplied by $\sin kt$.

   The trial particular integral is

$$y = \sum_{j=0}^{n} (B_j \sin kt + C_j \cos kt)t^j.$$

Can we try

$$y = (B \sin kt + C \cos kt) \sum_{j=0}^{n} D_j \ t^j$$

The answer is no. Do you know why?

If the trial particular integral or part of it is identical to one of the terms of the complementary function, then the trial particular integral must be multiplied by an extra power of t. Therefore, we need to find the complementary function before we try to work on the particular integral. What do we mean by "identical in form"? It means that the ratio of their t-dependencies is a constant. Thus, $-2e^{-t}$ and $A\ e^{-t}$ are identical in form, but e^{-t} and e^{-2t} are not.

PROBLEMS

1. Given the vector $\boldsymbol{A} = (2, 2, -1)$ and $\boldsymbol{B} = (6, -3, 2)$, determine
   (a) $6\boldsymbol{A} - 3\boldsymbol{B}$
   (b) $A^2 + B^2$
   (c) $\boldsymbol{A} \cdot \boldsymbol{B}$
   (d) The angle between $\boldsymbol{A}$ and $\boldsymbol{B}$
   (e) The direction cosines of $\boldsymbol{A}$
   (*f*) The component of $\boldsymbol{B}$ in the direction of $\boldsymbol{A}$
2. Find a unit vector perpendicular to the plane of $\boldsymbol{A} = (2, -6, -3)$ and $\boldsymbol{B} = (4, 3, -1)$.
3. Prove that
   (a) The median to the base of an isosceles triangle is perpendicular to the base.
   (b) An angle inscribed in a semicircle is a right angle.
4. (a) Given two vectors $\boldsymbol{A} = (2, 1, -1)$ and $\boldsymbol{B} = (1, -1, 2)$, find
   (a) $\boldsymbol{A} \times \boldsymbol{B}$
   (b) Find a unit vector perpendicular to the plane containing vectors $\boldsymbol{A}$ and $\boldsymbol{B}$
5. Prove (a) the law of sines for plane triangles, and (b) Equation A1.16.
6. Evaluate $\left(2\hat{e}_1 - 3\hat{e}_2\right) \cdot \left[\left(\hat{e}_1 + \hat{e}_2 - \hat{e}_3\right) \times \left(3\hat{e}_1 - \hat{e}_3\right)\right]$.
7.
   (a) Prove that a necessary and sufficient condition for the vectors $\boldsymbol{A}$, $\boldsymbol{B}$, and $\boldsymbol{C}$ to be coplanar is that $\boldsymbol{A} \cdot (\boldsymbol{B} \times \boldsymbol{C}) = 0$.
   (b) Find an equation for the plane determined by the three points $P_1(2, -1, 1)$, $P_2(3, 2, -1)$, and $P_3(-1, 3, 2)$.
8. Solve the following equations:
   (a) $x\, \mathrm{d}y/\mathrm{d}x + y^2 = 1$
   (b) $\mathrm{d}y/\mathrm{d}x = (x + y)^2$
9. Show that $(3x^2 + y \cos x)\mathrm{d}x + (\sin x - 4y^3)\mathrm{d}y = 0$ is an exact differential equation and find its general solution.
10. Find a constant α such that $(x + y)^\alpha$ is an integrating factor of the equation

$$(4x^2 + 2xy + 6y)dx + (2x^2 + 9y + 3x)dy = 0.$$

    What is the solution of this equation?
11. Solve $\mathrm{d}y/\mathrm{d}x + y = y^3 x$.

12. Solve
    (a) The equation $(D^2 - D - 12)y = 0$ with boundary conditions $y = 0$, $Dy = 3$ when $t = 0$.
    (b) The equation $(D^2 + 2D + 3)y = 0$ with boundary conditions $y = 2$, $Dy = 0$ when $t = 0$.
    (c) The equation $(D^2 - 2D + 1)y = 0$ with boundary conditions $y = 5$, $Dy = 3$ when $t = 0$.
13. Find the particular integral of $(D^2 + 2D - 1)y = 3 + t^3$ (Atkin).
14. Find the particular integral of $(2D^2 + 5D + 7) = 3e^{2t}$.
15. Find the particular integral of $(3D^2 + D - 5)y = \cos 3t$.

Appendix 2: D'Alembert's Principle and Lagrange's Equations

A2.1 VIRTUAL DISPLACEMENT, VIRTUAL WORK, AND PRINCIPLE OF VIRTUAL WORK

D'Alembert's principle is based on the concepts of virtual displacement and virtual work. What do we mean by virtual displacement and virtual work?

A2.1.1 Virtual Displacement

A virtual displacement, denoted by δr_1, is defined as an imagined, infinitesimal, instantaneous displacement of the coordinates that is consistent with the constraints. In other words, virtual displacements have three characteristics: (1) They are infinitesimal, (2) they occur at a given instant in time (i.e., no time changes $\delta t = 0$), and (3) they are to be carried out in a manner consistent with the constraints. Virtual displacement $\delta \boldsymbol{r}_j$ will not necessarily correspond to any actual displacement $d\boldsymbol{r}_j$ occurring in dt during which the forces and constraints may be changing.

A2.1.2 Virtual Work and Principle of Virtual Work

Now consider a time-independent (scleronomic) system of N particles. If the system is in equilibrium, then the total force on each particle must be zero: $\boldsymbol{F}_j = 0$ for all j; evidently the scalar product $\boldsymbol{F}_j \cdot \delta \boldsymbol{r}_j$ that is the virtual work done on the jth particle by $\boldsymbol{F}_j$ in a virtual displacement $\delta \boldsymbol{r}_j$ is also zero. The total virtual work done on the system, which is the sum of the preceding vanishing products over all N particles, must likewise vanish:

$$\delta W = \sum_{j=1}^{N} \vec{F}_j \cdot \delta \vec{r}_j = 0. \tag{A2.1}$$

The total force $\boldsymbol{F}_j$ acting on the jth particle of the system may be an externally applied force, $\boldsymbol{F}_j^{(e)}$, or forces of constraints, $\boldsymbol{f}_j$. Equation A2.1 now becomes

$$\delta W = \sum_{j=1}^{N} \vec{F}_j^{(e)} \cdot \delta \vec{r}_j + \sum_{j=1}^{N} \vec{f}_j \cdot \delta \vec{r}_j = 0. \tag{A2.2}$$

Many of the constraints that commonly occur, such as sliding on a frictionless surface and rolling contact without slipping, do no work under a virtual displacement, that is,

$$\vec{f}_j \cdot \delta \vec{r}_j = 0. \tag{A2.3}$$

This is practically the only possible situation we can imagine where the forces of constraints must be perpendicular to $\delta \boldsymbol{r}_j$; otherwise, the system could be spontaneously accelerated by the forces of constraint alone, and we know that this does not occur. Equation A2.3 will no longer hold if frictional forces are taken into account. If we exclude frictional force from consideration and restrict ourselves to a system for which the virtual work of the forces of constraint vanishes, then we have

$$\sum_{j=1}^{N} \vec{f}_j \cdot \delta\vec{r}_j = 0 \tag{A2.4}$$

and as a result, Equation A2.2 reduces to

$$\delta W = \sum_{j=1}^{N} \vec{F}_j^{(e)} \cdot \delta\vec{r}_j = 0. \tag{A2.5}$$

Thus, if for any arbitrary virtual displacement $\delta \boldsymbol{r}_j$, the virtual work of the forces of constraint vanishes; the virtual work of the external applied force $\boldsymbol{F}_j^{(e)}$ also vanishes, and the system is in equilibrium. Equation A2.5 is known as the principle of virtual work.

It should be noted that the coefficients of $\delta \boldsymbol{r}_j$ can no longer be set equal to zero. This is because $\delta \boldsymbol{r}_j$ are not all independent but are now connected by the equations of constraints. To equate the coefficients to zero, we transform Equation A2.5 into a form involving the virtual displacements δq_i, which are independent. From Equation 4.1, we have

$$\delta\vec{r}_i = \sum_{j=1}^{n} \frac{\partial \vec{r}_i}{\partial q_j} \delta q_j. \tag{A2.6}$$

Note that no variation of time, δt, is involved here by definition of virtual displacement. Substituting Equation A2.6 into Equation A2.5, we obtain

$$\delta W = \sum_{i=1}^{N} \vec{F}_i^{(e)} \cdot \delta\vec{r}_i = \sum_{i=1}^{N}\sum_{j=1}^{n} \vec{F}_i^{(e)} \cdot \frac{\partial \vec{r}_i}{\partial q_j} \delta q_j = \sum_{j=1}^{n} Q_j \delta q_j = 0 \tag{A2.7}$$

where

$$Q_j = \sum_{i=1}^{N} \vec{F}_i^{(e)} \cdot \frac{\partial \vec{r}_i}{\partial q_j} \tag{A2.8}$$

is called the generalized force corresponding to the coordinate q_j. In general, Q_j will not have the dimensions of force; however, the product $Q_j q_j$ will always have the dimensions of work.

Because δq_j are arbitrary and independent, the condition of equilibrium is

$$Q_j = 0, \text{ for all } j. \tag{A2.9}$$

The importance of the principle of virtual work is that it constitutes a single principle upon which all statics can be based. The following two examples will suffice to fix the basic ideas of the principle just outlined. After the study of these two examples, we should be able to apply it with ease and confidence.

Example A2.1

A uniform plank of mass M is leaning against a smooth wall and makes an angle α with the smooth floor as shown in Figure A1.1. The lower end of the plank is connected to the base of the wall with an inextensible string whose mass is negligible. What is the tension in the string?

Solution:

This example shows a scleronomic system with workless constraints. The external constraint forces are the wall and floor reactions $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$. They do no work.

We take the length of the plank as $2b$. The midpoint O of the plank is at a distance y above the floor, and the lower end of the plank is at a distance x from the wall, where

$$x = 2b \cos\alpha,\ y = b \sin\alpha. \qquad \text{(A2.10)}$$

The forces acting on the plank are

T = tension in the string
Mg = weight of the plank
N_1, N_2 = normal reactions of the wall and the floor, respectively

We now imagine a horizontal displacement δx of the lower end of the plank from B to D. The change in the level of O is δy. From Equation A2.10, we see that

$$|\delta y| = b \cos \alpha\ \delta\alpha,\ |\delta x| = 2b \sin \alpha\ \delta\alpha. \qquad \text{(A2.11)}$$

Now the constraint forces $\boldsymbol{N}_1$ and $\boldsymbol{N}_2$ do no work because the wall and the floor are frictionless. Then the principle of virtual work gives

$$-T|\delta x| + Mg|\delta y| = 0 \qquad \text{(A2.12)}$$

where the sign convention is that the tension T and the horizontal virtual δx are in opposite directions, so it does negative work. Conversely, while δy and Mg are in the same direction, Mg does positive work.

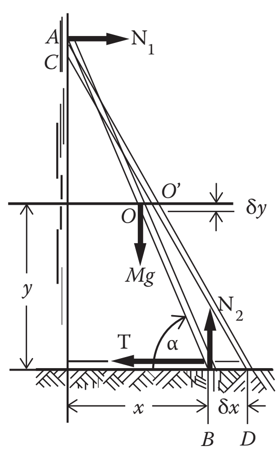


FIGURE A2.1 Example 2.1.

Substituting Equation A2.11 into Equation A2.12 and solving for T, we find

$$T = -2\,Mg \cot \alpha.$$

Example A2.2

A ring of mass m slides on a smooth vertical rod (Figure A2.2). Attached to the ring is a light, inextensible string passing over a smooth peg distance a from the rod. At the other end of the string is a mass M ($M > m$). The ring is released from rest at the same level as the peg. If h is the maximum distance that the ring will fall, determine h in terms of M, a, and m.

Solution:

If b is the length of the string, then clearly

$$x + \sqrt{h^2 + a^2} = b. \tag{A2.13}$$

Now, imagine a vertical displacement δh of the ring along the rod. The change in x is connected to h by the following equation:

$$\delta x + (h^2 + a^2)^{-1/2} h \delta h = 0$$

or

$$|\delta x| = (h^2 + a^2)^{-1/2} h |\delta h|. \tag{A2.14}$$

The constraints over the rod and peg do no work, so the principle of virtual work gives

$$Mg|\delta x| + (-mg|\delta h|) = 0 \tag{A2.15}$$

where the sign convention is that x and Mg are in the same direction, so Mg does positive work; while h and mg are in opposite directions, mg does negative work.

Substituting Equation A2.14 into Equation A2.15, we obtain

$$h = \frac{ma}{\sqrt{M^2 - m^2}}.$$

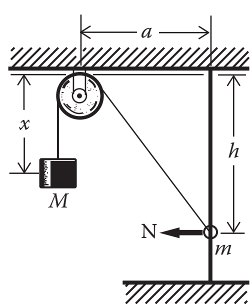


FIGURE A2.2 Example 2.2.

A2.2 D'ALEMBERT'S PRINCIPLE

The principle of virtual work deals only with statics. We are tempted at this point to find a principle that involves the general motion of the system. Such a principle was first suggested by Bernoulli and then developed by D'Alembert. We first write Newton's second law of motion in the following form:

$$\vec{F}_i - \mathrm{d}\vec{p}_i/\mathrm{d}t = 0\,.$$

If we regard $-\mathrm{d}\boldsymbol{p}_i/\mathrm{d}t$ as a force, an inertial or reversed effective force (named by Bernoulli and D'Alembert), which added to $\boldsymbol{F}_i$ produces equilibrium, then dynamics will reduce to statics. Now, instead of Equation A2.5, we have

$$\sum_{i=1}^{N} (\vec{F}_i - \dot{\vec{p}}_i)\cdot \delta\vec{r}_i = 0. \tag{A2.16}$$

We now resolve $\boldsymbol{F}_i$ into applied force $\boldsymbol{F}_j^{(e)}$ and forces of constraint $\boldsymbol{f}_i$, and if we again restrict ourselves to a system for which the virtual work of the forces of constraint vanishes, we obtain

$$\sum_{i=1}^{N} (\vec{F}_i^{(e)} - \dot{\vec{p}}_i)\cdot \delta\vec{r}_i = 0. \tag{A2.17}$$

This equation is the Lagrangian form of D'Alembert's principle. The superscript e in Equation A2.17 can now be dropped without ambiguity.

A2.3 LAGRANGE'S EQUATIONS: FROM D'ALEMBERT'S PRINCIPLE

Lagrange selected D'Alembert's principle as the starting point of his "Mecanique Analytique" and obtained the equations of motion, now known as Lagrange's equations, from it. This is what we now proceed to do. We shall first transform Equation A2.17 into an equation involving virtual displacements of the generalized coordinates δq_i, which are independent of each other. In terms of the generalized coordinates, the virtual work done by the force $\boldsymbol{F}_i$ (the external applied force) becomes $\sum_j Q_j \delta q_j$ as shown in Equation A2.7.

$$\delta W = \sum_{i=1}^{N} \vec{F}_i \cdot \delta\vec{r}_i = \sum_{j=1}^{n} Q_j \delta q_j \tag{A2.18}$$

and Q_j is given by Equation A2.8. We now write the inertial force term in Equation A1.17 as

$$\begin{aligned}\sum_{i=1}^{N} \dot{\vec{p}}_i \cdot \delta\vec{r}_i &= \sum_{i=1}^{N}\sum_{j=1}^{n} m_i \ddot{\vec{r}}_i \cdot \frac{\partial \vec{r}_i}{\partial q_j}\delta q_j \\ &= \sum_{j=1}^{n} \delta q_j \sum_{i=1}^{N} m_i \left[\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{r}}_i \cdot \frac{\partial \vec{r}_i}{\partial q_j}\right) - \dot{\vec{r}}_i \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \vec{r}_i}{\partial q_j}\right)\right]\end{aligned} \tag{A2.19}$$

and from Equation 4.1, we find

$$\mathrm{d}\vec{r}_i = \sum_{j=1}^{n} \frac{\partial \vec{r}_i}{\partial q_j}\mathrm{d}q_j + \frac{\partial \vec{r}_i}{\partial t}\mathrm{d}t \quad \text{and} \quad \dot{\vec{r}}_i = \sum_{j=1}^{n} \frac{\partial \vec{r}_i}{\partial q_j}\mathrm{d}\dot{q}_j + \frac{\partial \vec{r}_i}{\partial t}. \tag{A2.20}$$

The partial derivatives in Equation A2.20 are themselves functions of the generalized coordinates q_i and the time. As a result, the particle velocities have the following functional form:

$$\dot{\vec{r}}_i = \dot{\vec{r}}_i(q_1,\ldots,\ q_n;\ \dot{q}_1,\ldots,\ \dot{q}_n), \quad i = 1,\ 2,\ldots,N.$$

Moreover, Equation A2.20 provides an explicit function of the indicated variables and shows that $\dot{\vec{r}}_i$ in fact depends linearly on the generalized velocities $\dot{q}_j$. Thus, we can readily evaluate the partial derivative $\partial\dot{\vec{r}}_i/\partial\dot{q}_j$ to obtain

$$\partial\dot{\vec{r}}_i/\partial\dot{q}_j = \partial\vec{r}_i/\partial q_j. \tag{A2.21}$$

Note that now the independent variables appearing in parentheses in Equation A2.19 are physically independent in the sense that each can be specified independently at a given instant of time. The subsequent motion of the system is then, of course, determined by the equations of motion. We now substitute Equation A2.21 into the first term on the right-hand side of Equation 5.24, with the result that

$$\sum_{i=1}^{N} m_i \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{r}}_i \cdot \frac{\partial \vec{r}_i}{\partial q_j}\right) = \sum_{i=1}^{N} m_i \frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{\vec{r}}_i \cdot \frac{\partial \dot{\vec{r}}_i}{\partial \dot{q}_j}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial}{\partial \dot{q}_j}\left(\frac{1}{2}\sum_{i=1}^{N} m_i \left|\dot{\vec{r}}_i\right|\right) = \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} \tag{A2.22}$$

where $T = 2m_i^2\dot{r}^2$ is the total kinetic energy of the system.

We can also rewrite the second term on the right-hand side of Equation A2.19 as

$$\begin{aligned}\sum_{i=1}^{N} m_i\dot{\vec{r}}_i \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \vec{r}_i}{\partial q_j}\right) &= \sum_{i=1}^{N} m_i\dot{\vec{r}}_i \cdot \left[\sum_k \frac{\partial^2 \vec{r}_i}{\partial q_k \partial q_j}\dot{q}_j + \frac{\partial^2 \vec{r}_i}{\partial q_j \partial t}\right] \\ &= \sum_{i=1}^{N} m_i\dot{\vec{r}}_i \cdot \frac{\partial \dot{\vec{r}}}{\partial q_j} = \frac{\partial}{\partial q_j}\left(\sum_i \frac{1}{2} m_i \dot{r}_i^2\right) = \frac{\partial T}{\partial q_j}.\end{aligned} \tag{A2.23}$$

With Equations A2.22 and A2.23, Equation A2.19 becomes

$$\sum_i^N \dot{\vec{p}}_i \cdot \delta\vec{r}_i = \sum_j \delta q_j \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j}\right). \tag{A2.24}$$

Combining this with Equation A2.18, D'Alembert's principle gives

$$\sum_j^n \left(\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} - Q_j\right)\delta q_j = 0.$$

δq_j are all independent for a holonomic system, and each of the coefficients must be zero separately, from which it follows that

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} - \frac{\partial T}{\partial q_j} = Q_j, \quad j = 1, 2, \ldots, n \tag{A2.25}$$

there being n equation in all and n being the number of degrees of freedom of the system.

Equation A2.25 can be simplified further if the external forces $\boldsymbol{F}_i$ are conservative: $\boldsymbol{F}_i = -\nabla_i V$. Then,

$$Q_j = \sum_{i=1}^{N} \vec{F}_i \cdot \frac{\partial \vec{r}_i}{\partial q_j} = -\sum_{i=1}^{N} \nabla_i V \cdot \frac{\partial \vec{r}_i}{\partial q_j} = -\frac{\partial V}{\partial q_j} \tag{A2.26}$$

and Equation A2.25 becomes

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial T}{\partial \dot{q}_j} - \frac{\partial (T - V)}{\partial q_j} = 0, \quad j = 1, 2, \ldots, n.$$

Now, if the potential function V is a function of position only, then it is independent of the generalized velocities $\dot{q}_j$. We can now include a term in V in the first term on the right-hand side of the preceding equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial (T - V)}{\partial \dot{q}_j} - \frac{\partial (T - V)}{\partial q_j} = 0, \quad j = 1, 2, \ldots, n.$$

Following Lagrange, we introduce a new function L defined by

$$L(q_i, \dot{q}_i, t) = T(q_i, \dot{q}_i, t) - V(q_i). \tag{A2.27}$$

This function is called the Lagrangian function (or the Lagrangian) of the system. In terms of this function, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = 0, \quad j = 1, 2, \ldots, n \tag{A2.28}$$

where n is the number of degrees of freedom of the system. These n second-order differential equations are called Lagrange's equations for a conservative, holonomic, dynamic system. If some of the forces acting on the system are not conservative, Lagrange's equation can be written in the form

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_j} - \frac{\partial L}{\partial q_j} = Q'_j, \quad j = 1, 2, \ldots, n \tag{A2.29}$$

where L contains the potential of the conservative forces as before, and Q'_j represent the nonconservative forces that do not arise from a potential. Time-varying force functions and frictional forces are typical nonconservative forces.

Appendix 3: Derivation of Hamilton's Principle from D'Alembert's Principle

We consider, for simplicity, a system of N particles of masses m_i, located at the point $\boldsymbol{r}_i$ and acted upon by external forces $\boldsymbol{F}_i$. By D'Alembert's principle, we have

$$\sum_{i=1}^{N}(\vec{F}_i - m_i\ddot{\vec{r}}_i)\cdot\delta\vec{r}_i = 0. \tag{A3.1}$$

This is the dynamic condition to be satisfied by the impressed forces and the internal forces for arbitrary $\delta\vec{r}_i$ consistent with the constraints. We may interpret the virtual displacements $\delta\vec{r}_i$ as the small variations in the following sense: a particle in its motion will take a certain path described by $\boldsymbol{r} = \boldsymbol{f}(t)$ under the action of external forces. Varied paths described in the same time between points in space may be imagined by changing the actual path by small variation $\delta\boldsymbol{r}$ for corresponding times.

We observe that the first term on the right-hand side is the virtual work done by the external applied force $\boldsymbol{F}_I$

$$\delta W = \sum_{i=1}^{N}\vec{F}_i\cdot\delta\vec{r}_i. \tag{A3.2}$$

Furthermore, the second term in Equation A3.1 can be expressed as

$$\sum_{i=1}^{N} m_i\ddot{\vec{r}}_i\cdot\delta\vec{r}_i = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\vec{r}_i\right) - \sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\frac{\mathrm{d}}{\mathrm{d}t}\delta\vec{r}_i. \tag{A3.3}$$

Now, because each virtual displacement is to take place without the passage of time, the variation and differentiation operations are interchangeable with respect to time:

$$\frac{\mathrm{d}}{\mathrm{d}t}(\delta\vec{r}_i) = \delta\dot{\vec{r}}_i$$

Hence, Equation A3.3 becomes

$$\begin{aligned}\sum_{i=1}^{N} m_i\ddot{\vec{r}}_i\cdot\delta\vec{r}_i &= \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\vec{r}_i\right) - \sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\dot{\vec{r}}_i = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\vec{r}_i\right) \\ &\quad - \delta\left(\frac{1}{2}\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\dot{\vec{r}}\right) = \frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\vec{r}_i\right) - \delta T.\end{aligned} \tag{A3.4}$$

Putting this and Equation A3.2 into Equation A3.1, we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\vec{r}_i\right)=\delta T+\delta W.$$

Integrating with respect to time t between the limits t_1 and t_2,

$$\int_{t_1}^{t_2}(\delta T+\delta W)\,\mathrm{d}t=\left[\sum_{i=1}^{N} m_i\dot{\vec{r}}_i\cdot\delta\vec{r}_i\right]_{t_1}^{t_2}. \tag{A3.5}$$

If the configuration of the system is specified at times t_1 and t_2 such that the dynamic path and all imagined variations of this path coincide, then $\delta \boldsymbol{r}_I(t_1) = \delta \boldsymbol{r}_I(t_2) = 0$; Equation A3.5 then becomes

$$\int_{t_1}^{t_2}(\delta T+\delta W)\,\mathrm{d}t=0. \tag{A3.6}$$

Up to this point, the configuration of the system is given in terms of the position vectors $\boldsymbol{r}_i$ of the particles. In order to make the values of δT and δW independent of the coordinates for a given virtual displacement and time, we need to transform to the generalized coordinates q_i. From the transformation equations

$$r_i = r_i\,(q_1, q_2,\ldots,q_n, t)$$

we have

$$\delta W=\sum_{i=1}^{N}\vec{F}_i\cdot\delta\vec{r}_i=\sum_{i=1}^{N}\vec{F}_i\cdot\sum_{j=1}^{n}\frac{\delta\vec{r}_i}{\delta q}\delta q_j=\sum_{j=1}^{n}Q_j\delta q_j \tag{A3.7}$$

where

$$Q_j=\sum_{i=1}^{N}F_i\cdot\frac{\delta\vec{r}_i}{\delta q_j}. \tag{A3.8}$$

T can also be expressed in terms of the generalized coordinates:

$$T=T(q_i,\dot{q}_i;t).$$

With these, we can rewrite Equation A3.6 in the form

$$\int_{t_1}^{t_2}\left(\delta T+\sum_{j=1}^{n}Q_j\delta q_j)\mathrm{d}t\right)=0 \tag{A3.9}$$

where $\delta q_j(t_1) = \delta q_j(t_2) = 0$. For a constrained system, δq's must also conform to the instantaneous constraints.

Equation A3.9 is often considered to be a generalized version of Hamilton's principle. It can also be considered an integral form of D'Alembert's principle. Equation A3.9 is independent of the choice of coordinates with which we describe the system; this is its advantage over D'Alembert's principle.

If the external forces are conservative forces, $\vec{F}_i = -\nabla_i V$, then

$$\delta W = \sum_j Q_j \delta q_j = \sum_{j,i} \vec{F}_i \cdot \frac{\partial \vec{r}_i}{\partial q_j} \delta q_j = -\sum_j \frac{\partial V}{\partial q_j} \delta q_j = -\delta V$$

and Equation A3.9 becomes

$$\int_{t_1}^{t_2} (\delta T - \delta V)\, \mathrm{d}t = \int_{t_1}^{t_2} \delta L\, \mathrm{d}t = 0 \tag{A3.10}$$

where $\delta q_j(t_1) = \delta q_j(t_2) = 0$.

For a holonomic system, the variation and integration operations can be interchanged. Then Equation A2.10 can be written as

$$\delta I = \delta \int_{t_1}^{t_2} L\, \mathrm{d}t = 0 \tag{A3.11}$$

with $\delta q_j(t_1) = \delta q_j(t_2) = 0$. This is Hamilton's principle for holonomic, conservative systems.

Appendix 4: Noether's Theorem

In Chapters 4 and 5, we note that symmetries of the Lagrangian or Hamiltonian gave rise to constants of the motion; such parameters are of utmost importance in the analysis. However, the constants of the motion do not always come from the obvious symmetries of the Lagrangian nor do they always have a simple form. Often they are expressed by complicated functions of coordinates and momenta, which are invariant in time.

It is therefore desirable to develop a general approach that is not limited by the specific details of a given situation. Such a formalism was found in 1918 by the noted mathematician Emmy Noether, now known as Noether's theorems. Her approach has been simplified and popularized by several writers in recent years. Here we are only concerned with Noether's first theorem, which can be stated as

> To every infinitesimal transformation that leaves invariant the action integral there corresponds a conserved quantity.

Such transformations, which leave invariant the action integral (and thus leave the equation of motion invariant in form), are called canonical or symmetry transformations.

Now we have a physical system described by generalized coordinates q_i and velocities $\dot{q}_i$. The Lagrangian $L(q, \dot{q}, t)$ is assumed to be known and will provide the correct equations of motion for the system when substituted into Lagrange's equations. Suppose that it is possible to find a new set of generalized coordinates Q_i and T through infinitesimal symmetry transformations:

$$q_i' = q_i + \delta q_i \quad t' = t + \delta t. \tag{A4.1}$$

In general, δt and δq_i are of the form

$$\delta q_i = \varepsilon Q(q_i, \dot{q}_i, t) \quad \delta t = \varepsilon T(q_i, \dot{q}_i, t) \tag{A4.2}$$

where ε is an infinitesimally small parameter of the transformations and Q and T are functions of the corresponding variables.

The action integral in terms of the primed variables is

$$I' = \int_{t_1'}^{t_2'} L'(q_i', \dot{q}_i', t')\,\mathrm{d}t'. \tag{A4.3}$$

Note that the Lagrangian of a system is defined only to within an additive total time derivative of any function of coordinates and time, and the equations of motion are unchanged if the Lagrangian is modified as follows:

$$L \to L(q_i, \dot{q}_i, t) + \mathrm{d}f(q_i, t)/\mathrm{d}t.$$

We therefore require that L' arising from transformations (Equation A4.1) differs from the original L by

$$L'(q_i', \dot{q}_i', t') = L(q_i, \dot{q}_i, t) + \mathrm{d}f(q_i, t)/\mathrm{d}t. \tag{A4.4}$$

It follows that the change in the action integral resulting from the transformations (Equation A4.1) is given by

$$\Delta I = \int_{t_1'}^{t_2'} L'(q_i', \dot{q}_i', t')\,\mathrm{d}t' - \int_{t_1}^{t_2} L(q_i, \dot{q}_i, t)\,\mathrm{d}t$$
$$= \int_{t_1'}^{t_2'} L(q_i', \dot{q}_i', t')\,\mathrm{d}t' + \int_{t_1'}^{t_2'} \frac{\mathrm{d}}{\mathrm{d}t'} f(q_i', t')\,\mathrm{d}t' - \int_{t_1}^{t_2} L(q_i, \dot{q}_i, t)\,\mathrm{d}t$$

which vanishes provided the transformations (Equation A4.1) are canonical transformations.

Replacing the primed variables according to Equation A4.1, the last equation becomes

$$\Delta I = \int_{t_1}^{t_2} L(q_i + \delta q_i, \dot{q}_i + \delta\dot{q}_i, t + \delta t)\left(1 + \frac{\mathrm{d}\delta t}{\mathrm{d}t}\right)\mathrm{d}t - \int_{t_1}^{t_2} L(q_i, \dot{q}_i, t)\,\mathrm{d}t$$
$$+ \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t} f(q_i + \delta q_i, t + \delta t)\frac{\mathrm{d}t}{\mathrm{d}t'}\,\mathrm{d}t'.$$

Performing Taylor's expansion of the integrands of the last equation about the point $q_i(t)$ yields

$$\Delta I = \int_{t_1}^{t_2} \left\{ L(q_i, \dot{q}_i, t) + \sum_i \frac{\partial L}{\partial q_i}\Big|_{(q_i,\dot{q}_i,t)} \delta q_i + \sum_i \frac{\partial L}{\partial \dot{q}_i}\Big|_{(q_i,\dot{q}_i,t)} \delta\dot{q}_i + \frac{\partial L}{\partial t}\Big|_{(q_i,\dot{q}_i,t)} \delta t + \text{H.O.} \right\}\left(1 + \frac{\mathrm{d}\delta t}{\mathrm{d}t}\right)\mathrm{d}t$$
$$- \int_{t_1}^{t_2} L(q_i, \dot{q}_i, t)\,\mathrm{d}t + \int_{t_1}^{t_2} \frac{\mathrm{d}}{\mathrm{d}t}\left\{ f + \sum_i \frac{\partial f}{\partial q_i}\Big|_{(q_i,\dot{q}_i,t)} \delta q_i + \frac{\partial f}{\partial t}\Big|_{(q_i,\dot{q}_i,t)} \delta t \right\}\mathrm{d}t$$

where H.O. stands for the infinitesimal terms of higher order. Neglecting the H.O. terms, there remains

$$\Delta I = \int \left\{ \sum_i \frac{\partial L}{\partial q_i}\delta q_i + \sum_i \frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i + \frac{\partial L}{\partial t}\delta t + L\frac{\mathrm{d}\delta t}{\mathrm{d}t} \right\}\mathrm{d}t$$
$$+ \int \frac{\mathrm{d}}{\mathrm{d}t}\delta f(q_i, t)\,\mathrm{d}t + \left[f(q(t_2), t - f(q(t_1), t) \right].$$

The last term in the square bracket gives zero on variation and thus it can be dropped, and there remains

$$\Delta I = \int \left[\sum_i \frac{\partial L}{\partial q_i}\delta q_i + \sum_i \frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i + \frac{\partial L}{\partial t}\delta t + L\frac{\mathrm{d}\delta t}{\mathrm{d}t} \right]\mathrm{d}t + \int \frac{\mathrm{d}}{\mathrm{d}t}\delta f(q_i, t)\,\mathrm{d}t. \qquad \text{(A4.5)}$$

Now,

$$\delta\dot{q}_i = \delta\left(\frac{\mathrm{d}q_i(t)}{\mathrm{d}t}\right) \equiv \frac{\mathrm{d}q_i'(t')}{\mathrm{d}t'} - \frac{\mathrm{d}q_i}{\mathrm{d}t} \qquad \text{(A4.6)}$$

$$\frac{\mathrm{d}q_i'(t')}{\mathrm{d}t'} = \frac{\mathrm{d}}{\mathrm{d}t}(q_i + \delta q_i)\frac{\mathrm{d}t}{\mathrm{d}t'} = \left(\frac{\mathrm{d}q_i}{\mathrm{d}t} + \frac{\mathrm{d}\delta q_i}{\mathrm{d}t}\right)\left(1 - \frac{\mathrm{d}\delta t}{\mathrm{d}t'}\right).$$

Comparing this equation with Equation A4.6, we obtain, to first order in the infinitesimal,

$$\delta\frac{\mathrm{d}q_i}{\mathrm{d}t} = \frac{\mathrm{d}\delta q_i}{\mathrm{d}t} - \frac{\mathrm{d}q_i}{\mathrm{d}t}\frac{\mathrm{d}\delta t}{\mathrm{d}t'}. \tag{A4.7}$$

Using Equation A4.7 in Equation A4.5 yields

$$\Delta I = \int\left[\sum_i\left(\frac{\partial L}{\partial q_i} + \frac{\partial L}{\partial \dot{q}_i}\frac{\mathrm{d}}{\mathrm{d}t}\right)\delta q_i + \frac{\partial L}{\partial t}\delta t - \dot{q}_i\frac{\partial L}{\partial \dot{q}_i}\frac{\mathrm{d}\delta t}{\mathrm{d}t} + L\frac{\mathrm{d}\delta t}{\mathrm{d}t}\right]\mathrm{d}t + \int\frac{\mathrm{d}}{\mathrm{d}t}\delta f(q_i,t)\,\mathrm{d}t$$

which may be rewritten as

$$\begin{aligned}\Delta I = &\int\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i\frac{\partial L}{\partial \dot{q}_i}\delta q_i + \left(L - \sum_i\frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i\right)\delta t + \delta f\right]\mathrm{d}t \\ &+ \int\sum_i\left(\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}\right)\delta q_i\delta t + \int\left(\frac{\partial L}{\partial t} + \frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i\dot{q}_i\frac{\partial L}{\partial \dot{q}_i}\right] - \frac{\partial L}{\partial t}\right)\delta t\mathrm{d}t.\end{aligned} \tag{A4.8}$$

But

$$\frac{\partial L}{\partial t} = \frac{\mathrm{d}L}{\mathrm{d}t} - \sum_i\frac{\partial L}{\partial q_i}\dot{q}_i - \sum_i\frac{\partial L}{\partial \dot{q}_i}\ddot{q}_i.$$

Substituting this last equation into Equation A4.8, we obtain

$$\begin{aligned}\Delta I = &\int\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i\frac{\partial L}{\partial \dot{q}_i}\delta q_i + \left(L - \sum_i\frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i\right)\delta t + \delta f\right]\mathrm{d}t \\ &+ \int\sum_i\left(\frac{\partial L}{\partial q_i} - \frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial L}{\partial \dot{q}_i}\right)(\delta q_i - \dot{q}_i)\mathrm{d}t\end{aligned}.$$

The second integrand of the last equation vanishes by making use of the Lagrange's equations of motion, and there remains

$$\Delta I = \int\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i\frac{\partial L}{\partial \dot{q}_i}\delta q_i + \left(L - \sum_i\frac{\partial L}{\partial \dot{q}_i}\delta\dot{q}_i\right)\delta t + \delta f\right]\mathrm{d}t. \tag{A4.9}$$

Now Noether's theorem states that if the action is invariant under the infinitesimal transformations (Equation A4.1), that is, $\Delta I = 0$, then it follows from Equation A4.9 that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left[\sum_i \delta q_i \frac{\partial L}{\partial \dot{q}_i} + \left(L - \sum_i \delta \dot{q}_i \frac{\partial L}{\partial \dot{q}_i}\right)\delta t + \delta f\right] = 0 \tag{A4.10}$$

and

$$\sum_i \delta q_i \frac{\partial L}{\partial \dot{q}_i} + \left(L - \sum_i \delta \dot{q}_i \frac{\partial L}{\partial \dot{q}_i}\right)\delta t + \delta f = \text{constant}. \tag{A4.11}$$

Equation A4.10 is the conservation law with the term in the brackets being the conserved quantity. Thus, for every infinitesimal symmetry transformation, there is a corresponding conservation law. By means of the following formula that is extracted from Equation A4.5:

$$\sum_i \left(\frac{\partial L}{\partial q_i}\delta q + \frac{\partial L}{\partial \dot{q}_i}\delta \dot{q}_i\right)_i + \frac{\partial L}{\partial t}\delta t + L\frac{\mathrm{d}\delta t}{\mathrm{d}t} = -\frac{\mathrm{d}}{\mathrm{d}t}\delta f(q_i, t), \tag{A4.12}$$

we can test whether a given transformation is a symmetrical one.

The test consists in demonstrating that the terms on the left-hand side of Equation A4.12 can be cast into a form similar to that on the right-hand side, that is, the total derivative with respect to the independent variable. If the left-hand side of Equation A4.12 vanishes identically, it follows immediately from Equation A4.5 that the Lagrangian is form-invariant under the transformation.

We now illustrate Noether's theorem with three familiar examples.

Example A4.1: A Pure Spatial Displacement of the Origin of the Reference System

Consider a Cartesian reference system, and the displacement of the origin is in the x_i direction by a finite amount. Then,

$$t' - t = \delta t = 0, \quad x_i'^{\alpha} - x_i^{\alpha} = \delta x_i^{\alpha} = C_i, \quad \delta \dot{x}_i^{\alpha} = 0 \tag{A4.13}$$

where C_i is a constant, and $x_i^{\alpha}(i = 1,2,3)$ is the x_i coordinate of the αth particle. Putting these into the test (Equation A4.12), we obtain

$$\sum_{\alpha=1}^{N} C_i \frac{\partial L}{\partial x_i^{\alpha}} = -\frac{\mathrm{d}}{\mathrm{d}t}\delta f\left(x_i^{\alpha}, t\right).$$

Thus, Equation A4.12 is a symmetry transformation leaving the Lagrangian form-invariant if the force $\partial L/\partial x_i^{\alpha}$ in the x_i direction vanishes. The conservation law associated with the symmetry transformation (Equation A4.12) is given by Equation A4.11, which takes the simple form:

$$\sum_{\alpha=1}^{N} C_i \frac{\partial L}{\partial \dot{x}_i^{\alpha}} = \text{constant}$$

or

$$\sum_{\alpha=1}^{N} p_i^{\alpha} = \text{constant}$$

which expresses the familiar result of the constancy of the component of the total linear momentum of the system along the x_i-axis. Thus, the total momentum of the system is conserved if space is homogeneous.

Example A4.2: A Pure Time Translation

We now consider displacement of the time origin. The infinitesimal translation in this case is

$$t' - t = \delta t_0, \quad \delta q_i = 0.\text{s} \tag{A4.14}$$

Applying the test (Equation A4.12), we obtain

$$\delta t_0 \frac{\partial L}{\partial t} = -\frac{\text{d}}{\text{d}t}\delta f(q_i, t).$$

Thus, if $\partial L/\partial t = 0$, that is, the Lagrangian L does not depend explicitly on time, then the time translation is a symmetry transformation under which the Lagrangian L is form-invariant.

Equation A4.11 now reduces to the simple relationship

$$L - \sum_i \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} = \text{constant}.$$

If we restrict to $L = T - V$ and $V = V(\dot{q}, t)$, then Euler's theorem gives

$$\sum_i \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} = 2T$$

and

$$L - \sum_i \dot{q}_i \frac{\partial L}{\partial \dot{q}_i} = (T - V) - 2T = -(T + V) = \text{constant}.$$

That is, energy conservation follows from the invariance of the laws of motion under a displacement of system in time (homogeneity of time).

Similarly, we can prove that if space is isotropic, angular momentum of the system follows from the invariance of the laws of motion under an infinitesimal rotation of the system.

Appendix 5: Conic Sections, Ellipse, Parabola, and Hyperbola

This appendix has two parts: the algebraic equation of a conic section and polar coordinate representation of conic sections. The reader who is not interested in algebraic equations of conic sections can go directly to the polar coordinate representations of conic sections, which is well suited for solving orbit problems.

A5.1 ALGEBRAIC EQUATIONS OF CONIC SECTIONS

The circle, ellipse, parabola, and hyperbola are called conic sections because each shape can be obtained by using a plane to intersect a cone. Any conic section can be represented by a second-order algebraic equation:

$$Ax^2 + Bxy + Cy^2 + Dx + Ey + F = 0 \tag{A5.1}$$

where the coefficients A, B, C, D, E, and F are constants. When A, B, and C are not all zero, the above equation represents either a conic section or a degenerate case of a conic section. Degenerate cases are summarized as follows for different values of the product AC of coefficients A and C:

Conic section	AC	Degenerate cases
Parabola	$= 0$	One line (two coincident lines)
Ellipse	> 0	Circle Point
Hyperbola	< 0	Two intersecting lines

Except for the above degenerate cases, the graph of Equation A4.1 can be determined by calculating the discriminant $\mathcal{D}$.

$$\mathcal{D} = B^2 - 4AC.$$

Value of discriminant	Conic section
$\mathcal{D} = 0$	Parabola
$\mathcal{D} < 0$	Ellipse
$\mathcal{D} > 0$	Hyperbola

Sets of coefficients for common conic sections are summarized below.

1. **Circle**
   If $A = C$, $F = -r^2$, and $B = D = E = 0$, we obtain

$$x^2 + y^2 = r^2 \text{ (constant } r\text{)}$$

   which is a circle of radius r.

2. **Ellipse**
   If $A = 1/a^2$, $C = 1/b^2$, $F = -1$, and $B = D = E = 0$, we obtain

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1, \quad a > b.$$

   The semi-major axis is a, and the semi-minor axis is b (Figure A5.1).
3. **Parabola**
   If $C = 1$, $D = -k$, and $A = B = E = F = 0$, we obtain (Figure A5.2)

$$y^2 = kx, \text{ a parabola.}$$

4. **Hyperbola**
   If $A = 1/a^2$, $C = -1/b^2$, $F = -1$, and $B = D = E = 0$, we obtain (Figure A5.3)

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1, \quad \text{a hyperbola.}$$

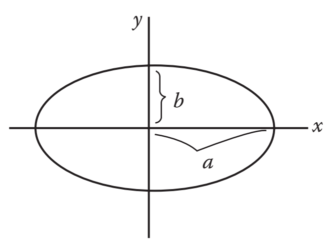


FIGURE A5.1 Ellipse.

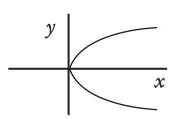


FIGURE A5.2 Parabola.

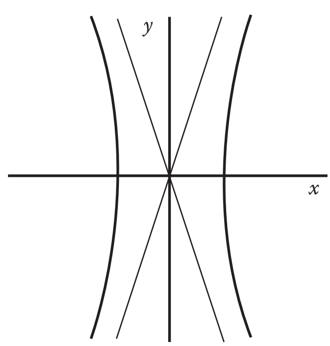


FIGURE A5.3 Hyperbola.

Notice that the only difference between an ellipse and a hyperbola is the sign of C. In general, a parabolic equation has the form $y = x^n$, and $n > 0$, and a hyperbola equation has the same form $y = x^n$ but with $n < 0$.

A5.2 POLAR COORDINATE REPRESENTATIONS OF CONIC SECTIONS

Consider a fixed point O and a fixed line AB distant D from O (Figure A5.4). By definition, a conic section is the curve traced by a point P in the plane of O, and AB moves so that the ratio of its distance r from O to its distance d from line AB is a positive constant ε: $r/d = \varepsilon$. The point O is called a focus, the line AB is called a directrix, the ratio ε is called the eccentricity, and the parameter p is called the latus rectum.

From Figure A5.4, we see that

$$D = d + r\cos\theta = \frac{r}{\varepsilon}\left(1+\varepsilon\cos\theta\right) \tag{A5.1}$$

or

$$r = \frac{\varepsilon D}{1+\varepsilon\cos\theta}. \tag{A5.2}$$

The angle θ is called the true anomaly. The parameter p equals r when $\theta = \pi/2$:

$$p = \frac{\varepsilon D}{1+\varepsilon\cos(\pi/2)} = \varepsilon D.$$

Accordingly, Equation A5.2 becomes

$$r = \frac{p}{1+\varepsilon\cos\theta}. \tag{A5.3}$$

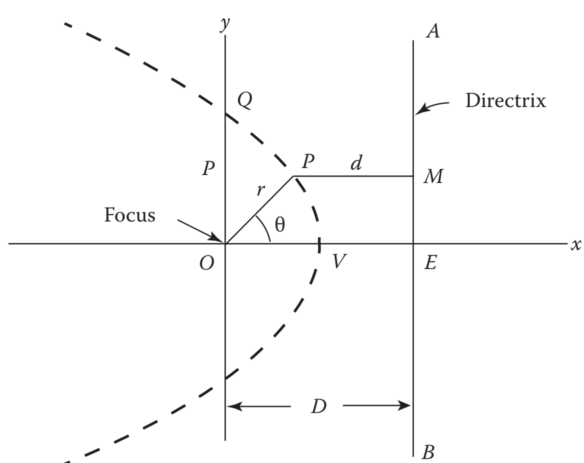


FIGURE A5.4 Conic sections.

Parameter p determines the extent of the orbit, and ε determines the shape. Shapes for different values of ε are tabulated below:

Eccentricity ε	Geometric Shape
0	Circle
$0 < \varepsilon < 1$	Ellipse
1	Parabola
$\varepsilon > 1$	Hyperbola

1. **Ellipse** $0 < \varepsilon < 1$

   An ellipse is a close curve traced by the moving point P such that the sum of the distance from two fixed points (O, O', the foci) is a constant, taken as $2a$: $r' + r = 2a$. Inspection of the figure shows that $2a$ is the major axis of the ellipse (Figure A5.5). Note that $2a$ is also equal to the sum of the minimum value of r_1 of the radius vector corresponding to $\theta = 0$ and the maximum value r_2 corresponding to $\theta = \pi$. We have, using Equation A5.3,

$$\frac{p}{1+\varepsilon} + \frac{p}{1-\varepsilon} = 2a.$$

   Solving p, we find

$$P = a(1 - \varepsilon^2). \tag{A5.4}$$

   And the equation of the ellipse is

$$r = \frac{a(1-\varepsilon^2)}{1+\varepsilon\cos\theta}. \tag{A5.5}$$

   Inspection of the figure also shows that $OS + O'S = 2a$ or $OS = a$. Let us denote OC by c; then $a = c + r(\theta = 0)$ or

$$a = c + \frac{a(1-\varepsilon^2)}{1+\varepsilon} = c + a(1-\varepsilon).$$

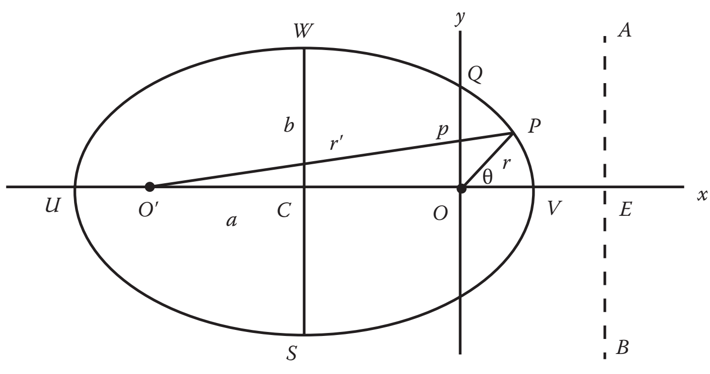


FIGURE A5.5 Ellipse, $0 < \varepsilon < 1$.

Solving for ε

$$\varepsilon = \frac{c}{a}. \quad \text{(A5.6)}$$

Inspection of the triangle *OCS* shows that

$$(OS)^2 = (OC)^2 + (CS)^2$$

or

$$a^2 = b^2 + c^2 = b^2 + a^2\varepsilon^2.$$

Solving for semi-minor axis *b*, we find

$$b = a(1 - \varepsilon^2)^{1/2}. \quad \text{(A5.7)}$$

In the special case of ε = 0, Equation A4.5 reduces to $r = a$; the ellipse degenerates into a circle.

2. **Parabola** ε = 1
The equation of the parabola is

$$r = \frac{p}{1 + \cos\theta} \quad \text{(A5.8)}$$

which is obtained by setting ε = 1 in Equation A5.3.

Recall the geometric definition of a parabola as the locus of points equally distant from a focus *O* and a fixed line *AB* (the directrix). In this case, it follows that

$$r_p = r(0°) = p/2.$$

Then we also have $d = r$.

From Equation A5.8, we observe that *r* goes to infinity as θ → 180° (Figure A5.6).

3. **Hyperbola** ε > 1
The hyperbola represents the special case of Equation A5.3 with an eccentricity ε > 1:

$$r = \frac{p}{1 + \varepsilon\cos\theta}, \quad \varepsilon > 1. \quad \text{(A5.9)}$$

The hyperbola consists of two branches as indicated in Figure A5.7 (next page).

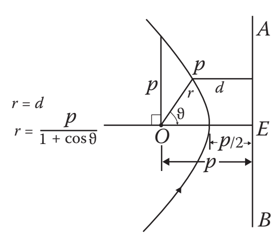


FIGURE A5.6 Parabola, ε = 1.

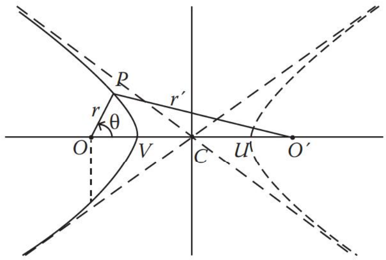


FIGURE A5.7 Hyperbola, $\varepsilon > 1$.

The hyperbola is asymptotic to the dashed lines, which are called its asymptotes. The intersection of the asymptotes is called the center. By analogy with the ellipse, the major axis $2a$ is the distance between vertices V and U. Then the distance from the center C to vertex V equal the semi-major axis a. And as in the case of the ellipse, we define

$$\varepsilon = 2c/a$$

where $2c$ is the interfocal distance. Now, in terms of these lengths, the hyperbola may be defined as the locus of points such that

$$r - r' = 2a.$$

From the last equation, we have

$$(r')^2 = (r + 2a)^2 = r^2 + 4ar + 4a^2. \tag{A5.10}$$

On the other hand, the law of cosines gives

$$(r')^2 = r^2 + c^2 - 2rc \cos\theta = r^2 + 4a^2\varepsilon^2 - 4ra\varepsilon \cos\theta.$$

Equating the right-hand side of the last equation with the right-hand side of Equation A5.10:

$$r^2 + 4a^2\varepsilon^2 - 4ra\varepsilon \cos\theta = r^2 + 4ar + 4a^2.$$

Solving for r, we find the equation of the hyperbola

$$r = \frac{a(\varepsilon^2 - 1)}{1 + \varepsilon\cos\theta}. \tag{A5.11}$$

Comparing this with Equation A5.9, we find

$$p = a(\varepsilon^2 - 1). \tag{A5.12a}$$

Note that Equation A5.11 gives only the left branch of the hyperbola shown in Figure A5.7. The right branch (shown dotted in the figure) can be represented by this equation only if r and θ are defined relative to O' instead of O. On the other hand, if r and θ are not redefined, the right branch is given by

$$r = \frac{a(\varepsilon^2 - 1)}{-1 + \varepsilon\cos\theta}. \tag{A5.12b}$$

MECHANICS

Classical Mechanics

Second Edition

Classical Mechanics, Second Edition presents a complete account of the classical mechanics of particles and systems for physics students at the advanced undergraduate level. The book evolved from a set of lecture notes for a course on the subject taught by the author at California State University, Stanislaus, for many years. It assumes the reader has been exposed to a course in calculus and a calculus-based general physics course. However, no prior knowledge of differential equations is required. Differential equations and new mathematical methods are developed in the text as the occasion demands.

The book begins by describing fundamental concepts, such as velocity and acceleration, upon which subsequent chapters build. The second edition has been updated with two new sections added to the chapter on Hamiltonian formulations, and the chapter on collisions and scattering has been rewritten. The book also contains three new chapters covering Newtonian gravity, the Hamilton-Jacobi theory of dynamics, and an introduction to Lagrangian and Hamiltonian formulations for continuous systems and classical fields. To help students develop more familiarity with Lagrangian and Hamiltonian formulations, these essential methods are introduced relatively early in the text.

The topics discussed emphasize a modern perspective, with special note given to concepts that were instrumental in the development of modern physics, for example, the relationship between symmetries and the laws of conservation. Applications to other branches of physics are also included wherever possible. The author provides detailed mathematical manipulations, while limiting the inclusion of the more lengthy and tedious ones. Each chapter contains homework problems of varying degrees of difficulty to enhance understanding of the material in the text. This edition also contains four new appendices on D'Alembert's principle and Lagrange's equations, derivation of Hamilton's principle, Noether's theorem, and conic sections.

CRC Press
Taylor & Francis Group
an informa business
www.taylorandfrancisgroup.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K16463
ISBN: 978-1-4665-6998-0
90000
9 781466 569980
www.crcpress.com